S0008M

Sannolikhetslära och statistik

Kurskompendium från Pelles Moln · 58 noter · 2026-10-19
pellesmoln.org/kurser/S0008M

Innehåll

SANSTAT - formelsammling 5

Lektion 1 – Sannolikhet introduktion · sid 43–57

Beskrivande statistik 8
Utfallsrum och händelser 13
Sannolikhetsmått 17
Klassiska sannolikhetsdefinitionen 26
Multiplikationsprincipen 33
Binomialkoefficient 38

Lektion 2 – Betingad sannolikhet · sid 43–69

Betingad sannolikhet 42
Lagen om total sannolikhet 48
Bayes sats 53

Lektion 3 – Oberoende händelser · sid 69–75

Oberoende händelser 58
Oberoende upprepningar 68

Lektion 4 – Diskreta stokastiska variabler · sid 81–97

Stokastisk variabel 72
Sannolikhetsfunktion 76
Fördelningsfunktion 79
Binomialfördelning 83
Hypergeometrisk fördelning 89
Poissonfördelning 94
Likformig fördelning 100
ffg-fördelning 104

Lektion 5 – Diskreta fördelningar, väntevärde och varians · sid 91–114

Väntevärde 107
Varians och standardavvikelse 113

Lektion 6 – Kontinuerliga stokastiska variabler · sid 115–136

Kontinuerliga stokastiska variabler 120
Frekvensfunktion 125
Mest sannolika värde 129

Lektion 7 – Kontinuerliga fördelningar, fortsättning · sid 136–150

Exponentialfördelning 133
Normalfördelning 141

Standardiserad normalfördelning 149
Median och kvantiler 153

Lektion 8 – Linjära funktioner och summor · sid 151–158, 177–181

Linjära funktioner av stokastiska variabler 157
Oberoende stokastiska variabler 162

Lektion 9 – Summor av normalfördelade stokastiska variabler · sid 181–189

Summor av normalfördelade stokastiska variabler 165
Medelvärdets fördelning 171

Lektion 10 – Centrala gränsvärdessatsen · sid 188–199

Centrala gränsvärdessatsen 175
Normalapproximation 179

Lektion 11 – Punktskattningar, introduktion till konfidensintervall · sid 201–211, 219–225, 226–234

Statistisk inferens 183
Punktskattning 187
Teckenintervall 192
Konfidensintervall 197
t-fördelningen 204

Lektion 12 – Konfidensintervall fortsättning, stickprov i par · sid 238–241

Stickprov i par 210

Lektion 13 – Stickprov i par och två stickprov · sid 241–247

Två stickprov 216

Lektion 14 – Enkla hypoteser · sid 261–267

Hypotesprövning 223
Enkla hypoteser 228
Signifikansnivå och styrka 232

Lektion 15 – Sammansatta hypoteser · sid 268–286

Sammansatta hypoteser 237
Styrkefunktion 243

Lektion 16 – Sammansatta hypoteser, direktmetoden · sid 278–299

Direktmetoden 246
Konfidensintervall och hypotesprövning 251

Lektion 17 – Flerdimensionella fördelningar, inledning · sid 359–364, 371–376

Flerdimensionella fördelningar 254
Marginalfördelning 265

Betingad fördelning och betingat väntevärde 270

Lektion 18 – Korrelation och tvådimensionell normalfördelning · sid 377–385

Kovarians 281

Korrelation 287

Tvådimensionell normalfördelning 293

Lektion 19 – Icke-linjära funktioner av stokastiska variabler · sid 161–164, 365–369

Icke-linjära funktioner av stokastiska variabler 302

Max och min av stokastiska variabler 312

Gamla tentor

S0008M – sammanfattning inför tentan 318

SANSTAT - formelsammling

19 okt. 2026, 2 min läsning

UTKAST

#matematik #statistik #formelsamling

Def 2.1 Den klassiska sannolikhetsdefinitionen

$$P(A) = \frac{g}{m}$$

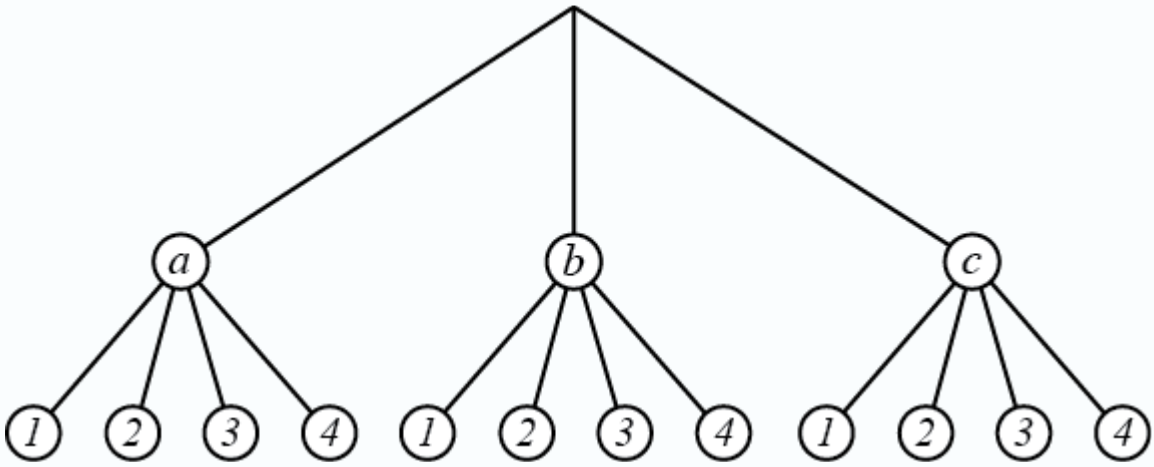
Där:

- g är gynsamma utfall
- m är möjliga utfall

Def 2.2 Multiplikationsprincipen

$$n_1 * n_2 * \cdots * n_{k-1} * n_k$$

Om man genomför k antal operationer, och varje operation har n antal möjliga utfall så kan man multiplicera dessa möjliga utfall för att få totala mängden möjliga utfall



Sats 2.1

n \\ k \end{pmatrix}= \frac{n!}{k!(n-k)!}$$ Antalet möjliga sätt att välja ut k antalet element bland n element utan återläggning mellan dragningarna och utan hänsyn till inebördens ordning bland de utvalda element

Def 2.3

En funktion P som till varje händelse A i ett utfallsrum Ω ordnar ett reelt tal $P(A)$ sägs vara ett sannolikhetsmått om P har en av följande egenskaper

- $0 \leq P(A) \leq 1$ för alla A
- $P(\Omega) = 1$
- $P(A \cup B) = P(A) + P(B)$ om A och B är disjunkta, dvs. inte kan inträffa samtidigt

Sats 2.3

Följande gäller: $P(A^C) = 1 - P(A)$

Sats 2.4

Följande gäller: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Betingad

Def 2.4

med den betingade sannolikheten $P(A|B)$ för händelsen A, givet att B, menas:
$P(A|B) = \frac{P(A \cap B)}{P(B)}$

Sats 2.7 Lagen om total sannolikhet

Antag att häldelserna $B_1, B_2, \ldots, B_n$ är disjunkta och tillsamman utgör hela utfallsrummet. Då gäller för varje händelse A att

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n)$$

Oberoende

Def 2.5 (sid: 70)

Om $P(A \cap B) = P(A) * P(B)$, säger vi att A och B är oberoende händelser

Sats 2.8 (sid: 74)

Betrakta ett försök bestående av n oberoende upprepningar av delförsök. Låt en händelse som inträffar vid delförsöket med sannolikhet p. Då är sannolikheten att A inträffar exakt k gånger i försöket

n \\ k \end{pmatrix}p^k(1-p)^{n-k}$$

Diskreta fördelningar

Def 2.5 (sid: 70)

Utfallet i ett slumpmässigt försök i form av ett Reelt tal, betrakta innan försöket är utfört försöket kallas observat värde på den stokastiska va

Beskrivande statistik

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M

Beskrivande statistik sammanfattar ett datamaterial med några få tal och diagram: var värdena ligger, hur utspridda de är och vilken form fördelningen har. Det är det första steget i varje analys och ämnet för laboration 1. Här modelleras ingenting ännu; allt räknas direkt på de observerade värdena.

1. Lägesmått och spridningsmått

Lägesmått säger *var* värdena ligger, **spridningsmått** *hur utspridda* de är.

Lägesmått	Spridningsmått
medelvärde $\bar{x} = \frac{1}{n}\sum x_i$	standardavvikelse s
median	variationsbredd, max − min
	kvartilavstånd

Standardavvikelsen räknas med nämnaren $n-1$, samma s som i Punktskattning:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Medianen påverkas inte av enstaka extrema värden, medelvärdet gör det. Ligger de nära varandra är materialet ungefär symmetriskt.

2. Median och kvartiler enligt Tukey

Sortera värdena. **Djupet** är hur långt in från kanten ett värde ligger i den sorterade listan.

$$\text{medianens djup} = \frac{n+1}{2}, \qquad \text{kvartilernas djup} = \frac{\lfloor \text{medianens djup} \rfloor + 1}{2}$$

Nedre kvartilen Q_1 räknas från botten, övre kvartilen Q_3 från toppen. Slutar ett djup på $,5$ tas medelvärdet av de två värdena runt det.

$$\text{kvartilavstånd} = Q_3 - Q_1$$

> **Andra program räknar annorlunda**
>
> Minitab och räknaren interpolerar kvartilerna på andra sätt och kan ge något andra värden än Tukeys djup. I laboration 1 gav Minitab $Q_1 = 6{,}825$ mot Tukeys $6{,}85$. Skillnaden är liten och ändrar sällan slutsatsen.

Motsvarande begrepp för en stokastisk variabel finns i Median och kvantiler.

---

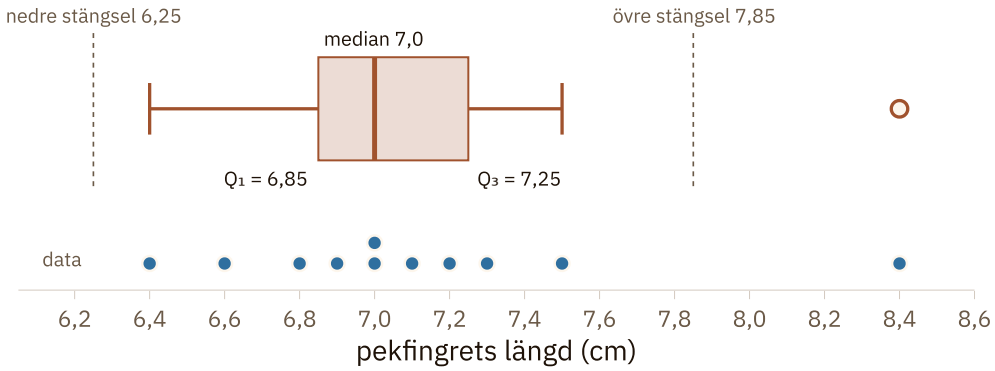
3. Stängsel, uteliggare och lådagram

Stängslen ligger $1{,}5$ kvartilavstånd utanför lådan:

$$\text{nedre stängsel} = Q_1 - 1{,}5\,(Q_3 - Q_1), \qquad \text{övre stängsel} = Q_3 + 1{,}5\,(Q_3 - Q_1)$$

Värden utanför stängslen är **uteliggare**.

I ett **lådagram** går lådan från Q_1 till Q_3 med ett streck vid medianen. Morrhåren går ut till det mest extrema värdet som fortfarande ligger *innanför* stängslen, och uteliggarna ritas som egna punkter. Själva stängslen ritas inte ut.



Figuren hör till exemplet nedan. Stängslen är utritade streckade för att visa var de ligger; i ett vanligt lådagram syns de inte.

Exempel — Tukeys mått för elva värden

Elva pekfingerlängder (cm), sorterade:

$$6{,}4 \quad 6{,}6 \quad 6{,}8 \quad 6{,}9 \quad 7{,}0 \quad 7{,}0 \quad 7{,}1 \quad 7{,}2 \quad 7{,}3 \quad 7{,}5 \quad 8{,}4$$

Bestäm median, kvartiler, kvartilavstånd, stängsel och uteliggare, och ange var morrhåren slutar. Beräkna också $\bar{x}$ och s.

Lösning

Median. Djupet är $(11+1)/2 = 6$; det sjätte värdet är $\boxed{7{,}0}$.

Kvartiler. Djupet är $(6+1)/2 = 3{,}5$, alltså medel av tredje och fjärde värdet från respektive ände:

$$Q_1 = \frac{6{,}8 + 6{,}9}{2} = \boxed{6{,}85}, \qquad Q_3 = \frac{7{,}3 + 7{,}2}{2} = \boxed{7{,}25}, \qquad Q_3 - Q_1 = 0{,}4$$

Stängsel. $1{,}5 \cdot 0{,}4 = 0{,}6$:

$$6{,}85 - 0{,}6 = \boxed{6{,}25}, \qquad 7{,}25 + 0{,}6 = \boxed{7{,}85}$$

$8{,}4$ ligger ovanför övre stängslet och är en **uteliggare**. Morrhåren går till $6{,}4$ och $7{,}5$.

Medelvärde och standardavvikelse.

$$\bar{x} = \frac{78{,}2}{11} \approx \boxed{7{,}11}, \qquad s \approx \boxed{0{,}53}$$

Medelvärdet ligger över medianen, dragen uppåt av uteliggaren.

På räknaren. One-Variable Statistics ger `x̄`, `sx` och `MedianX`. Kvartilerna `Q₁X` och `Q₃X` räknas med en annan metod och blir $6{,}8$ och $7{,}3$ här; använd Tukeys djup när uppgiften kräver det.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

4. Stam-bladdiagram och histogram

Ett **stam-bladdiagram** sorterar data för hand och visar formen utan att tappa några värden. För $7{,}3$ är 7 stammen och 3 bladet. Blir det för få stammar delas varje stam på flera rader.

Ett **histogram** delar in värdena i klasser och ritar en stapel per klass. För få klasser döljer formen, för många ger ett taggigt diagram. En tumregel är ungefär $\sqrt{n}$ klasser, och klassbredden bör vara en multipel av mätnoggrannheten så att inga klasser blir tomma av avrundningsskäl.

Snedhet. En lång svans åt höger betyder att fördelningen är **högerskev** (positivt sned), en lång svans åt vänster att den är **vänsterskev**.

5. Två stickprov eller par

Samma skillnad som i inferensdelen gäller redan här. Jämförs två olika grupper, till exempel män och kvinnor, är det två stickprov och vi ritar ett lådagram per grupp. Mäts samma person två gånger, till exempel dominant och icke-dominant hand, är det stickprov i par: räkna ut differensen per person och beskriv differenserna.

⚠ **Vanliga fel**

- **Blanda läges- och spridningsmått.** Kvartilerna själva är lägesmått; det är *avståndet* mellan dem som mäter spridning.
- **Morrhår till stängslen.** Morrhåren slutar vid det mest extrema *observerade* värdet innanför stängslen, inte vid stängslet.
- **`σx` i stället för `sx`.** Stickprovets standardavvikelse har nämnaren $n - 1$.
- **Jämföra par som två grupper.** Två mätningar per person är inte oberoende; titta på differenserna.

Se även

- Median och kvantiler
- Varians och standardavvikelse
- Punktskattning
- Två stickprov
- Stickprov i par
- S0008M Lab 1 ifylld

Resurser

- Khan Academy: Summarizing quantitative data
- Wikipedia: Box plot

Utfallsrum och händelser

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #grundbegrepp

> **Kurs:** S0008M **Förkunskaper:** Mängdlära

Utfallsrummet är mängden av alla möjliga utfall i ett slumpförsök. En händelse är en delmängd av utfallsrummet, alltså en samling utfall som vi vill tilldela en sannolikhet.

---

1. Slumpförsök och utfallsrum

Ett **slumpmässigt försök** är ett försök vars resultat vi inte kan förutsäga med säkerhet. Mängden av alla tänkbara utfall kallas **utfallsrummet** och betecknas Ω.

Köper vi tre maskiner och noterar för var och en om den fungerar (F) eller inte (O) efter sex månader är

$$\Omega = \{FFF,\ FFO,\ FOF,\ FOO,\ OFF,\ OFO,\ OOF,\ OOO\}$$

Utfallsrummet delas in efter hur många utfall det har:

Typ	Utfall	Exempel
Diskret	ändligt eller uppräkneligt många	tre maskiner ovan; antal pulser per 5 s, $\Omega = \{0, 1, 2, \dots\}$
Kontinuerligt	ett eller flera intervall	kraften i ett balkstöd, $\Omega = \{x \mid 0 \le x \le 100\}$

Typen avgör hur vi räknar sannolikheter längre fram: summor i det diskreta fallet, integraler i det kontinuerliga.

---

2. Händelser

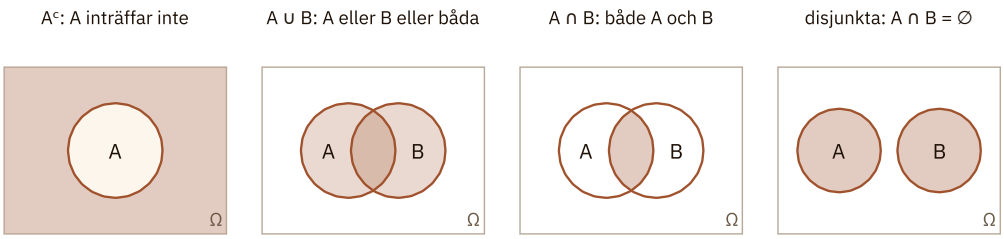
En **händelse** är en delmängd av Ω och betecknas med stor bokstav, A, B, C. Att ”A inträffar” betyder att utfallet hamnar i A. Att exakt en av de tre maskinerna fungerar är till exempel

$$A = \{FOO,\ OFO,\ OOF\}$$

3. Mängdoperationer på händelser

Händelser kombineras med mängdoperationer:

Beteckning	Namn	Inträffar när
A^c	komplementhändelse	A inte inträffar
$A \cup B$	union	A eller B eller båda inträffar
$A \cap B$	snitt	både A och B inträffar
$A \cap B = \varnothing$	disjunkta händelser	A och B kan inte inträffa samtidigt



Rita figur

Nästan varje uppgift med två eller tre händelser blir enklare med ett mängddiagram. Händelsen "exakt en av A och B" är svår att skriva direkt, men i figuren syns den som $(A \cap B^c) \cup (B \cap A^c)$ — de två månskärorna utanför snittet.

Hur vi sätter sannolikheter på händelserna och räknar med dem beskrivs i Sannolikhetsmått.

Exempel — Två tärningar

Vi kastar en röd och en blå tärning. Låt A = "summan är 7" och B = "minst en sexa".

a) Beskriv utfallsrummet. Hur många utfall har det?

b) Skriv upp A, $A \cap B$ och B^c. Hur många utfall har $A \cup B$?

Lösning

a) Ett utfall är ett par (r, b) med röda tärningens resultat först:

$$\Omega = \{(r, b) : r, b = 1, \dots, 6\}, \qquad |\Omega| = 6 \cdot 6 = 36$$

Att hålla isär tärningarna gör alla 36 utfall lika sannolika. Räknar vi i stället summan som utfall, $\{2, \dots, 12\}$, är utfallen *inte* lika sannolika.

b)

$$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}, \qquad A \cap B = \{(1,6), (6,1)\}$$

B^c = "ingen sexa" = alla par med $r, b \leq 5$, alltså 25 utfall. Därmed har B $36 - 25 = 11$ utfall, och

$$|A \cup B| = |A| + |B| - |A \cap B| = 6 + 11 - 2 = \boxed{15}$$

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **"Eller" betyder "minst en av".** $A \cup B$ inträffar också när båda inträffar. "Exakt en av" är en annan händelse, se tipset ovan.
- **Välja ett utfallsrum där utfallen inte är lika sannolika** och sedan räkna g/m. Summan av två tärningar har elva värden, men de är inte lika sannolika. Håll isär tärningarna.

Se även

- Mängdlära
- Sannolikhetsmått
- Klassiska sannolikhetsdefinitionen
- Betingad sannolikhet
- Stokastisk variabel

Resurser

- Khan Academy: Probability
- Wikipedia: Sample space

Sannolikhetsmått

19 okt. 2026, 8 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Utfallsrum och händelser

---

Ett sannolikhetsmått tilldelar varje händelse ett tal mellan 0 och 1 så att hela utfallsrummet får 1 och disjunkta händelser adderas. Ur de tre kraven följer alla räkneregler, bland dem komplementregeln och additionssatsen.

---

1. Definitionen

Den klassiska definitionen $P(A) = g/m$ gäller bara när utfallen är ändligt många och lika sannolika. I stället för att definiera *vad* $P(A)$ är i allmänhet kräver vi att sannolikheter har vissa egenskaper.

Definition 2.3. En funktion P som till varje händelse A i ett utfallsrum Ω ordnar ett reellt tal $P(A)$ kallas ett **sannolikhetsmått** om

a) $0 \le P(A) \le 1$ för alla A

b) $P(\Omega) = 1$

c) $P(A \cup B) = P(A) + P(B)$ om A och B är disjunkta

$P(A)$ kallas då **sannolikheten för** A.

---

2. Masstolkningen

Sannolikhet som massa

Tänk på sannolikhet som **massa** utspridd över Ω, med total massa 1. Då är $P(A)$ massan på A, och de tre egenskaperna blir självklara: massan är aldrig negativ och aldrig mer än helheten, hela Ω väger 1, och två områden som inte överlappar väger tillsammans summan av sina massor.

Masstolkningen är också det enklaste sättet att troliggöra räknereglerna nedan: rita mängddiagrammet och fråga var massan hamnar.

---

3. Komplementregeln

Sats 2.3.

$$\boxed{P(A^c) = 1 - P(A)}$$

A och A^c är disjunkta och fyller tillsammans hela Ω, så deras massor summerar till 1.

"Minst en" via komplementet

"Minst en av händelserna inträffar" är $A \cup B$, och "ingen inträffar" är dess komplement. Därför

$$P(\text{ingen}) = P((A \cup B)^c) = 1 - P(A \cup B)$$

När "minst en" är krånglig att räkna direkt är "ingen" ofta enkel, och tvärtom.

---

4. Additionssatsen

Sats 2.4.

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Adderar vi massan på A och massan på B har snittet $A \cap B$ räknats två gånger. Vi drar bort det en gång. Är A och B disjunkta är $P(A \cap B) = 0$ och vi är tillbaka i egenskap c).

Tre händelser

Samma räkning med tre händelser: snitten parvis har räknats två gånger och dras bort, men då har trippelsnittet först lagts till tre gånger och dragits bort tre gånger, så det måste läggas tillbaka:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C$$

⚠ Vanliga fel

- **Addera utan att dra bort snittet.** $P(A \cup B) = P(A) + P(B)$ gäller bara för disjunkta händelser. Ett svar större än 1 är ett säkert tecken.
- **Blanda ihop disjunkt och oberoende.** Disjunkt betyder $P(A \cap B) = 0$. Oberoende betyder $P(A \cap B) = P(A)\,P(B)$, se Oberoende händelser.
- **Räkna “minst en” direkt** när komplementet “ingen” är en enda term.

---

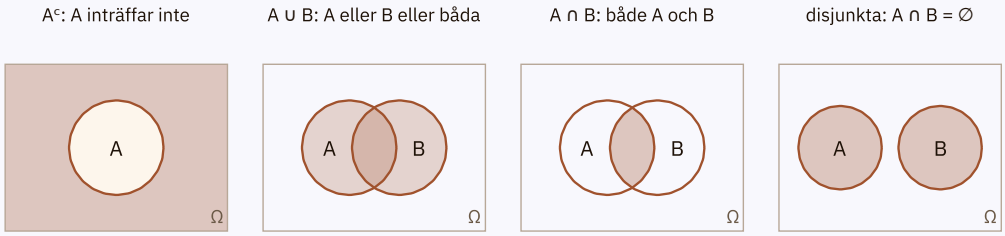
5. Lösta lektionsuppgifter

Lektion 1, övning 7 — Satserna 2.3 och 2.4 med masstolkning

Troliggör komplementregeln (Sats 2.3) och additionssatsen (Sats 2.4) med hjälp av masstolkningen eller en egen tolkning. Rita figur.

Lösning

Tänk på sannolikheten som massa utspridd över Ω, med total massa 1 (avsnitt 2).



Sats 2.3. A och A^c överlappar inte och täcker tillsammans hela Ω. Deras massor summerar därför till hela massan: $P(A) + P(A^c) = 1$, alltså $P(A^c) = 1 - P(A)$.

Sats 2.4. Dela $A \cup B$ i tre bitar som inte överlappar: $A \cap B^c$, $A \cap B$ och $B \cap A^c$. Massan på A är de två första bitarna och massan på B de två sista. Adderar vi $P(A)$ och $P(B)$ har mittbiten $A \cap B$ kommit med två gånger, så vi drar bort den en gång:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Exempel. Med $P(A) = 0{,}1$, $P(B) = 0{,}2$ och $P(A \cap B) = 0{,}05$ (övning 8 nedan) blir $P(A \cup B) = 0{,}25$ – lägg ihop de tre områdena i figuren till övning 8 och se att det stämmer.

Resonemangsuppgift – det finns inget facit att jämföra med.

Lektion 1, övning 8 (Vännman 2.15) – Två fel på byggelement

Vid tillverkning av en viss typ av byggelement kan två slags fel, A och B, uppkomma. Vi vet att $P(A) = 0{,}1$, $P(B) = 0{,}2$ och $P(A \cap B) = 0{,}05$. Beräkna sannolikheten att en tillverkad enhet har

a) åtminstone något av felen,

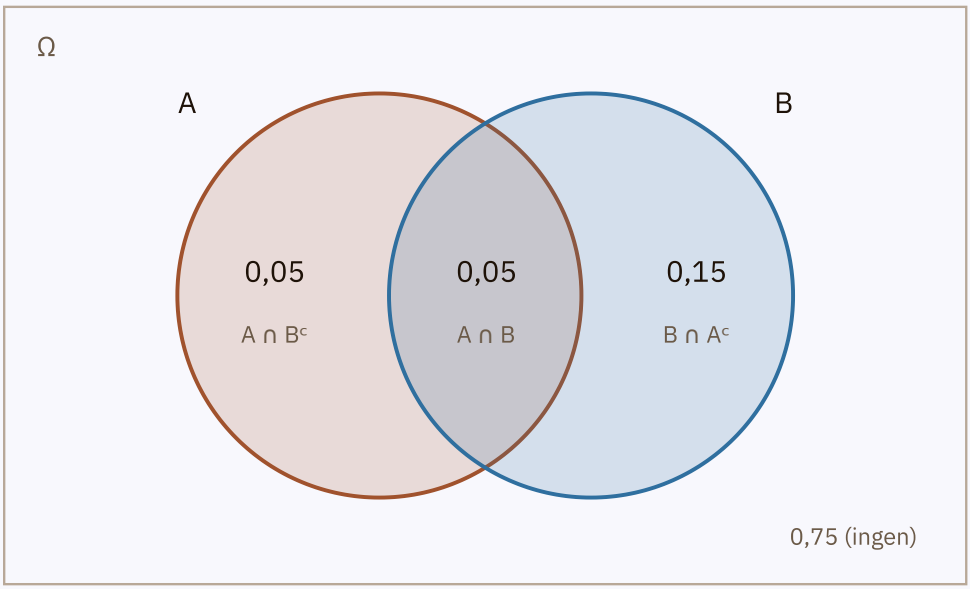
b) felet A men inte felet B,

c) felet B men inte felet A,

d) exakt ett av felen A och B.

Lösning

Figuren fördelar sannolikhetsmassan på de fyra områdena; b)–d) kan läsas av direkt.



a) "Åtminstone något" är $A \cup B$. Additionssatsen:

$$P(A \cup B) = 0{,}1 + 0{,}2 - 0{,}05 = \boxed{0{,}25}$$

b) A delas i två disjunkta bitar, $A \cap B$ och $A \cap B^c$, så $P(A) = P(A \cap B) + P(A \cap B^c)$:

$$P(A \cap B^c) = P(A) - P(A \cap B) = 0{,}1 - 0{,}05 = \boxed{0{,}05}$$

c) På samma sätt:

$$P(B \cap A^c) = 0{,}2 - 0{,}05 = \boxed{0{,}15}$$

d) "Exakt ett" är de två disjunkta månskärorna från b) och c):

$$P((A \cap B^c) \cup (B \cap A^c)) = 0{,}05 + 0{,}15 = \boxed{0{,}20}$$

Rimlighetskoll. "Exakt ett" plus "båda" ska vara "minst ett": $0{,}20 + 0{,}05 = 0{,}25$, som i a).

Svar i boken (2:a uppl., 2.15): a) 0,25, b) 0,05, c) 0,15, d) 0,20.

Lektion 1, övning 9 (Vännman 2.16) — Tre tidningar

I en tätort med 30 000 invånare finns tre tidningar, A, B och C. En undersökning visade att

Läser	Antal	Läser	Antal
A	12 000	A och B	7 000
B	8 000	A och C	4 500
C	6 000	B och C	1 000
		A, B och C	500

Hur stor är sannolikheten att en slumpmässigt vald invånare läser

a) minst en av tidningarna,

b) ingen av tidningarna,

c) en och endast en av tidningarna?

Lösning

Med invånaren vald på måfå är sannolikheten för varje grupp dess andel av 30 000:

$$P(A) = \frac{12}{30}, \quad P(B) = \frac{8}{30}, \quad P(C) = \frac{6}{30}, \quad P(A \cap B) = \frac{7}{30}, \quad P(A \cap C) = \frac{4{,}5}{30}, \quad P($$

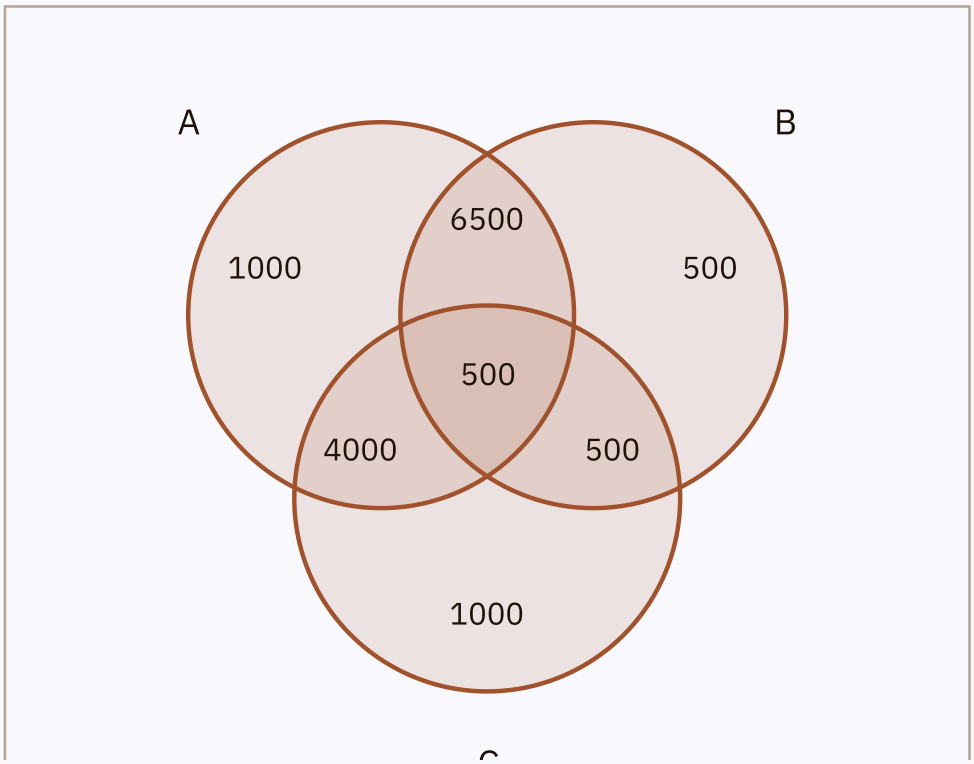
a) Additionssatsen för tre händelser:

$$P(A \cup B \cup C) = \frac{12 + 8 + 6 - 7 - 4{,}5 - 1 + 0{,}5}{30} = \frac{14}{30} = \boxed{\frac{7}{15}} \approx 0{,}47$$

b) Komplementet till a):

$$P((A \cup B \cup C)^c) = 1 - \frac{7}{15} = \boxed{\frac{8}{15}} \approx 0{,}53$$

c) Fyll i mängddiagrammet inifrån och ut: trippelsnittet först, sedan de parvisa snitten minus trippelsnittet, sist de tre ytterområdena.



De som läser exakt en tidning är de tre ytterområdena: $1000 + 500 + 1000 = 2500$.

$$P(\text{exakt en}) = \frac{2500}{30\,000} = \boxed{\frac{1}{12}} \approx 0{,}083$$

Rimlighetskoll. Summan av alla sju områden i figuren är $14\,000$, samma tal som gav $7/15$ i a).

Svar i boken (2:a uppl., 2.16): a) 0,47, b) 0,53, c) 0,08. Stämmer också med det handskrivna lösningsförslaget.

Lektion 2, övning 2 (Vännman 2.10 och 2.11) — Disjunkta händelser

2.10 Beräkna $P(B)$ om A och B är disjunkta händelser sådana att $P(A) = 0{,}25$ och $P(A \cup B) = 0{,}75$.

2.11 Två händelser A och B har sannolikheterna $0{,}6$ respektive $0{,}7$. Kan händelserna vara disjunkta?

Lösning

2.10 Disjunkta händelser har $P(A \cap B) = 0$, så egenskap c) i definitionen ger $P(A \cup B) = P(A) + P(B)$:

$$P(B) = P(A \cup B) - P(A) = 0{,}75 - 0{,}25 = \boxed{0{,}5}$$

2.11 Om de vore disjunkta skulle $P(A \cup B) = 0{,}6 + 0{,}7 = 1{,}3 > 1$, vilket strider mot $P(A \cup B) \leq 1$. **Nej**, de kan inte vara disjunkta. Additionssatsen ger dessutom att snittet måste väga minst $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq 1{,}3 - 1 = 0{,}3$.

I masstolkningen: två bitar med massorna $0{,}6$ och $0{,}7$ får inte plats i en total massa 1 utan att överlappa.

Svar i boken (2:a uppl., 2.10 och 2.11): 0,5 och Nej.

Lektion 2, övning 3 — Fel på växellådor

Vid tillverkning av växellådor kan två typer av fel uppkomma, fel A och fel B. Sannolikheten för fel A är $0{,}03$ och för fel B är den $0{,}05$. Båda felen kan uppkomma samtidigt, med sannolikheten $0{,}015$. Vad är sannolikheten att en slumpmässigt utvald växellåda har

a) minst ett av felen,

b) inte något fel,

c) exakt ett fel?

Lösning

Givet: $P(A) = 0{,}03$, $P(B) = 0{,}05$, $P(A \cap B) = 0{,}015$.

a) Additionssatsen:

$$P(A \cup B) = 0{,}03 + 0{,}05 - 0{,}015 = \boxed{0{,}065}$$

b) Komplementet till a):

$$P((A \cup B)^c) = 1 - 0{,}065 = \boxed{0{,}935}$$

c) “Exakt ett” är “minst ett” minus “båda”:

$$P(\text{exakt ett}) = P(A \cup B) - P(A \cap B) = 0{,}065 - 0{,}015 = \boxed{0{,}05}$$

Ett annat sätt, via månskärorna: $P(A \cap B^c) + P(B \cap A^c) = (0{,}03 - 0{,}015) + (0{,}05 - 0{,}015) = 0{,}015 + 0{,}035 = 0{,}05$.

Facit: a) 0,065, b) 0,935, c) 0,05.

Se även

- Utfallsrum och händelser
- Klassiska sannolikhetsdefinitionen
- Betingad sannolikhet
- Oberoende händelser
- Mängdlära

Resurser

- Khan Academy: Probability
- Wikipedia: Probability axioms

Klassiska sannolikhetsdefinitionen

19 okt. 2026, 6 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Utfallsrum och händelser

---

Om alla utfall i ett ändligt utfallsrum är lika sannolika ges en händelses sannolikhet av antalet gynnsamma utfall dividerat med antalet möjliga. Definitionen förutsätter alltså symmetri och duger inte när utfallen har olika sannolikhet.

---

1. Definitionen

Antag att utfallsrummet Ω består av m möjliga utfall och att alla är lika sannolika. Om g av dem är **gynnsamma**, det vill säga medför händelsen A, så är

$$\boxed{P(A) = \frac{g}{m}}$$

Två villkor

1. **Ändligt många utfall** m.
2. **Alla utfall lika sannolika.**

Antalet pulser från ett radioaktivt preparat under 5 sekunder har utfallsrummet $\{0, 1, 2, \dots\}$, alltså oändligt många utfall. Där fungerar definitionen inte, och vi får i stället skatta sannolikheten ur relativa frekvenser.

---

2. Att räkna g och m

Hela svårigheten i en uppgift ligger i att **räkna** g och m. Två verktyg räcker långt:

- Multiplikationsprincipen räknar **ordnade** val — "först en, sedan en till".
- Binomialkoefficienten $\binom{n}{k}$ räknar val **utan hänsyn till ordning**.

⚠ Välj ett sätt och håll fast vid det

Samma problem kan lösas både **med** och **utan** hänsyn till ordning, men g och m måste räknas på samma sätt. Att ta m som ordnade val och g som oordnade ger fel svar.

---

3. Dragning ur en urna

Den vanligaste tillämpningen: en låda med N föremål, varav a är av en viss sort (“defekta”) och $N - a$ av en annan. Vi drar n utan återläggning och frågar efter sannolikheten att få exakt k av den första sorten. Utan hänsyn till ordning är

$$m = \binom{N}{n}, \qquad g = \binom{a}{k}\binom{N-a}{n-k}$$

där g följer av multiplikationsprincipen: först väljer vi vilka k av de a defekta, sedan vilka $n - k$ av de $N - a$ felfria. Samma resonemang blir senare den hypergeometriska fördelningen.

---

4. Lösta lektionsuppgifter

Lektion 1, övning 1 — Tolv lappar i en låda

I en låda finns tolv lappar, numrerade från 1 till 12. Vi drar en lapp slumpmässigt. Definiera utfallsrummet och bestäm, med den klassiska sannolikhetsdefinitionen, sannolikheten att vi drar en lapp med nummer fyra eller lägre. Vilken händelse har sannolikheten beräknats för, och vilka förutsättningar krävs för att definitionen ska få användas?

Lösning

Utfallsrum. $\Omega = \{1, 2, \dots, 12\}$, där till exempel talet 7 står för att vi drar lappen med siffran 7. Alltså $m = 12$.

Händelse. A = “lapp med nummer fyra eller lägre”, formellt $A = \{1, 2, 3, 4\}$, så $g = 4$.

$$P(A) = \frac{g}{m} = \frac{4}{12} = \boxed{\frac{1}{3}}$$

Förutsättningar. Ändligt många utfall som alla kan antas lika sannolika — vilket de är om lapparna är likadana och dragningen görs på måfå.

Facit: $1/3$.

Lektion 1, övning 3 (Vännman 2.1) — Defekta muttrar

Bland 10 muttrar finns 4 defekta. Vi tar 4 muttrar på måfå utan återläggning. Vad är sannolikheten att vi får

a) alla 4 defekta,

b) först 2 defekta och sedan 2 korrekta,

c) exakt 2 defekta?

Vad blir g och m i varje fall?

Lösning

a) Ordningen spelar ingen roll. $m = \binom{10}{4} = 210$ och bara ett urval består av alla fyra defekta: $g = \binom{4}{4} = 1$.

$$P = \frac{1}{210} \approx \boxed{0{,}0048}$$

b) Här frågar vi efter en **ordning**, så vi räknar ordnade val. $m = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$. Gynnsamt: två defekta först ($4 \cdot 3$ sätt) och sedan två korrekta ($6 \cdot 5$ sätt), alltså $g = 4 \cdot 3 \cdot 6 \cdot 5 = 360$.

$$P = \frac{360}{5040} = \frac{1}{14} \approx \boxed{0{,}071}$$

c) Ordningen spelar ingen roll igen. Välj 2 av de 4 defekta och 2 av de 6 korrekta: $g = \binom{4}{2}\binom{6}{2} = 6 \cdot 15 = 90$, och $m = 210$.

$$P = \frac{90}{210} = \frac{3}{7} \approx \boxed{0{,}429}$$

Rimlighetskoll. b) är en av $\binom{4}{2} = 6$ lika sannolika ordningar för två defekta och två korrekta, så $6 \cdot \frac{1}{14} = \frac{3}{7}$ ska vara svaret i c). Det stämmer.

På räknaren. `nCr(4,2)*nCr(6,2)/nCr(10,4)` ger $\frac{3}{7}$ exakt; `ctrl` + `enter` ger närmevärdet.

Svar i boken (2:a uppl., 2.1): a) 0,0048, b) 0,071, c) 0,429.

Lektion 1, övning 6 (Vännman 2.2) — Glödlampor

I en låda ligger 15 glödlampor, varav 5 är defekta. Vi tar 5 lampor på måfå utan återläggning. Beräkna sannolikheten att vi får

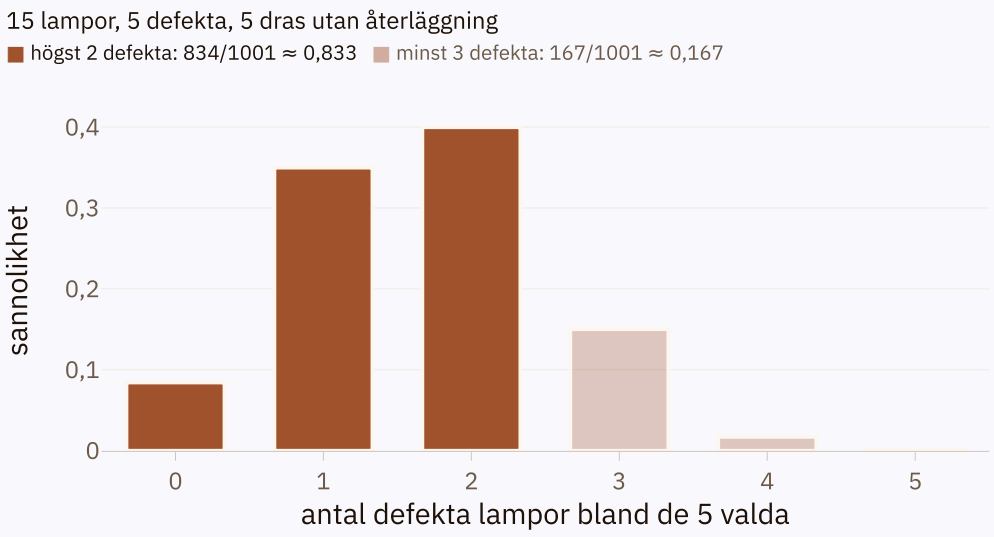
a) ingen defekt lampa, b) exakt 1 defekt, c) exakt 2 defekta,

d) högst 2 defekta, e) minst 3 defekta, f) minst 2 defekta.

Vad blir g och m?

Lösning

Ordningen spelar ingen roll, så $m = \binom{15}{5} = 3003$ i alla deluppgifter. Att få exakt k defekta ger $g = \binom{5}{k}\binom{10}{5-k}$, enligt avsnitt 3. Figuren visar alla sex sannolikheterna; d) och e) delar upp dem i två komplementära händelser.



a) $P = \dfrac{\binom{5}{0}\binom{10}{5}}{\binom{15}{5}} = \dfrac{252}{3003} = \boxed{\dfrac{12}{143}} \approx 0{,}084$

b) $P = \dfrac{\binom{5}{1}\binom{10}{4}}{\binom{15}{5}} = \dfrac{5 \cdot 210}{3003} = \boxed{\dfrac{50}{143}} \approx 0{,}350$

c) $P = \dfrac{\binom{5}{2}\binom{10}{3}}{\binom{15}{5}} = \dfrac{10 \cdot 120}{3003} = \boxed{\dfrac{400}{1001}} \approx 0{,}400$

d) “Högst 2” är 0, 1 eller 2 defekta. De tre händelserna är disjunkta, så vi kan summera a)–c):

$$P = \frac{12}{143} + \frac{50}{143} + \frac{400}{1001} = \boxed{\frac{834}{1001}} \approx 0{,}833$$

e) “Minst 3” är 3, 4 eller 5 defekta:

$$P = \frac{\binom{5}{3}\binom{10}{2} + \binom{5}{4}\binom{10}{1} + \binom{5}{5}\binom{10}{0}}{\binom{15}{5}} = \frac{450 + 50 + 1}{3003} = \boxed{\frac{167}{1001}} \approx 0{,}167$$

f) “Minst 2” är komplementet till “högst 1”, alltså

$$P = 1 - \left(\frac{12}{143} + \frac{50}{143}\right) = \boxed{\frac{81}{143}} \approx 0{,}566$$

Rimlighetskoll. “Högst 2” och “minst 3” är komplement: $\frac{834}{1001} + \frac{167}{1001} = 1$.

På räknaren. d) i ett steg med summamallen: `Σ(nCr(5,k)*nCr(10,5-k),k,0,2)/nCr(15,5)`.

Svar i boken (2:a uppl., 2.2): a) 0,084, b) 0,350, c) 0,400, d) 0,833, e) 0,167, f) 0,566. Stämmer också med det handskrivna lösningsförslaget.

Lektion 2, övning 1 — Röda och svarta kulor

I en urna ligger 25 kulor, 10 röda och 15 svarta. Utan återläggning och på måfå väljs 5 kulor. Vad är sannolikheten att få

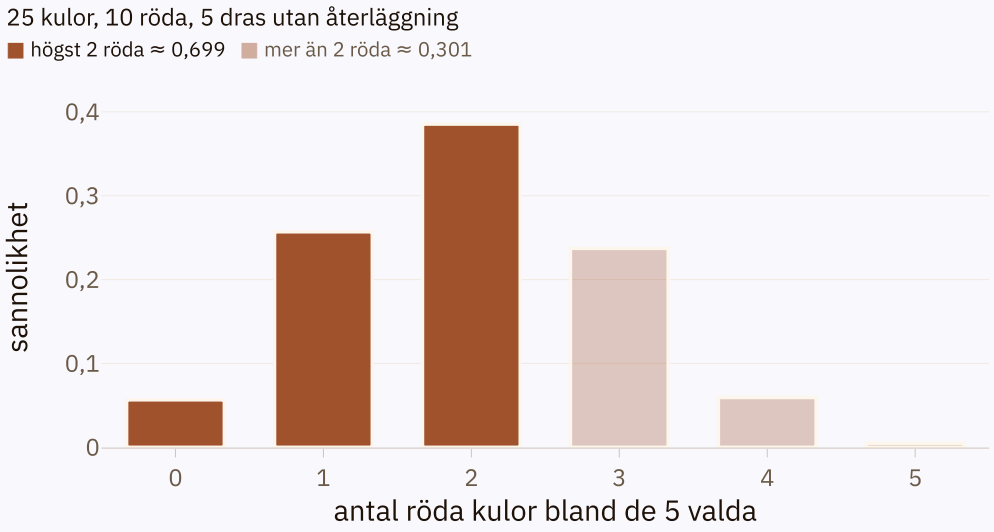
a) exakt 2 röda kulor,

b) högst 2 röda kulor,

c) mer än 2 röda kulor i urvalet?

Lösning

Ordningen spelar ingen roll, så $m = \binom{25}{5} = 53\,130$. Exakt k röda ger $g = \binom{10}{k}\binom{15}{5-k}$.



a)

$$P(\text{exakt } 2) = \frac{\binom{10}{2}\binom{15}{3}}{\binom{25}{5}} = \frac{45 \cdot 455}{53\,130} = \frac{20\,475}{53\,130} \approx \boxed{0{,}385}$$

b) 0, 1 eller 2 röda är disjunkta händelser, så sannolikheterna adderas:

$$P(\text{högst } 2) = \frac{\binom{10}{0}\binom{15}{5} + \binom{10}{1}\binom{15}{4} + \binom{10}{2}\binom{15}{3}}{\binom{25}{5}} = \frac{3003 + 13\,650 + 20\,475}{53\,130} = \frac{37\,128}{53\,130} \approx \boxed{0{,}699}$$

c) “Mer än 2” är komplementet till “högst 2”:

$$P(\text{mer än } 2) = 1 - 0{,}699 = \boxed{0{,}301}$$

På räknaren. `Σ(nCr(10,k)*nCr(15,5-k),k,0,2)/nCr(25,5)` ger b) direkt.

Facit: a) 0,385, b) 0,699, c) 0,301.

Se även

- Utfallsrum och händelser
- Sannolikhetsmått
- Multiplikationsprincipen
- Binomialkoefficient
- Hypergeometrisk fördelning

Resurser

- Wikipedia: Sannolikhet

Multiplikationsprincipen

19 okt. 2026, 4 min läsning

UTKAST

#matematik #kombinatorik #princip

> **Kurs:** S0008M **Förkunskaper:** Klassiska sannolikhetsdefinitionen

Görs flera operationer i tur och ordning multipliceras antalet sätt att göra var och en. Principen räknar ordnade val och är grunden för all kombinatorik i kursen.

1. Principen

Definition 2.2. Ska vi göra k operationer i tur och ordning, där den första kan göras på n_1 sätt, den andra på n_2 sätt och så vidare, så är det totala antalet sätt

$$\boxed{n_1 \cdot n_2 \cdots n_k}$$

Väljer vi två dosor ur en förpackning med fem kan den första väljas på 5 sätt och den andra på 4, så det finns $5 \cdot 4 = 20$ möjliga utfall.

2. Träddiagram

Principen syns tydligast i ett **träddiagram**: varje nivå är en operation, och antalet grenar multipliceras nivå för nivå. En planka som bedöms på tre stationer med 3, 2 och 2 alternativ ger ett träd med $3 \cdot 2 \cdot 2 = 12$ löv.

Trädet är också användbart för att *lista* utfallen och räkna de gynnsamma direkt, vilket ofta är säkrare än att räkna med formler när antalet utfall är litet.

3. Med hänsyn till ordning

Principen räknar ordnade val

Att först dra dosa 1 och sedan dosa 2 är ett annat utfall än att först dra dosa 2 och sedan dosa 1. Det är rätt när frågan gäller en ordning (“först en fungerande och sedan en trasig”), men dubbelräknar när ordningen inte spelar någon roll. Då delar vi bort de inbördes ordningarna, och det leder till binomialkoefficienten.

Att välja k element bland n kan göras på fyra sätt, beroende på två frågor: läggs det valda tillbaka, och spelar ordningen roll? Tre av dem förekommer i kursen:

	Ordningen spelar roll	**Ordningen spelar ingen roll**
Med återläggning	n^k	*(förekommer inte i kursen)*
Utan återläggning	$n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$	$\binom{n}{k}$

Båda talen i vänsterkolumnen är multiplikationsprincipen direkt: med återläggning har varje val n alternativ, utan återläggning ett färre för varje val. Specialfallet $k = n$ utan återläggning ger $n!$, antalet sätt att ordna n saker.

Exempel — Kod till ett cykellås

Ett cykellås har fyra ringar med siffrorna 0–9.

a) Hur många koder finns?

b) Hur många koder har fyra olika siffror?

c) Vi väljer en kod på måfå. Vad är sannolikheten att alla fyra siffror är olika?

Lösning

a) Varje ring har 10 lägen oberoende av de andra, och ordningen spelar roll (1234 är inte 4321). Det är ordnat med återläggning:

$$10^4 = \boxed{10\,000}$$

b) Nu får en siffra som redan använts inte väljas igen: ordnat utan återläggning.

$$10 \cdot 9 \cdot 8 \cdot 7 = \boxed{5040}$$

c) Alla $10\,000$ koder är lika sannolika, så klassiska sannolikhetsdefinitionen ger

$$P = \frac{5040}{10\,000} = \boxed{0{,}504}$$

Nästan varannan kod har alltså en upprepad siffra — fler än man kanske gissar.

På räknaren. `nPr(10,4)` ger `5040`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

⚠ **Vanliga fel**

- **Blanda ihop kolumnerna.** Fråga dig om $\{A, B\}$ och $\{B, A\}$ är *samma* utfall. Är de det ska ordningen bort, annars inte.
- **Räkna m och g på olika sätt.** Räknas de möjliga utfallen som ordnade måste de gynnsamma också räknas som ordnade, se Klassiska sannolikhetsdefinitionen.
- **Glömma att ett val påverkar nästa.** Utan återläggning minskar antalet alternativ för varje val, $n(n-1)\cdots$, inte $n \cdot n \cdots$.

---

4. Lösta lektionsuppgifter

Lektion 1, övning 2 — Tre schaktmaskiner

En maskintyp har $50\ \%$ chans att fungera i minst sex månader utan att några delar behöver bytas. Vi köper tre maskiner. Vad är sannolikheten att

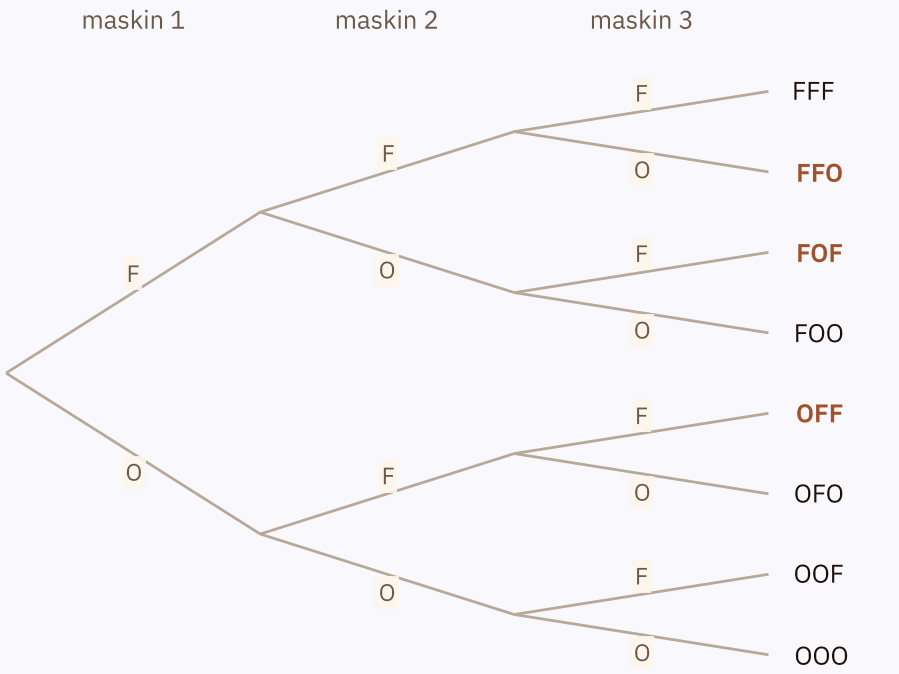
a) alla tre,

b) exakt två

har fungerat utan delbyte efter sex månader?

Lösning

Låt F = maskinen fungerar efter sex månader och O = den fungerar inte. Varje maskin ger F eller O, så enligt multiplikationsprincipen finns $m = 2 \cdot 2 \cdot 2 = 8$ utfall. Eftersom F och O är lika troliga är alla åtta lika sannolika.



markerat: exakt en trasig (O)

a) Bara utfallet FFF ger att alla tre fungerar: $g = 1$.

$$P(\text{alla tre}) = \frac{1}{8} = \boxed{0{,}125}$$

b) Utfallen med exakt två F är FFO, FOF och OFF (markerade i trädet): $g = 3$.

$$P(\text{exakt två}) = \frac{3}{8} = \boxed{0{,}375}$$

Facit: a) 0,125, b) 0,375.

Se även

- Klassiska sannolikhetsdefinitionen
- Binomialkoefficient
- Utfallsrum och händelser

Resurser

- Khan Academy: Counting, permutations, and combinations
- Wikipedia: Rule of product

Binomialkoefficient

19 okt. 2026, 4 min läsning

UTKAST

#matematik #kombinatorik #formel

> **Kurs:** S0008M **Förkunskaper:** Multiplikationsprincipen

---

Binomialkoefficienten $\binom{n}{k}$ är antalet sätt att välja k element bland n utan återläggning och utan hänsyn till ordningen. Den följer ur multiplikationsprincipen genom att vi delar bort de inbördes ordningarna.

---

1. Sats 2.1

Antalet sätt att välja ut k element bland n **utan återläggning** och **utan hänsyn till ordningen** bland de utvalda är

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{\overbrace{n(n-1)\cdots(n-k+1)}^{k \text{ faktorer}}}{\underbrace{k(k-1)\cdots 2\cdot 1}_{k \text{ faktorer}}}$$

med $k! = k(k-1)\cdots 2\cdot 1$ och konventionerna $0! = 1$ och $\binom{n}{0} = 1$. Beteckningen $\binom{n}{k}$ läses ”n över k” och är bara ett kortare sätt att skriva bråket.

---

2. Varför

Anta att det går att välja k element bland n utan hänsyn till ordning på x sätt. Vi räknar de **ordnade** valen på två olika sätt:

- Direkt med multiplikationsprincipen: $n(n-1)\cdots(n-k+1)$ sätt.
- Först välja de k elementen på x sätt och sedan ordna dem inbördes på $k!$ sätt: $x \cdot k!$.

Båda räknar samma sak, så

$$n(n-1)\cdots(n-k+1) = x\cdot k! \quad\Longrightarrow\quad x = \frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{k}$$

---

3. Räkna för hand

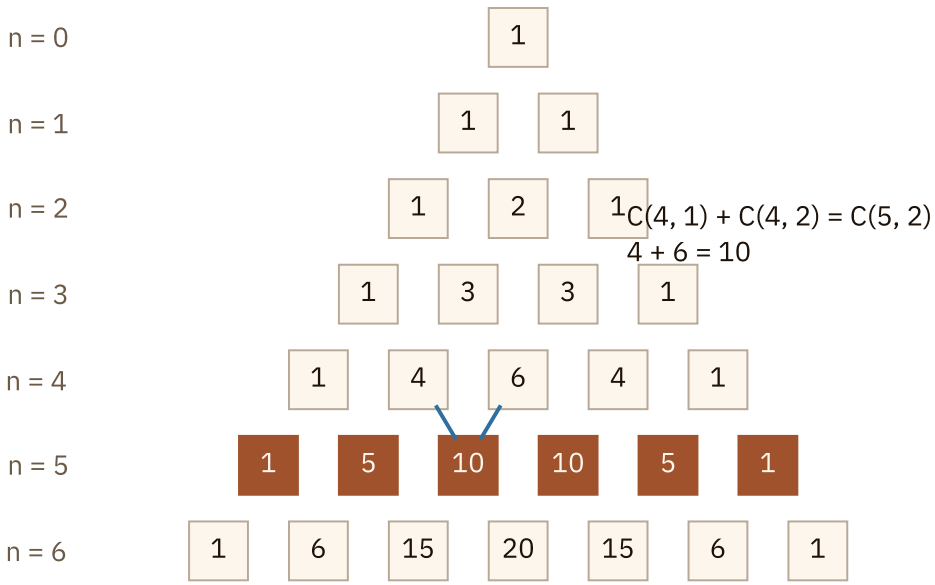
Bråket i mitten är det som räknas för hand

$\binom{20}{5}$ blir $\dfrac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$ — fem faktorer uppe, fem nere, och det mesta förkortas bort innan vi multiplicerar. Fakultetsformen är bra för bevis, men ingen räknar ut $20!$.

Två egenskaper sparar räknearbete:

$$\binom{n}{k} = \binom{n}{n-k}, \qquad \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Den första säger att välja ut k är samma sak som att välja bort $n-k$, så $\binom{35}{33} = \binom{35}{2} = 595$. Den andra säger att ett visst element antingen är med (då ska $k-1$ till väljas bland de andra $n-1$) eller inte (då ska alla k väljas bland de $n-1$). Den bygger upp **Pascals triangel**, där varje tal är summan av de två ovanför:



Rad n innehåller $\binom{n}{0}, \ldots, \binom{n}{n}$ och summerar till 2^n, antalet delmängder av en mängd med n element.

På räknaren heter funktionen `nCr` (`menu` → Probability → Combinations): `nCr(20,5)` ger $15\,504$.

4. Lösta lektionsuppgifter

Lektion 1, övning 4 — Lottorader

På hur många sätt kan vi välja 7 nummer bland 35 utan återläggning och utan hänsyn till ordningen bland de utvalda? Resonera utifrån multiplikationsprincipen: hur blir det om vi *tar* hänsyn till ordningen?

Lösning

Med hänsyn till ordning. Första numret kan väljas på 35 sätt, andra på 34 och så vidare: $35 \cdot 34 \cdot 33 \cdot 32 \cdot 31 \cdot 30 \cdot 29$ ordnade rader.

Utan hänsyn till ordning. Varje uppsättning av 7 nummer har räknats $7!$ gånger ovan, en gång per inbördes ordning. Dela bort det:

$$\binom{35}{7} = \frac{35 \cdot 34 \cdot 33 \cdot 32 \cdot 31 \cdot 30 \cdot 29}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \boxed{6\,724\,520}$$

Samma resonemang för k nummer bland n ger Sats 2.1, se avsnitt 1 och 2 ovan.

På räknaren. `nCr(35,7)`.

Facit: 6 724 520.

Lektion 1, övning 5 — Från lottoraden till Sats 2.1

Generalisera resonemanget i övning 4 så att det gäller för urval av k nummer bland n. Vad är viktigt i Sats 2.1? Varför definierar vi $\binom{n}{0} = 1$?

Lösning

Generalisering. Med hänsyn till ordning kan k nummer bland n väljas på $n(n-1)\cdots(n-k+1)$ sätt — k faktorer, en per val, och varje val har ett alternativ färre än det förra eftersom vi inte lägger tillbaka. Varje oordnad uppsättning förekommer där $k!$ gånger, en gång per inbördes ordning. Alltså är antalet oordnade urval

$$\frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{k}$$

vilket är Sats 2.1 (avsnitt 2 ovan).

Det viktiga i satsen är förutsättningarna: urvalet görs **utan återläggning** (annars blir det inte $n-1$ alternativ i andra valet) och **utan hänsyn till ordningen** (annars ska vi inte dela med $k!$).

Varför $\binom{n}{0} = 1$. Att välja noll element kan göras på precis ett sätt — att inte välja något. Definitionen gör också att fakultetsformen fungerar utan undantag: $\frac{n!}{0!\,n!} = 1$ med $0! = 1$.

Resonemangsuppgift — det finns inget facit att jämföra med.

Se även

- Multiplikationsprincipen
- Klassiska sannolikhetsdefinitionen
- Hypergeometrisk fördelning
- Binomialfördelning

Resurser

- Khan Academy: Counting, permutations, and combinations
- Wikipedia: Binomialkoefficient

Betingad sannolikhet

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Sannolikhetsmått

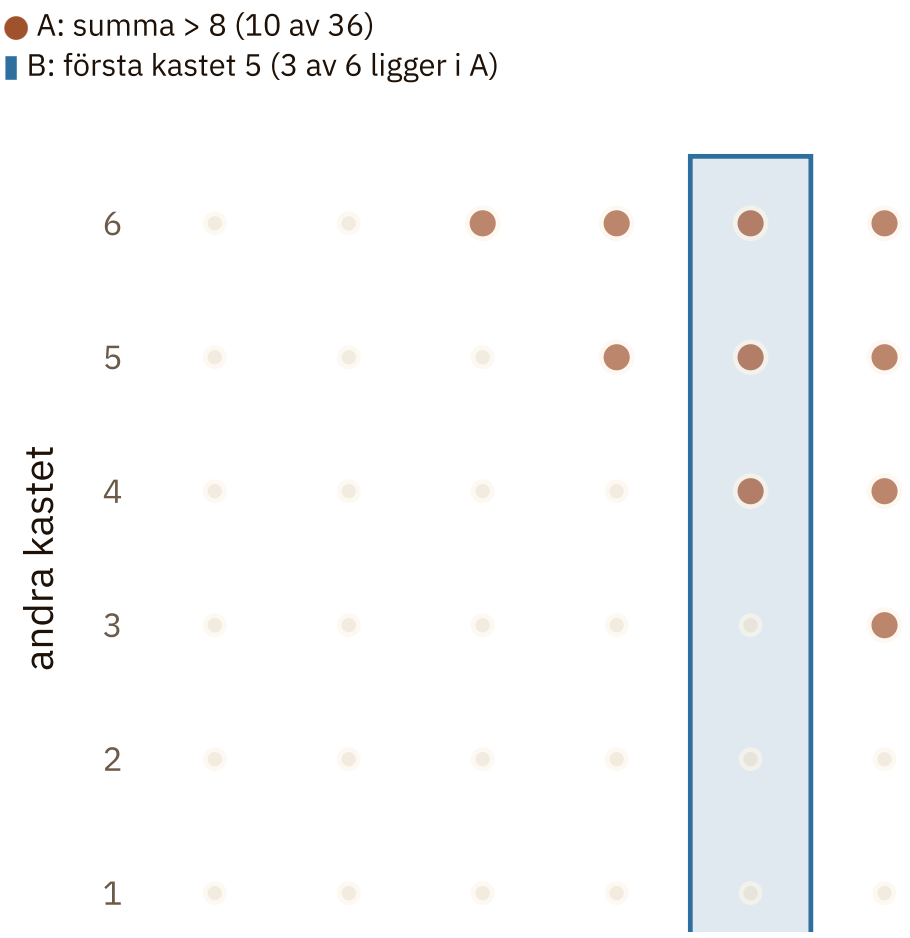
Betingad sannolikhet är sannolikheten för en händelse givet att en annan redan har inträffat, $P(A \mid B) = P(A \cap B)/P(B)$. Betingningen innebär att vi krymper utfallsrummet till det som är förenligt med den givna informationen.

---

1. Att krympa utfallsrummet

Två personer kastar en tärning två gånger var. Låt A vara att tärningssumman blir större än 8 och B att första kastet blir 5. Av de 36 lika sannolika utfallen ger 10 händelsen A, så $P(A) = 10/36$.

Får vi veta att första kastet blev 5 är bara de 6 utfallen i B längre möjliga. Av dem ger 3 händelsen A, och sannolikheten blir $3/6$. Vi har räknat som om B vore hela utfallsrummet.



I masstolkningen blir samma sak

$$P(A \mid B) = \frac{\text{massan på den del av } A \text{ som ligger i } B}{\text{massan på } B} = \frac{P(A \cap B)}{P(B)}$$

---

2. Definitionen

Definition 2.4. Med den **betingade sannolikheten** $P(A \mid B)$ för händelsen A givet händelsen B menas

$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}}$$

Förutsättning

Definitionen kräver $P(B) > 0$. Det går inte att betinga på något som aldrig inträffar.

$P(\cdot \mid B)$ uppfyller själv alla egenskaper hos ett sannolikhetsmått. Komplementregeln och additionssatsen gäller därför också betingat, till exempel $P(A^c \mid B) = 1 - P(A \mid B)$.

> ⚠ **$P(A \mid B)$ är inte $P(B \mid A)$**
>
> Täljaren är densamma, $P(A \cap B)$, men nämnarna skiljer sig. Sannolikheten att en skadad person är man är något helt annat än sannolikheten att en man skadas.

---

3. Multiplikationsformeln

Definitionen skriven baklänges ger snittets sannolikhet:

$$\boxed{P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A)}$$

Den används när det är lätt att tänka sig försöket i två steg. Väljer vi två dosor ur en förpackning med 5, varav 3 har trasig knapp, är sannolikheten att den första fungerar $2/5$, och givet det är sannolikheten att även den andra fungerar $1/4$:

$$P(\text{båda fungerar}) = \frac{2}{5} \cdot \frac{1}{4} = 0{,}1$$

Det är precis vad ett träddiagram gör: sannolikheterna längs en gren multipliceras, och varje sannolikhet efter första nivån är betingad på grenen fram till dit.

Med flera steg upprepas formeln och ger Lagen om total sannolikhet och Bayes sats.

---

4. Lösta lektionsuppgifter

> ☰ **Lektion 2, övning 4 och 5 (Vännman 2.18) — Malm från järnverket**
>
> Vid ett järnverk tas prover på den inkommande malmen. Låt
>
> A = Fe-halten är större än $66{,}5\ \%$,
>
> B = SiO_2-halten är mindre än $4{,}0\ \%$.
>
> Ur ett stort antal prover tar vi $P(A) = 0{,}32$, $P(B) = 0{,}54$ och $P(A \cap B) = 0{,}23$.

Övning 4. I ett nyligen analyserat prov vet vi att SiO_2-halten är mindre än $4\ \%$. Hur stor är då sannolikheten att Fe-halten är större än $66{,}5\ \%$? Skriv sannolikheten med formell beteckning.

Övning 5. Bestäm den betingade sannolikheten för B givet A och jämför med övning 4.

Lösning

Övning 4. Vi vet att B har inträffat och söker $P(A \mid B)$:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}23}{0{,}54} \approx \boxed{0{,}43}$$

Övning 5. Nu är det A som är given:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0{,}23}{0{,}32} \approx \boxed{0{,}72}$$

Jämförelse. Samma täljare, olika nämnare. Eftersom A är ovanligare än B blir $P(B \mid A)$ större: bland prover med hög Fe-halt har de flesta låg SiO_2-halt, men bland prover med låg SiO_2-halt är det färre än hälften som har hög Fe-halt. Båda är större än de obetingade $P(A) = 0{,}32$ och $P(B) = 0{,}54$, så de två egenskaperna tenderar att förekomma tillsammans.

Facit (övning 5): 0,72. Svar i boken (2:a uppl., 2.18): 0,43.

Lektion 2, övning 9 — Tomas två barn

Du stöter på en gammal bekant, Tomas, som berättar att han har två barn. När ni skiljs åt säger han: “Nu ska jag gå och hämta Anna på dagis.” Vad är sannolikheten att Tomas andra barn är en pojke?

Lösning

Utfallsrum. De två barnens kön: $\Omega = \{FF,\ FP,\ PF,\ PP\}$, där vi antar att vart och ett av de fyra utfallen har sannolikhet $1/4$.

Händelser. Vi vet att minst ett barn är en flicka: $A = \{FF,\ FP,\ PF\}$. Vi söker sannolikheten att det finns en pojke, alltså ett barn av vardera kön: $B = \{FP,\ PF\}$.



äldsta barnet först: F = flicka, P = pojke

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{2/4}{3/4} = \boxed{\frac{2}{3}}$$

Samma sak direkt ur utfallsrummet: av de tre utfall som är förenliga med informationen innehåller två en pojke.

Facit: 2/3. Stämmer med det handskrivna lösningsförslaget.

Lektion 2, övning 10 — Bil-get-problemet

I ett TV-program får du välja en av tre dörrar. Bakom två står en get och bakom den tredje en bil, som blir din om du väljer rätt dörr. Du pekar på en dörr. Programledaren öppnar en av de två andra, och där står en get. Du erbjuds att byta dörr. Ska du byta?

Lösning

Byt dörr.

Den som håller fast vid sitt första val vinner bara om den valde bildörren från början, vilket har sannolikheten $1/3$. Programledarens dörröppning ändrar inte på det.

Den som byter vinner om den från början valde en getdörr, vilket har sannolikheten $2/3$. Programledaren måste nämligen öppna den *andra* getdörren, så den kvarvarande stängda dörren döljer då bilen.

$$P(\text{vinst} \mid \text{byter}) = \boxed{\frac{2}{3}}, \qquad P(\text{vinst} \mid \text{byter inte}) = \frac{1}{3}$$

Facit: byta dörr, med samma motivering.

Se även

- Sannolikhetsmått
- Lagen om total sannolikhet
- Bayes sats
- Oberoende händelser
- Multiplikationsprincipen

Resurser

- 3Blue1Brown: Bayes theorem, the geometry of changing beliefs — betingning som att krympa området man tittar på
- Khan Academy: Conditional probability
- Wikipedia: Betingad sannolikhet

Lagen om total sannolikhet

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Betingad sannolikhet

Delar vi upp utfallsrummet i disjunkta fall kan sannolikheten för en händelse räknas fall för fall och sedan summeras. Varje term är fallets sannolikhet gånger den betingade sannolikheten för händelsen i just det fallet.

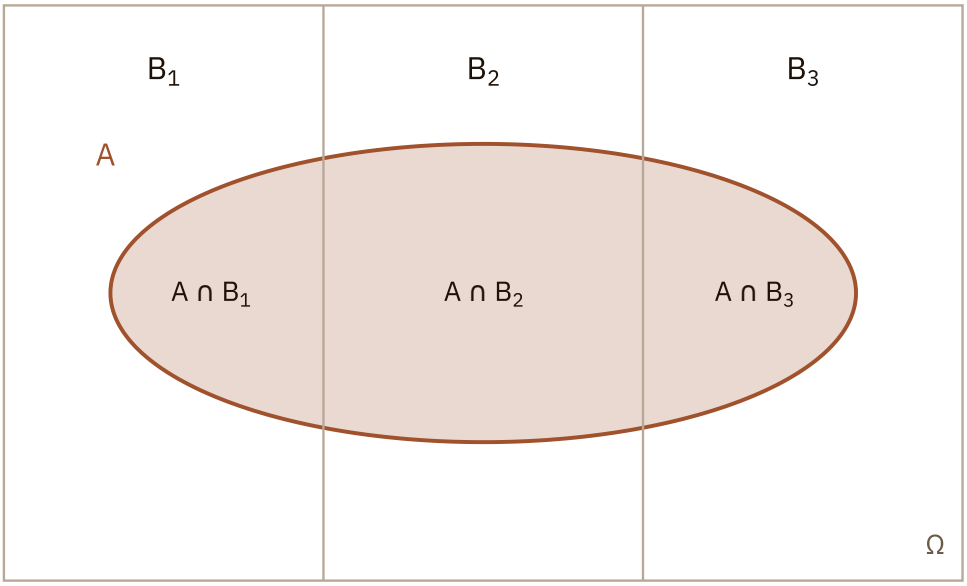
1. Satsen

Sats 2.7. Antag att händelserna $B_1, B_2, \dots, B_n$ är disjunkta och tillsammans utgör hela utfallsrummet. Då gäller för varje händelse A att

$$\boxed{P(A) = P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2) + \cdots + P(B_n)\,P(A \mid B_n)}$$

2. Varför

$B_1, \dots, B_n$ skär A i disjunkta bitar $A \cap B_1,\ \dots,\ A \cap B_n$ som tillsammans är hela A:



Disjunkta bitar adderas enligt sannolikhetsmåttets egenskap c), och varje bit skrivs om med multiplikationsformeln $P(A \cap B_i) = P(B_i)\,P(A \mid B_i)$:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)\,P(A \mid B_i)$$

3. I ett träddiagram

Satsen är exakt det vi gör i ett träddiagram: första nivån är fallen B_i, andra nivån är de betingade sannolikheterna för A i varje fall. Vi multiplicerar längs varje gren som slutar i A och adderar grenarna.

När satsen behövs

Signalen i en uppgift är att sannolikheten för A är given *per grupp* — per fabrik, per transportsätt, per skickat tecken — tillsammans med hur stor varje grupp är. Då är fallen B_i grupperna.

Frågar uppgiften därefter “givet att A inträffade, vilken grupp kom det från?” är nästa steg Bayes sats.

Vanliga fel

- **Fallen täcker inte allt eller överlappar.** $B_1, \ldots, B_n$ måste vara disjunkta och tillsammans vara hela Ω. Kontroll: $P(B_1) + \cdots + P(B_n) = 1$.
- **Summera de betingade sannolikheterna utan vikter.** $P(A \mid B_1) + P(A \mid B_2)$ är ingen sannolikhet. Varje term ska vägas med $P(B_i)$.

4. Lösta lektionsuppgifter

Lektion 2, övning 6a (Vännman 2.25) — Batterier från tre fabriker

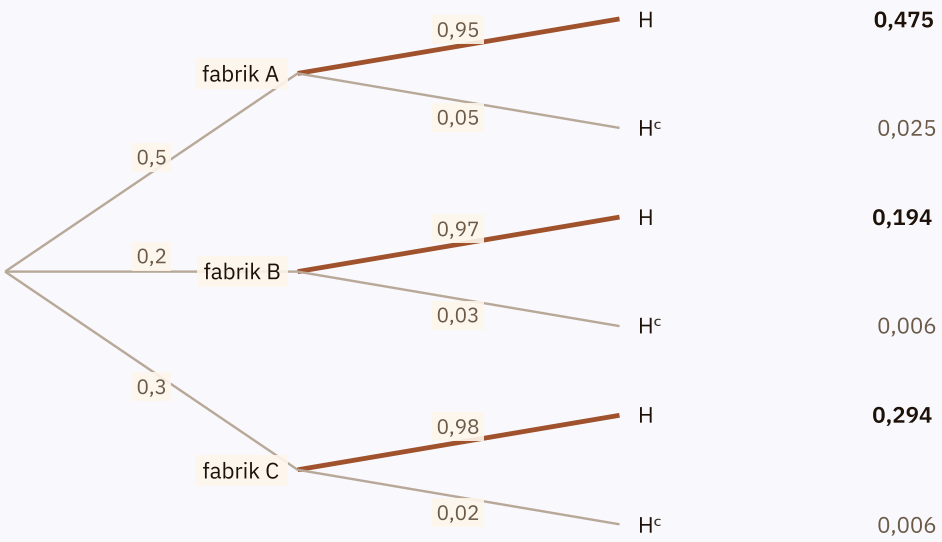
Tillverkningen av ett batteri är förlagd till tre fabriker: fabrik A står för $50\ \%$, B för $20\ \%$ och C för $30\ \%$. Ett batteri från fabrik A har $95\ \%$ sannolikhet att räcka mer än 100 drifttimmar. Motsvarande för B och C är $97\ \%$ respektive $98\ \%$. Batterierna blandas i ett stort centrallager. Vad är sannolikheten att ett batteri som tas på måfå ur lagret räcker mer än 100 drifttimmar?

Lösning

Beteckningar. A, B, C = batteriet kommer från respektive fabrik; H = batteriet räcker mer än 100 timmar. Fabrikerna är disjunkta och täcker hela lagret.

$$P(A) = 0{,}50, \quad P(H \mid A) = 0{,}95$$
$$P(B) = 0{,}20, \quad P(H \mid B) = 0{,}97$$
$$P(C) = 0{,}30, \quad P(H \mid C) = 0{,}98$$

Lagen om total sannolikhet — i trädet är det summan av de grenar som slutar i H:



P(H) = summan av de markerade grenarna = 0,963

$$P(H) = 0{,}50 \cdot 0{,}95 + 0{,}20 \cdot 0{,}97 + 0{,}30 \cdot 0{,}98 = 0{,}475 + 0{,}194 + 0{,}294 = \boxed{0{,}963}$$

Rimlighetskoll. Svaret är ett viktat medelvärde av $0{,}95$, $0{,}97$ och $0{,}98$ och måste därför ligga mellan dem. Det gör det, och närmast $0{,}95$ eftersom fabrik A väger tyngst.

Fortsättningen, b), finns i Bayes sats.

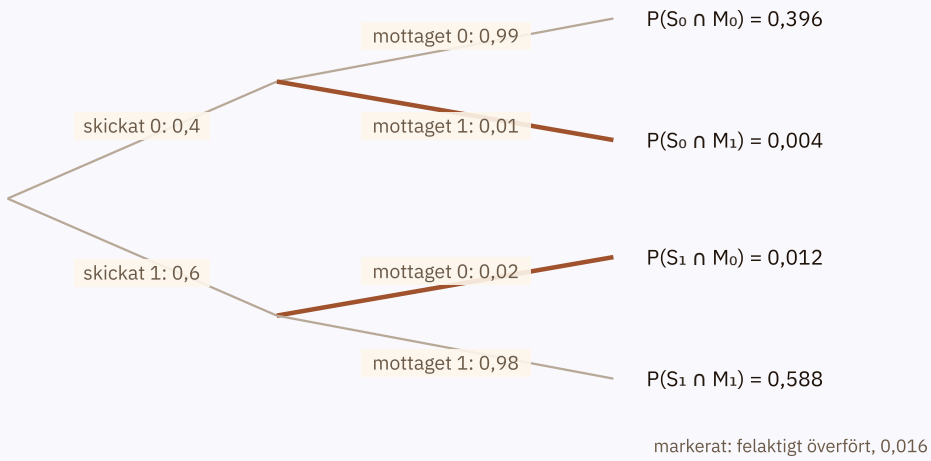
Svar i boken (2:a uppl., 2.24 = 3:e uppl. 2.25): 0,96.

Lektion 2, övning 8b (Vännman 2.27) — Felaktigt överförda tecken

Meddelanden kodade i binära tecken 0 och 1 överförs i ett telekommunikationssystem. Ett utsänt tecken 0 mottas som 1 med sannolikhet $0{,}01$ (och som 0 med sannolikhet $0{,}99$). Ett utsänt tecken 1 mottas som 0 med sannolikhet $0{,}02$ (och som 1 med sannolikhet $0{,}98$). Tecknet 1 förekommer i proportion $0{,}6$ och tecknet 0 i proportion $0{,}4$. Hur stor andel av tecknen överförs felaktigt?

Lösning

Beteckningar. S_0, S_1 = tecknet 0 respektive 1 skickas; M_0, M_1 = tecknet 0 respektive 1 mottas.



Ett fel är att skicka 0 och ta emot 1, eller att skicka 1 och ta emot 0. De två grenarna är disjunkta:

$$P(\text{fel}) = P(S_0)\,P(M_1 \mid S_0) + P(S_1)\,P(M_0 \mid S_1) = 0{,}4 \cdot 0{,}01 + 0{,}6 \cdot 0{,}02 = 0{,}004 + 0{,}0$$

Deluppgift a) finns i Bayes sats.

Svar i boken (2:a uppl., 2.26 = 3:e uppl. 2.27): 0,016. Stämmer också med det handskrivna lösningsförslaget.

Se även

- Betingad sannolikhet
- Bayes sats
- Sannolikhetsmått
- Multiplikationsprincipen

Resurser

- Khan Academy: Probability
- Wikipedia: Law of total probability

Bayes sats

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Betingad sannolikhet, Lagen om total sannolikhet

---

Bayes sats vänder på en betingning: vet vi hur sannolik händelsen A är i varje fall B_i, kan vi räkna ut hur sannolikt varje fall är givet att A har inträffat. Nämnaren är lagen om total sannolikhet.

---

1. Satsen

Låt $B_1, \dots, B_n$ vara disjunkta och tillsammans utgöra hela utfallsrummet, och låt $P(A) > 0$. Då är

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)\,P(A \mid B_j)}$$

---

2. Varför

Det är ingen ny idé, bara två kända formler efter varandra. Definitionen ger

$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)}$$

Täljaren skrivs om med multiplikationsformeln, $P(B_i \cap A) = P(B_i)\,P(A \mid B_i)$, och nämnaren med lagen om total sannolikhet.

I träddiagrammet

Täljaren är **en** gren som slutar i A. Nämnaren är **alla** grenar som slutar i A. $P(B_i \mid A)$ är alltså den andel av A-massan som kommer via gren i.

---

3. Det vi har och det vi söker

Uppgifter ger nästan alltid $P(A \mid B_i)$ — hur ofta en fabrik tillverkar defekta enheter, hur ofta ett tecken störs — men frågar efter $P(B_i \mid A)$: givet en defekt enhet, vilken fabrik kom den från?

Basfrekvensen avgör

Även om $P(A \mid B_i)$ är liten kan $P(B_i \mid A)$ bli stor om B_i är vanlig, och tvärtom. Att glömma faktorn $P(B_i)$ i täljaren är det vanligaste felet.

Exempel — Ett positivt test för en ovanlig sjukdom

En sjukdom finns hos $1\ \%$ av befolkningen. Ett test ger positivt svar hos $95\ \%$ av de sjuka och hos $10\ \%$ av de friska. En slumpmässigt vald person testar positivt. Vad är sannolikheten att personen är sjuk?

Lösning

Beteckningar. S = personen är sjuk, $+$ = testet är positivt. Givet: $P(S) = 0{,}01$, $P(+ \mid S) = 0{,}95$, $P(+ \mid S^c) = 0{,}10$.

Bayes sats med fallen S och S^c:

$$P(S \mid +) = \frac{0{,}01 \cdot 0{,}95}{0{,}01 \cdot 0{,}95 + 0{,}99 \cdot 0{,}10} = \frac{0{,}0095}{0{,}1085} \approx \boxed{0{,}088}$$

Tolkning med naturliga frekvenser. Tänk dig $10\,000$ personer: 100 är sjuka, och 95 av dem testar positivt. Av de 9900 friska testar 990 positivt. Bland de 1085 positiva är alltså bara 95 sjuka:



Testet är bra, men sjukdomen är så ovanlig att de falska positiva från den stora friska gruppen blir tio gånger fler än de sanna positiva.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

---

4. Lösta lektionsuppgifter

Lektion 2, övning 6b (Vännman 2.25) — Vilken fabrik?

Fortsättning från Lagen om total sannolikhet: fabrik A, B och C står för $50\ \%$, $20\ \%$ och $30\ \%$ av batterierna, och andelen som räcker mer än 100 drifttimmar är $95\ \%$, $97\ \%$ och $98\ \%$. Vi tar på måfå ett batteri ur lagret och finner att det räcker mer än 100 drifttimmar. Vad är sannolikheten att det är tillverkat i fabrik A?

Lösning

Med H = batteriet räcker mer än 100 timmar gav lagen om total sannolikhet $P(H) = 0{,}963$. Bayes sats:

$$P(A \mid H) = \frac{P(A)\,P(H \mid A)}{P(H)} = \frac{0{,}50 \cdot 0{,}95}{0{,}963} = \frac{0{,}475}{0{,}963} \approx \boxed{0{,}49}$$

Rimlighetskoll. Fabrik A gör hälften av batterierna men har den lägsta hållbarheten, så bland de hållbara batterierna bör andelen från A vara lite under $0{,}50$. Det är den.

Svar i boken (2:a uppl., 2.24 = 3:e uppl. 2.25): 0,49.

Lektion 2, övning 7 (Vännman 2.23) — Skador på en arbetsplats

På en arbetsplats skadades $1\ \%$ av personalen under ett år. $60\ \%$ av alla skadade var män. $30\ \%$ av de anställda var kvinnor. Är det de manliga eller de kvinnliga anställda som löper störst risk att råka ut för en skada enligt denna undersökning?

Lösning

Beteckningar. S = personen skadas; M = personen är man; $K = M^c$ = personen är kvinna.

Givet. $P(S) = 0{,}01$, $P(M \mid S) = 0{,}60$, $P(K) = 0{,}30$, alltså $P(M) = 0{,}70$ och $P(K \mid S) = 0{,}40$.

Vi söker risken per kön, $P(S \mid M)$ och $P(S \mid K)$ — betingningen åt andra hållet mot det som är givet. Bayes sats, här med nämnaren $P(M)$ given direkt:

$$P(S \mid M) = \frac{P(S)\,P(M \mid S)}{P(M)} = \frac{0{,}01 \cdot 0{,}60}{0{,}70} \approx 0{,}0086$$

$$P(S \mid K) = \frac{P(S)\,P(K \mid S)}{P(K)} = \frac{0{,}01 \cdot 0{,}40}{0{,}30} \approx 0{,}0133$$

Svar. $\boxed{\text{De kvinnliga anställda}}$ löper störst risk, trots att de flesta skadade är män. Männen är fler, och därför är de också fler bland de skadade.

Svar i boken (2:a uppl., 2.22 = 3:e uppl. 2.23): kvinnor.

Lektion 2, övning 8a (Vännman 2.27) — Mottagen etta

Samma telekommunikationssystem som i Lagen om total sannolikhet: tecknet 1 skickas i proportion $0{,}6$ och 0 i proportion $0{,}4$; ett utsänt 0 mottas som 1 med sannolikhet $0{,}01$ och ett utsänt 1 mottas som 0 med sannolikhet $0{,}02$. Om 1 har mottagits, vad är den betingade sannolikheten att 1 har sänts?

Lösning

Med S_0, S_1 = skickat 0 respektive 1 och M_1 = mottaget 1 ger lagen om total sannolikhet nämnaren:

$$P(M_1) = P(S_0)\,P(M_1 \mid S_0) + P(S_1)\,P(M_1 \mid S_1) = 0{,}4 \cdot 0{,}01 + 0{,}6 \cdot 0{,}98 = 0{,}004 + 0{,}5$$

Bayes sats:

$$P(S_1 \mid M_1) = \frac{P(S_1)\,P(M_1 \mid S_1)}{P(M_1)} = \frac{0{,}588}{0{,}592} \approx \boxed{0{,}993}$$

I trädiagrammet i Lagen om total sannolikhet är det grenen $S_1 \cap M_1$ delad med summan av de två grenar som slutar i M_1.

Svar i boken (2:a uppl., 2.26 = 3:e uppl. 2.27): 0,993. Stämmer också med det handskrivna lösningsförslaget.

Se även

- Betingad sannolikhet
- Lagen om total sannolikhet
- Sannolikhetsmått

Resurser

- 3Blue1Brown: Bayes theorem, the geometry of changing beliefs — bästa visuella förklaringen av varför basfrekvensen spelar roll
- Wikipedia: Bayes sats

Oberoende händelser

19 okt. 2026, 8 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Betingad sannolikhet

Två händelser är oberoende om vetskapen om att den ena har inträffat inte ändrar sannolikheten för den andra. Då är sannolikheten att båda inträffar produkten av de enskilda sannolikheterna.

---

1. Definitionen

Kastar vi en tärning två gånger och låter A = femma i första kastet och B = tvåa i andra är

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/36}{1/6} = \frac{1}{6} = P(B)$$

Informationen att A har inträffat ändrar inte vår kännedom om B. Multiplikationsformeln $P(A \cap B) = P(A)\,P(B \mid A)$ blir då $P(A \cap B) = P(A)\,P(B)$, och det tar vi som definition.

Definition 2.5. Om

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

säger vi att A och B är **oberoende** händelser.

> **Två sätt att säga samma sak**
>
> Om $P(B) > 0$ gäller
>
> $$P(A \mid B) = P(A) \quad \Longleftrightarrow \quad P(A \cap B) = P(A)\,P(B)$$
>
> Framåt: multiplicera $P(A \mid B) = P(A \cap B)/P(B) = P(A)$ med $P(B)$. Bakåt: dela produkten med $P(B)$. Produktformen är symmetrisk i A och B och kräver inte att någon sannolikhet är positiv, och därför används den som definition.

Är händelserna **beroende** måste vi känna till någon av de betingade sannolikheterna $P(A \mid B)$ eller $P(B \mid A)$ för att kunna räkna ut $P(A \cap B)$. Är de oberoende räcker $P(A)$ och $P(B)$.

---

2. Oberoende är inte disjunkt

⚠ Disjunkta händelser är beroende

Disjunkta händelser kan inte inträffa samtidigt, så $P(A \cap B) = 0$. Har båda positiv sannolikhet är $P(A)\,P(B) > 0$, och produktregeln bryts. Att veta att A inträffat säger tvärtom väldigt mycket om B: den inträffade *inte*.

Disjunkt är en egenskap hos mängderna och syns i ett mängddiagram. Oberoende är en egenskap hos sannolikheterna och syns inte i figuren.

---

3. Fler än två händelser

Tre händelser A, B och C är oberoende om produktregeln gäller för **alla** kombinationer:

$$
\begin{aligned}
P(A \cap B \cap C) &= P(A)\,P(B)\,P(C) \\
P(A \cap B) &= P(A)\,P(B) \\
P(A \cap C) &= P(A)\,P(C) \\
P(B \cap C) &= P(B)\,P(C)
\end{aligned}
$$

och motsvarande för fler händelser. Är $A, B, C, \ldots$ oberoende så är också komplementen $A^c, B^c, C^c, \ldots$ oberoende, liksom alla blandningar som $A, B^c, C, \ldots$

"Minst en" för oberoende händelser

Kombinera komplementregeln med produktregeln:

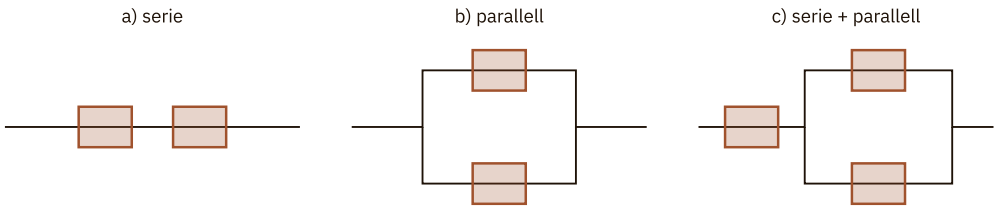
$$
P(\text{minst en av } A_1, \ldots, A_n) = 1 - P(A_1^c \cap \cdots \cap A_n^c) = 1 - P(A_1^c) \cdots P(A_n^c)
$$

Det är standardgreppet i nästan varje uppgift om oberoende händelser.

---

4. Serie- och parallellsystem

Komponenter som fungerar oberoende av varandra kan kopplas på två grundsätt:



Koppling	Fungerar om	Sannolikhet, komponenter med p_1, p_2
Serie	alla komponenter fungerar	$p_1\,p_2$
Parallell	minst en komponent fungerar	$1-(1-p_1)(1-p_2)$

Sammansatta system räknas inifrån och ut: ersätt varje parallellgren med dess sannolikhet och multiplicera sedan i serie.

---

5. Lösta lektionsuppgifter

Lektion 3, övning 1–3 — Justering av utombordsmotorer

En tillverkare av utombordsmotorer vet av erfarenhet att $40\ \%$ av motorerna behöver justering av förgasaren, $30\ \%$ behöver justering av tändningen och $21\ \%$ behöver justering av båda. En slumpmässigt vald motor visade sig behöva justering av tändningen.

Övning 1. Givet den informationen, vad är sannolikheten att motorns förgasare behöver justering? Inför och definiera lämpliga händelser.

Övning 2. Lös samma uppgift men anta att $12\ \%$ av motorerna behöver justering av både förgasare och tändning. Jämför med sannolikheten att en slumpmässigt vald motor behöver justering av förgasaren. Vilken roll spelar informationen om tändningen här, och vilken roll spelade den i övning 1?

Övning 3. Är händelserna i övning 1 oberoende? Är de i övning 2 det?

Lösning

Beteckningar. F = motorn behöver justering av förgasaren; T = motorn behöver justering av tändningen. $P(F) = 0{,}40$ och $P(T) = 0{,}30$.

Övning 1. $P(F \cap T) = 0{,}21$:

$$P(F \mid T) = \frac{P(F \cap T)}{P(T)} = \frac{0{,}21}{0{,}30} = \boxed{0{,}70}$$

Informationen höjer sannolikheten från $0{,}40$ till $0{,}70$: motorer som behöver ny tändinställning behöver oftare också förgasarjustering.

Övning 2. $P(F \cap T) = 0{,}12$:

$$P(F \mid T) = \frac{0{,}12}{0{,}30} = \boxed{0{,}40} = P(F)$$

Här spelar informationen om tändningen ingen roll alls.

Övning 3. I övning 1 är $P(F)\,P(T) = 0{,}40 \cdot 0{,}30 = 0{,}12 \neq 0{,}21 = P(F \cap T)$, så F och T är **beroende**. I övning 2 är $P(F \cap T) = 0{,}12 = P(F)\,P(T)$, så de är **oberoende** — samma slutsats som $P(F \mid T) = P(F)$.

Facit: övning 1: 0,70; övning 2: 0,40, informationen spelar ingen roll.

Lektion 3, övning 4 (Vännman 2.30) — Disjunkt och oberoende?

Två händelser A och B har sannolikheter skilda från noll. A och B är disjunkta. Kan A och B vara oberoende?

Lösning

Nej. Disjunkta händelser har $P(A \cap B) = P(\varnothing) = 0$. Eftersom båda sannolikheterna är skilda från noll är $P(A)\,P(B) > 0$. Alltså är

$$P(A \cap B) = 0 \neq P(A)\,P(B)$$

och produktregeln gäller inte. I figuren: om A har inträffat ligger utfallet utanför B, så $P(B \mid A) = 0$, vilket skiljer sig från $P(B) > 0$.

Svar i boken (2:a uppl., 2.29 = 3:e uppl. 2.30): Nej.

Lektion 3, övning 5 (Vännman 2.34) — Maskin med tre delar

En maskin består av tre delar som är felfria med sannolikheterna $0{,}9$, $0{,}85$ respektive $0{,}95$. Delarnas funktionsduglighet antas oberoende av varandra. Bestäm sannolikheten att maskinen fungerar, om den fungerar

a) endast om alla tre delarna fungerar,

b) om minst en av de tre delarna fungerar.

Lösning

Låt A_i = del i fungerar; A_1, A_2, A_3 är oberoende.

a) Seriesystem — alla måste fungera:

$$P(A_1 \cap A_2 \cap A_3) = 0{,}9 \cdot 0{,}85 \cdot 0{,}95 \approx \boxed{0{,}727}$$

b) Parallellsystem — minst en ska fungera. Via komplementet, där komplementen också är oberoende:

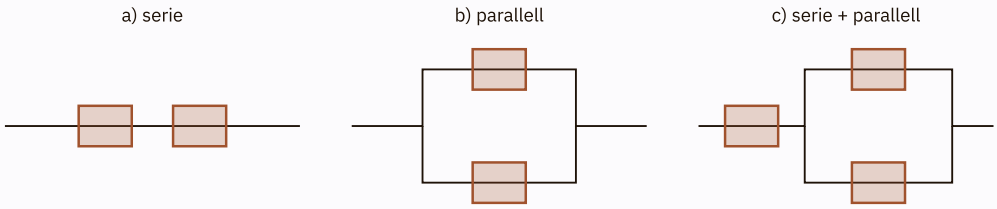
$$P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1^c)\,P(A_2^c)\,P(A_3^c) = 1 - 0{,}1 \cdot 0{,}15 \cdot 0{,}05 = 1 - 0{,}00075 \approx \boxed{0{,}}$$

Tolkning. a) motsvarar figur a) i avsnitt 4 med tre block i rad; b) motsvarar figur b) med tre block parallellt. Serie gör systemet sämre än varje enskild del, parallell gör det bättre.

Svar i boken (2:a uppl., 2.33 = 3:e uppl. 2.34): a) 0,73, b) 0,9993.

Lektion 3, övning 11 (Vännman 2.36) — Tre systemkopplingar

Varje komponent i systemen nedan fungerar med sannolikheten p, oberoende av de andra. Hur stor är sannolikheten att systemet fungerar?



Lösning

a) Två komponenter i serie:

$$P = p \cdot p = \boxed{p^2}$$

b) Två komponenter parallellt — systemet går bara sönder om båda går sönder:

$$P = 1 - (1-p)^2 = \boxed{2p - p^2}$$

c) En komponent i serie med parallellgrenen från b):

$$P = p\,(1 - (1-p)^2) = \boxed{2p^2 - p^3}$$

Rimlighetskoll. Med $p = 0{,}9$: a) $0{,}81$, b) $0{,}99$, c) $0{,}891$. Parallellkopplingen är bäst och c) ligger mellan a) och b), som väntat.

Svar i boken (2:a uppl., 2.35 = 3:e uppl. 2.36): a) p^2, b) $2p - p^2$, c) $2p^2 - p^3$.

Lektion 3, övning 8 — Minst en dubbelsexa

Vad är sannolikheten att få minst en dubbelsexa vid 24 kast med två tärningar?

a) Bestäm sannolikheten.

b) Bestäm det minsta antal kast som krävs för att sannolikheten att få minst en dubbelsexa ska bli större än $0{,}5$.

Lösning

Låt A_i = dubbelsexa i kast i. $P(A_i) = 1/36$ och kasten är oberoende.

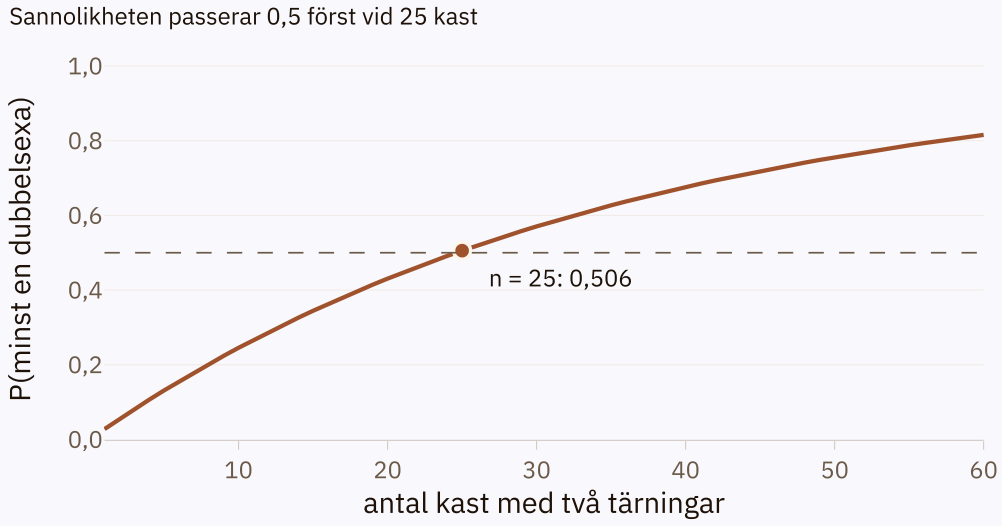
a) Via komplementet:

$$P(\text{minst en}) = 1 - P(A_1^c \cap \cdots \cap A_{24}^c) = 1 - \left(\frac{35}{36}\right)^{24} \approx \boxed{0{,}4914}$$

b) Vi söker minsta n med

$$1 - \left(\frac{35}{36}\right)^{n} > 0{,}5 \iff \left(\frac{35}{36}\right)^{n} < 0{,}5 \iff n > \frac{\ln 0{,}5}{\ln(35/36)} \approx 24{,}6$$

Olikheten vänder när vi delar med $\ln(35/36) < 0$. Alltså krävs $\boxed{n = 25}$ kast.



På räknaren. `solve(1-(35/36)^n>0.5,n)` ger $n > 24{,}6$.

Facit: a) 0,4914, b) n = 25. Stämmer med det handskrivna lösningsförslaget.

Lektion 3, övning 9 (Vännman 2.45) — Kasserade byggelement

Byggelement kan vara behäftade med fel av tre typer: typ I med sannolikheten $0{,}05$, typ II med $0{,}10$ och typ III med $0{,}03$. Ett element kan ha flera feltyper samtidigt, och feltyperna förekommer oberoende av varandra. Ett element måste kasseras om minst ett fel förekommer.

a) Bestäm sannolikheten att ett element måste kasseras.

b) En byggmästare köper 20 element. Vad är sannolikheten att han måste kassera minst 2?

Lösning

a) "Minst ett fel" via komplementet, med oberoende feltyper:

$$p = 1 - 0{,}95 \cdot 0{,}90 \cdot 0{,}97 = 1 - 0{,}82935 \approx \boxed{0{,}171}$$

b) De 20 elementen är oberoende upprepningar där varje element kasseras med sannolikheten $p = 0{,}17065$. Låt A_k = exakt k kasseras. Via komplementet "högst 1":

$$P(\text{minst } 2) = 1 - P(A_0) - P(A_1) = 1 - (1-p)^{20} - 20\,p\,(1-p)^{19} \approx \boxed{0{,}88}$$

På räknaren. `binomCdf(20,0.17065,2,20)`.

Svar i boken (2:a uppl., 2.44 = 3:e uppl. 2.45): a) 0,17, b) 0,88.

Lektion 3, övning 10 — Födelsedagsproblemet

Vad är det minsta antal studenter som behövs i en klass för att det ska vara mer än $50\ \%$ chans att minst två fyller år på samma dag? Gör lämpliga antaganden och ställ upp den olikhet som löser problemet.

Lösning

Antaganden. Inga tvillingar eller trillingar i klassen, ingen är född den 29 februari, och var och en av årets 365 dagar är en lika sannolik födelsedag, oberoende mellan studenterna.

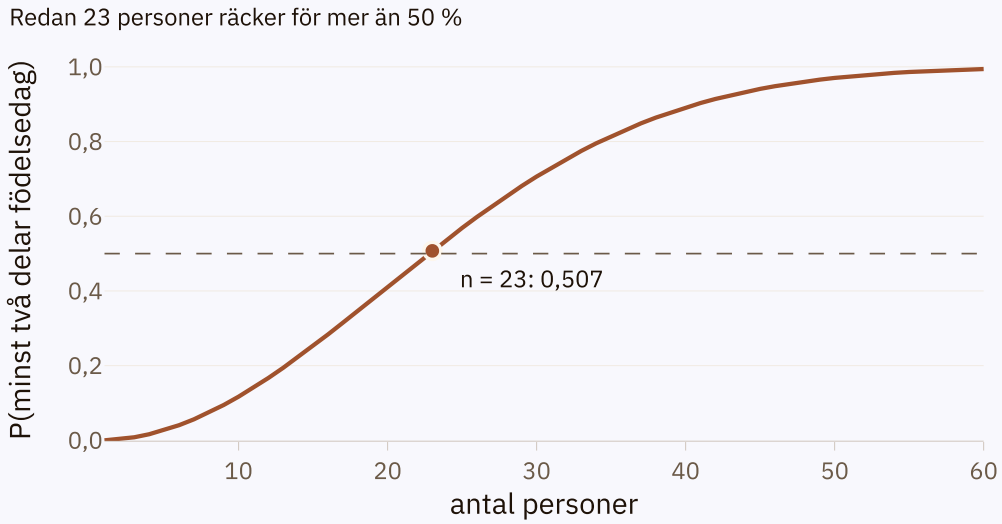
Komplementet. “Minst två delar födelsedag” är komplementet till “alla har olika födelsedagar”. Ta studenterna en i taget: den första kan ha vilken dag som helst, den andra måste undvika 1 dag, den tredje 2 dagar och så vidare. Med n studenter:

$$P(\text{alla olika}) = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{366-n}{365}$$

Olikheten. Vi söker minsta heltal n med

$$1 - 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{366-n}{365} > 0{,}5$$

Räknar vi faktor för faktor passeras $0{,}5$ först vid $n = 23$, där sannolikheten blir $0{,}507$. Svaret är $\boxed{23 \text{ studenter}}$.



På räknaren. `1-Π((365-i)/365,i,0,22)` ger $0{,}507$ och samma med `21` i stället för `22` ger $0{,}476$.

Facit: olikheten ovan med samma antaganden; lektionsbladet anger svaret 23.

Se även

- Betingad sannolikhet
- Sannolikhetsmått

- Oberoende upprepningar
- Binomialfördelning
- Oberoende stokastiska variabler

Resurser

- Khan Academy: Probability
- Wikipedia: Oberoende (sannolikhetsteori)
- Wikipedia: Birthday problem

Oberoende upprepningar

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Oberoende händelser, Binomialkoefficient

---

Upprepar vi samma delförsök n gånger oberoende av varandra, och händelsen A har samma sannolikhet p varje gång, är sannolikheten att A inträffar exakt k gånger $\binom{n}{k}p^k(1-p)^{n-k}$. Resultatet bär upp hela binomialfördelningen.

---

1. Oberoende upprepningar av ett delförsök

Ett slumpförsök kan ofta ses som en upprepning av n likadana **delförsök**, där händelserna i de olika delförsöken är oberoende. Vi säger då att försöket består av n **oberoende upprepningar av ett delförsök**. Att kasta en tärning fem gånger är ett sådant försök; delförsöket är ett kast.

---

2. Resonemanget

Kasta en tärning 5 gånger och sök sannolikheten att få exakt 3 ettor. Låt A = etta i ett kast, så $P(A) = 1/6$.

Varje kastserie med exakt 3 ettor är en följd som $AAAA^cA^c$ eller AA^cAA^cA. Två saker räknas:

- **Hur många följder?** Lika många som sätten att placera 3 stycken A på 5 platser, alltså $\binom{5}{3}$ enligt Sats 2.1.
- **Hur sannolik är var och en?** På grund av oberoendet multipliceras sannolikheterna: $(1/6)^3(5/6)^2$, samma för alla följder.

Alltså

$$P(\text{exakt 3 ettor}) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 \approx 0{,}032$$

---

3. Sats 2.8

Ersätt tärningskastet med ett godtyckligt delförsök och ettan med en godtycklig händelse A.

Sats 2.8. Betrakta ett försök som består av n oberoende upprepningar av ett delförsök, och låt A vara en händelse som inträffar vid varje delförsök med sannolikheten p. Då är sannolikheten att A inträffar exakt k gånger i försöket

$$\boxed{P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

Två villkor

1. **Samma** p i varje upprepning.
2. **Oberoende** upprepningar.

Drar vi utan återläggning bryts båda: sannolikheten ändras efter varje dragning, och dragningarna beror av varandra. Då gäller i stället urnmodellen.

Sannolikheterna $\binom{n}{k}p^k(1-p)^{n-k}$ kallas **binomialsannolikheter**, eftersom binomialkoefficienterna ingår. I kapitel 3 får de ett eget namn: binomialfördelningen.

---

4. Lösta lektionsuppgifter

Lektion 3, övning 6 — Fyra sexor på sju kast

Pitepilten Pelle kastar en symmetrisk tärning 7 gånger. Betrakta kasten som 7 oberoende upprepningar av att kasta en gång. Vad är sannolikheten att han får exakt 4 sexor? Var kommer oberoendeantagandet in? Generalisera sedan till n kast och k sexor, och därefter till en godtycklig händelse A med sannolikhet p.

Lösning

Låt A = sexa i ett kast, $P(A) = 1/6$.

Antalet följder. Antalet kastserier med 4 sexor och 3 icke-sexor är antalet sätt att välja vilka 4 av de 7 kasten som ger sexa: $\binom{7}{4} = 35$.

Sannolikheten för varje följd. Här kommer oberoendet in: sannolikheten för en viss följd är produkten av sannolikheterna i varje kast, $(1/6)^4(5/6)^3$.

$$P(\text{exakt 4 sexor}) = \binom{7}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^3 \approx \boxed{0{,}0156}$$

Generalisering. Med n kast och k sexor blir samma resonemang $\binom{n}{k}(1/6)^k(5/6)^{n-k}$, och med godtyckligt A och p blir det Sats 2.8 ovan.

På räknaren. `binomPdf(7,1/6,4)`.

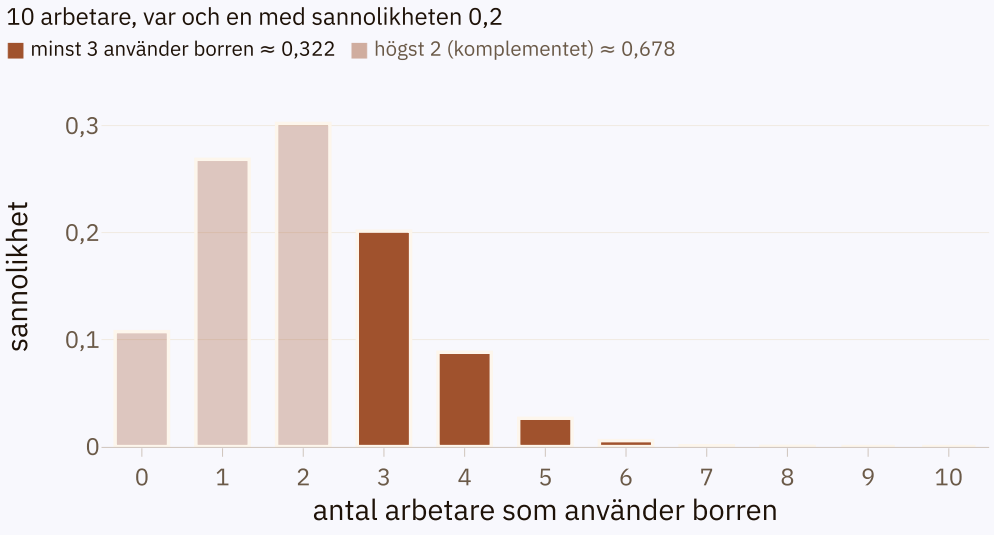
Facit: 0,0156. Stämmer med det handskrivna lösningsförslaget.

Lektion 3, övning 7 (Vännman 2.43) — Handborrmaskiner

Vid ett monteringsband på en verkstad finns 10 arbetare. De använder alla under $20\ \%$ av arbetstiden en handborrmaskin. Händelserna att olika arbetare i ett visst ögonblick använder sin handborrmaskin antas vara oberoende. Vad är sannolikheten att minst tre av arbetarna i ett visst ögonblick använder sin handborrmaskin?

Lösning

Varje arbetare är en oberoende upprepning med $p = 0{,}2$ och $n = 10$. Låt A_k = exakt k använder sin maskin. Via komplementet “högst 2”:



$$P(A_0) = 0{,}8^{10} \approx 0{,}1074$$
$$P(A_1) = \binom{10}{1} 0{,}2 \cdot 0{,}8^{9} \approx 0{,}2684$$
$$P(A_2) = \binom{10}{2} 0{,}2^{2} \cdot 0{,}8^{8} \approx 0{,}3020$$

$$P(\text{minst 3}) = 1 - (0{,}1074 + 0{,}2684 + 0{,}3020) \approx \boxed{0{,}322}$$

På räknaren. `binomCdf(10,0.2,3,10)`.

Svar i boken (2:a uppl., 2.42 = 3:e uppl. 2.43): 0,32.

Se även

- Oberoende händelser
- Binomialkoefficient
- Binomialfördelning
- Klassiska sannolikhetsdefinitionen

Resurser

- 3Blue1Brown: Binomial distributions — samma räkning som här, med bra intuition för varför binomialkoefficienten dyker upp
- Wikipedia: Bernoulli trial

Stokastisk variabel

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #grundbegrepp

> **Kurs:** S0008M **Förkunskaper:** Mängdlära

En stokastisk variabel är ingen variabel i vanlig mening — det är en **funktion** som översätter varje slumpmässigt utfall till ett tal. Genom den kan vi räkna med sannolikheter för tal i stället för för händelser.

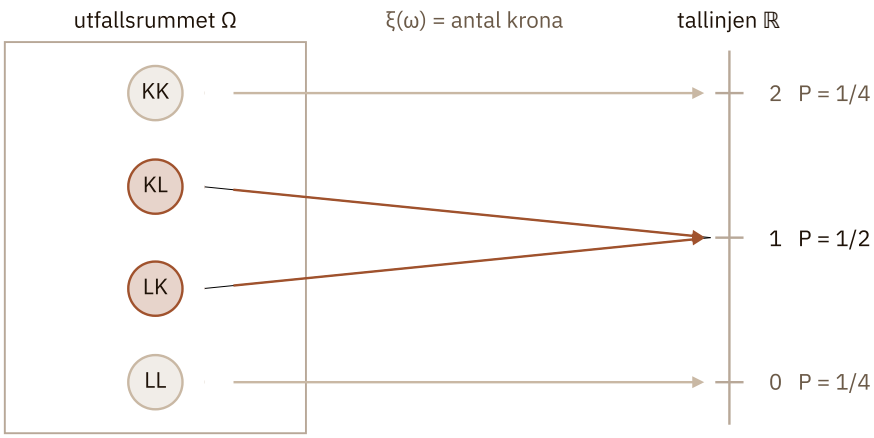
---

1. Definitionen

Ett slumpförsök har ett utfallsrum Ω: mängden av allt som kan hända. En stokastisk variabel ξ är en avbildning

$$\boxed{\xi : \Omega \to \mathrm{R}}$$

som ger varje utfall $\omega \in \Omega$ ett reellt tal $\xi(\omega)$. Vid ett tärningskast är Ω de sex sidorna, och ξ kan vara antalet prickar.



Figuren visar idén för två myntkast: fyra utfall i Ω, men bara tre värden på ξ. Sannolikheten för ett värde är summan av sannolikheterna för de utfall som avbildas dit.

Vi skriver själva variabeln med grekiska bokstäver — ξ, η, ζ — och de värden den antar med latinska småbokstäver. Skrivsättet $P(\xi = x)$ läses alltså "sannolikheten att den stokastiska variabeln ξ råkar landa på talet x".

Variabeln är inte sitt värde

ξ är funktionen, x är utfallet av ett enskilt försök. Skillnaden känns pedantisk tills man skriver $P(\xi \leq x)$ — där är ξ det slumpmässiga och x en fix gräns vi jämför mot.

---

2. Diskret eller kontinuerlig

Uppdelningen styr vilken maskin vi räknar med, och allt annat följer av den.

	Antar	Beskrivs av	Räknas med
Diskret	ändligt eller uppräkneligt många värden	sannolikhetsfunktion $p_\xi(x) = P(\xi = x)$	summa
Kontinuerlig	alla värden i ett intervall	frekvensfunktion $f_\xi(x)$	integral

Gemensamt för båda är fördelningsfunktionen $F_\xi(x) = P(\xi \leq x)$, som är definierad för alla reella x oavsett typ.

Stokastisk modell

Att välja en stokastisk variabel och en fördelning för den kallas att ställa upp en **stokastisk modell** för försöket. De vanligaste modellerna har egna namn: binomial-, hypergeometrisk-, Poisson-, likformig och ffg-fördelning. Att känna igen vilken situation som hör till vilken modell är en stor del av kursen.

Exempel — Två myntkast

Två symmetriska mynt kastas. Låt ξ vara antalet krona. Bestäm fördelningen för ξ och $P(\xi \geq 1)$.

Lösning

Utfallsrummet är $\Omega = \{KK, KL, LK, LL\}$ med fyra lika sannolika utfall (K = krona, L = klave). ξ avbildar dem på

$$\xi(KK) = 2, \quad \xi(KL) = \xi(LK) = 1, \quad \xi(LL) = 0$$

Varje värde får summan av sina utfalls sannolikheter:

x	0	1	2
$P(\xi = x)$	1/4	1/2	1/4

$$P(\xi \geq 1) = 1 - P(\xi = 0) = \boxed{\frac{3}{4}}$$

Detta är $\mathrm{Bin}(2,\ 0{,}5)$; binomialfördelningen är just den här konstruktionen för n försök.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **Blanda ihop ξ och x.** $P(\xi = x)$ är en funktion av x; $P(x)$ betyder ingenting.
- **Tro att värdena är lika sannolika för att utfallen är det.** Utfallen i Ω är lika sannolika, men värdet 1 nås av två utfall och får dubbelt så stor sannolikhet.

Se även

- Sannolikhetsfunktion
- Fördelningsfunktion
- Kontinuerliga stokastiska variabler
- Väntevärde
- Varians och standardavvikelse
- Binomialfördelning
- Utfallsrum och händelser

Resurser

- Khan Academy: Random variables
- Wikipedia: Stokastisk variabel

Sannolikhetsfunktion

19 okt. 2026, 2 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Stokastisk variabel

Sannolikhetsfunktionen till en diskret stokastisk variabel anger sannolikheten för varje enskilt värde, $p_\xi(x) = P(\xi = x)$. Den kan ses som en massfördelning på reella axeln: massan $P(\xi = x)$ i varje punkt, med totalmassan 1.

---

1. Definitionen

För en diskret stokastisk variabel ξ som kan anta värdena $x_1, x_2, \ldots$ är **sannolikhetsfunktionen**

$$\boxed{p_\xi(x) = P(\xi = x)}$$

Den är noll för alla x som ξ inte kan anta. Vi skriver ofta bara $p(x)$ när det är tydligt vilken variabel som avses.

---

2. Två egenskaper

Varje sannolikhetsfunktion uppfyller

$$p_\xi(x) \geq 0 \quad \text{för alla } x, \qquad \sum_x p_\xi(x) = 1$$

där summan tas över alla värden ξ kan anta. Den andra egenskapen är ett bra kontrollsteg: har vi räknat fram en hel sannolikhetsfunktion ska värdena summera till 1.

---

3. Massfördelning och stolpdiagram

Sannolikhet som massa på reella axeln

Placera massan $P(\xi = x)$ i punkten x på tallinjen. Den totala massan är 1. Sannolikheten att ξ hamnar i en mängd A är massan som ligger där:

$$P(\xi \in A) = \sum_{x \in A} p_\xi(x)$$

Vi ritar sannolikhetsfunktionen som ett **stolpdiagram**, en stolpe per möjligt värde med höjden $p_\xi(x)$. Från stolpdiagrammet kan vi läsa av det mest sannolika värdet (högsta stolpen) och uppskatta väntevärdet (masscentrum).

Summerar vi stolparna från vänster får vi fördelningsfunktionen $F_\xi(x) = P(\xi \le x)$.

---

4. Lösta lektionsuppgifter

Lektion 4, övning 2 — Sökningar mot en databas

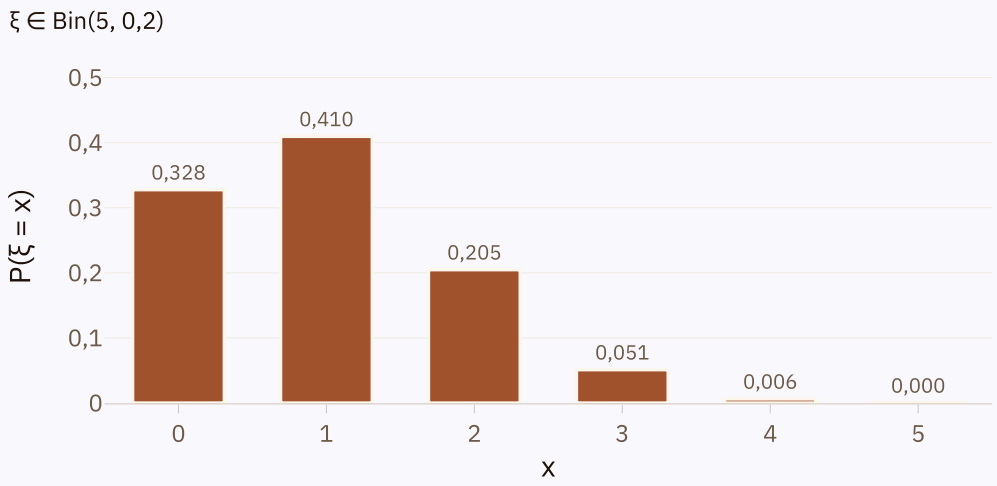
I ett rum står 5 datorer uppkopplade mot en server. Under $20\ \%$ av tiden gör var och en sökningar i en databas, oberoende av varandra. Låt ξ = antal sökningar som pågår samtidigt vid en given tidpunkt. Vilka värden kan ξ anta? Bestäm sannolikhetsfunktionen $P(\xi = x)$.

Lösning

Värden. Mellan ingen och alla fem datorer kan söka samtidigt, så ξ kan anta värdena $0, 1, 2, 3, 4, 5$.

Sannolikhetsfunktion. Varje dator är en oberoende upprepning där "söker" har sannolikheten $0{,}2$. Sats 2.8 ger

$$\boxed{P(\xi = x) = \binom{5}{x} 0{,}2^x\, 0{,}8^{5-x}, \qquad x = 0, 1, \dots, 5}$$



Kontroll. Stolparna summerar till 1: $0{,}328 + 0{,}410 + 0{,}205 + 0{,}051 + 0{,}006 + 0{,}000 = 1{,}000$.

Facit: ξ kan anta 0, 1, ..., 5 och $P(\xi = x) = \binom{5}{x} 0{,}2^x 0{,}8^{5-x}$.

Se även

- Stokastisk variabel
- Fördelningsfunktion
- Binomialfördelning
- Väntevärde
- Frekvensfunktion

Resurser

- Khan Academy: Random variables
- Wikipedia: Probability mass function

Fördelningsfunktion

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Sannolikhetsfunktion

Fördelningsfunktionen anger sannolikheten att en stokastisk variabel är högst x, $F_\xi(x) = P(\xi \leq x)$. Den är definierad för alla reella x, både för diskreta och kontinuerliga variabler, och växer från 0 till 1.

---

1. Definitionen

För en stokastisk variabel ξ är **fördelningsfunktionen**

$$\boxed{F_\xi(x) = P(\xi \leq x), \qquad -\infty < x < \infty}$$

Den beskriver samma fördelning som sannolikhetsfunktionen, men kumulativt: $F_\xi(x)$ är all massa som ligger till vänster om eller i punkten x.

---

2. Egenskaper

- F_ξ är **växande**: flyttar vi x åt höger kan massan till vänster bara öka.
- $F_\xi(x) \to 0$ när $x \to -\infty$ och $F_\xi(x) \to 1$ när $x \to \infty$.
- För en diskret variabel är F_ξ en **trappstegsfunktion** som hoppar i varje möjligt värde x, och hoppet är lika stort som $P(\xi = x)$.

---

3. Mellan sannolikhetsfunktion och fördelningsfunktion

Från sannolikhetsfunktionen summerar vi:

$$F_\xi(x) = \sum_{k \leq x} p_\xi(k)$$

Omvänt, för en variabel som bara antar heltal, är sannolikhetsfunktionen hoppen:

$$p_\xi(x) = F_\xi(x) - F_\xi(x-1)$$

Sannolikheter för intervall

$$P(a < \xi \leq b) = F_\xi(b) - F_\xi(a)$$

Var noga med olikheterna när ξ är diskret. För heltalsvärda ξ är $P(\xi < k) = F_\xi(k-1)$, $P(\xi \geq k) = 1 - F_\xi(k-1)$ och $P(\xi > k) = 1 - F_\xi(k)$.

Tabellerna ger F, inte p

Binomial- och Poissontabellerna i kursboken är tabeller över fördelningsfunktionen. Enskilda sannolikheter får vi som skillnader, $P(\xi = x) = F(x) - F(x-1)$. Miniräknarens `binomcdf` och `poissoncdf` ger också F.

Vanliga fel

- $\leq$ **eller** $<$ **för diskreta variabler.** $P(\xi < 3) = F_\xi(2)$ för en heltalsvariabel, inte $F_\xi(3)$. För kontinuerliga variabler spelar det ingen roll.
- **"Minst" blir komplement.** $P(\xi \geq 3) = 1 - F_\xi(2)$ för heltal, inte $1 - F_\xi(3)$.
- **Tabellen ger F.** Värdet i tabellen är den kumulativa sannolikheten; $P(\xi = k)$ får man som $F(k) - F(k-1)$.

---

4. Lösta lektionsuppgifter

Lektion 4, övning 5 — Tre jobb till en entreprenör

En entreprenör lägger anbud på tre olika jobb, A, B och C. Sannolikheterna att få dem är $P(A) = 0{,}5$, $P(B) = 0{,}8$ och $P(C) = 0{,}3$, oberoende av varandra. Låt ξ = totala antalet jobb entreprenören får.

a) Vilka värden kan ξ anta?

b) Bestäm sannolikhetsfunktionen och fördelningsfunktionen, och rita dem.

c) Beräkna $P(0 < \xi \leq 2)$.

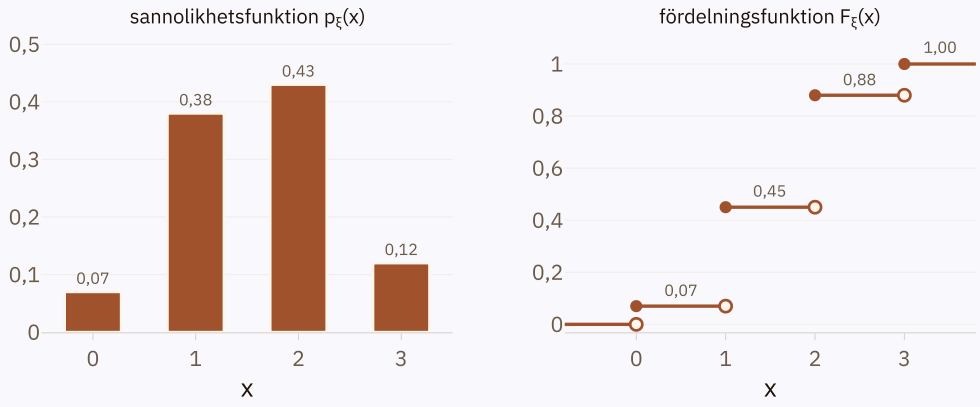
Lösning

a) ξ kan anta värdena $0, 1, 2, 3$.

b) Varje värde är en union av disjunkta händelser, och varje snitt räknas med produktregeln för oberoende händelser:

$$
\begin{aligned}
p(0) &= P(A^c \cap B^c \cap C^c) = 0{,}5 \cdot 0{,}2 \cdot 0{,}7 = 0{,}07 \\
p(1) &= P(A \cap B^c \cap C^c) + P(A^c \cap B \cap C^c) + P(A^c \cap B^c \cap C) \\
&= 0{,}5 \cdot 0{,}2 \cdot 0{,}7 + 0{,}5 \cdot 0{,}8 \cdot 0{,}7 + 0{,}5 \cdot 0{,}2 \cdot 0{,}3 = 0{,}38 \\
p(2) &= P(A^c \cap B \cap C) + P(A \cap B^c \cap C) + P(A \cap B \cap C^c) \\
&= 0{,}5 \cdot 0{,}8 \cdot 0{,}3 + 0{,}5 \cdot 0{,}2 \cdot 0{,}3 + 0{,}5 \cdot 0{,}8 \cdot 0{,}7 = 0{,}43 \\
p(3) &= P(A \cap B \cap C) = 0{,}5 \cdot 0{,}8 \cdot 0{,}3 = 0{,}12
\end{aligned}
$$

Summan är 1. Fördelningsfunktionen fås genom att summera: $F(0) = 0{,}07$, $F(1) = 0{,}45$, $F(2) = 0{,}88$, $F(3) = 1$.



c)

$$P(0 < \xi \leq 2) = p(1) + p(2) = 0{,}38 + 0{,}43 = \boxed{0{,}81}$$

eller med fördelningsfunktionen, $F(2) - F(0) = 0{,}88 - 0{,}07 = 0{,}81$.

Svar i boken (2:a uppl., 3.5): p = 0,07, 0,38, 0,43, 0,12; c) 0,81. Uppgiften är återgiven ur det handskrivna lösningsförslaget, och lösningen stämmer också med det.

Se även

- Sannolikhetsfunktion
- Stokastisk variabel
- Binomialfördelning
- Poissonfördelning
- Frekvensfunktion

Resurser

- Khan Academy: Random variables
- Wikipedia: Fördelningsfunktion

Binomialfördelning

19 okt. 2026, 5 min läsning

UTKAST

#matematik #statistik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Stokastisk variabel, Oberoende upprepningar

Binomialfördelningen räknar hur många av n oberoende försök som lyckas, när varje försök lyckas med samma sannolikhet p. Den är kursens vanligaste diskreta fördelning och grunden för både Poisson- och normalapproximationen.

1. Sannolikhetsfunktionen

Upprepa samma försök n gånger, oberoende av varandra, där varje försök antingen lyckas (sannolikhet p) eller misslyckas. Låt ξ vara **antalet lyckade försök**. Då är ξ binomialfördelad, $\xi \in \mathrm{Bin}(n, p)$, med

$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n$$

De tre faktorerna säger var sin sak:

- p^k — de k försök som lyckas. Oberoende, alltså multiplicerat.
- $(1-p)^{n-k}$ — de $n-k$ som misslyckas.
- $\binom{n}{k}$ — antalet sätt att välja *vilka* av de n försöken som är de lyckade. Utan den räknar vi bara en enda specifik ordningsföljd.

Fyra villkor måste hålla

1. **Fixt** n — antalet försök bestäms i förväg, inte av utfallet.
2. **Två utfall** per försök — lyckat eller inte.
3. **Samma** p i alla försök.
4. **Oberoende** försök.

Faller något av dem är det inte en binomialfördelning. Dragning *utan* återläggning bryter mot både 3 och 4.

2. Väntevärde och varians

$$\boxed{E(\xi) = np, \qquad V(\xi) = np(1-p)}$$

Båda följer direkt ur att ξ är en **summa av n oberoende Bernoullivariabler** $\xi = \eta_1 + \cdots + \eta_n$, där varje η_i är 1 vid lyckat försök och 0 annars. Eftersom $E(\eta_i) = p$ och $V(\eta_i) = p(1-p)$ ger väntevärdets linjäritet och variansens additivitet vid oberoende resultatet med n termer.

Variansen är som störst vid $p = 1/2$ och krymper mot noll när p närmar sig 0 eller 1 — vilket är rimligt, för då är utfallet nästan givet på förhand.

Exempel — exakt två sexor på fem tärningskast

Vi kastar en rättvis sexsidig tärning fem gånger. Vad är sannolikheten att exakt två av kasten ger en sexa?

Lösning

Setup. "Lyckat" = sexa, alltså $p = \frac{1}{6}$, $n = 5$, och vi söker $k = 2$. Kasten är oberoende och p är densamma i alla, så $\xi \in \mathrm{Bin}(5, \frac{1}{6})$.

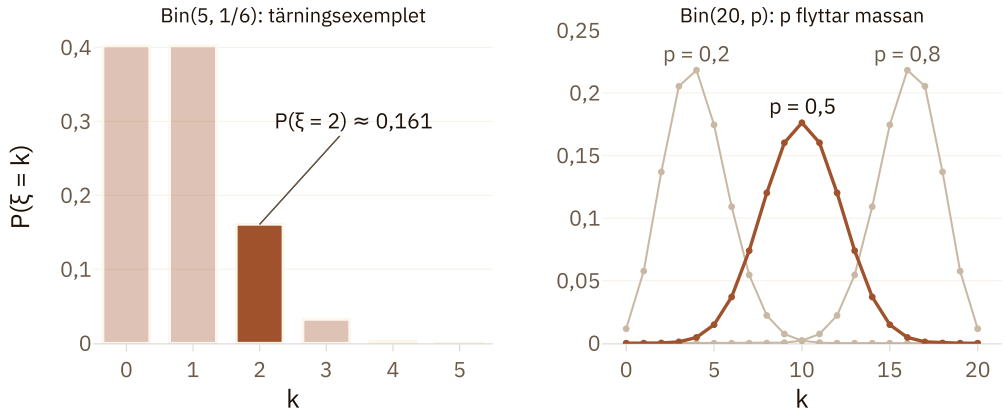
Räkningen.

$$P(\xi = 2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = 10 \cdot \frac{1}{36} \cdot \frac{125}{216} = \boxed{\frac{625}{3888}} \approx 0{,}161$$

Rimlighetskoll. Väntevärdet är $E(\xi) = 5 \cdot \frac{1}{6} \approx 0{,}83$ sexor. Två sexor ligger alltså över det förväntade men bara drygt en standardavvikelse bort, eftersom $\sigma = \sqrt{5 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 0{,}83$. Att det landar på ungefär $16\ \%$ är därför rimligt — ovanligt, men långt ifrån osannolikt.

På räknaren. `binomPdf(5,1/6,2)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.



Till vänster syns tärningsexemplet: massan ligger vid $k = 0$ och $k = 1$, precis som väntevärdet $0{,}83$ antyder. Till höger flyttar p tyngdpunkten i sidled, och staplarna är som bredast vid $p = 1/2$ — samma sak som att variansen $np(1-p)$ maximeras där.

---

3. Släktingar och gränsfall

Binomialfördelningen sitter mitt i ett nät av andra fördelningar, och det är ofta enklare att räkna med grannarna:

Situation	Vad som händer
$n = 1$	Fördelningen reduceras till en Bernoullifördelning $\mathrm{Be}(p)$
$p \le 0{,}1$ och $n \ge 10$	Poissonapproximation $\mathrm{Po}(np)$
$np(1-p) \ge 10$	Normalapproximation $N(np, \sqrt{np(1-p)})$ — se normalfördelningen
Dragning utan återläggning	Hypergeometrisk fördelning i stället — försöken är inte oberoende

Beteckning

Kursen skriver $\xi \in \mathrm{Bin}(n, p)$ med $\in$, som Vännman. I annan litteratur är $\xi \sim \mathrm{Bin}(n, p)$ vanligare.

Varför normalapproximationen dyker upp

Eftersom ξ är en summa av många oberoende likafördelade termer är centrala gränsvärdessatsen tillämplig. Det är samma mekanism som gör normalfördelningen till standardsvar överallt annars.

⚠ **Vanliga fel**

- **Dragning utan återläggning är inte binomisk.** Ändras sannolikheten mellan försöken är det hypergeometrisk fördelning.
- **Glömma binomialkoefficienten.** $P(\xi = k) = \binom{n}{k}p^k(1-p)^{n-k}$; utan $\binom{n}{k}$ får man sannolikheten för *en* bestämd ordning.
- **Fel gräns i `binomCdf`.** `binomCdf(n,p,a,b)` ger $P(a \le \xi \le b)$ med båda gränserna inräknade; "fler än 3" börjar på 4.

---

4. Lösta lektionsuppgifter

Lektion 4, övning 1 och 3 — Samtidiga databassökningar

I ett rum står 5 datorer uppkopplade mot en server. Under $20\ \%$ av tiden gör var och en sökningar i en databas, oberoende av varandra. Vad är sannolikheten att det vid en given tidpunkt pågår

a) exakt 2 sökningar samtidigt,

b) minst 2 sökningar samtidigt?

Vilken fördelning har ξ = antalet samtidiga sökningar, och med vilka parametrar?

Lösning

Modell. Fem oberoende datorer, var och en söker med sannolikheten $0{,}2$. Antalet som söker är därför binomialfördelat: $\xi \in \mathrm{Bin}(5,\ 0{,}2)$, med parametrarna $n = 5$ och $p = 0{,}2$.

a)

$$P(\xi = 2) = \binom{5}{2}\,0{,}2^2\,0{,}8^3 = 10 \cdot 0{,}04 \cdot 0{,}512 = \boxed{0{,}2048}$$

b) Via komplementet "högst 1":

$$P(\xi \ge 2) = 1 - P(\xi = 0) - P(\xi = 1) = 1 - 0{,}8^5 - 5 \cdot 0{,}2 \cdot 0{,}8^4 = 1 - 0{,}32768 - 0{,}4096$$

På räknaren. `binomPdf(5,0.2,2)` respektive `binomCdf(5,0.2,2,5)`.

Facit: a) 0,2048, b) 0,2627; $\xi \in \mathrm{Bin}(5,\ 0{,}2)$.

Lektion 4, övning 10 — Slumpvandring

En partikel utför slumpvandring på heltalen: den startar på talet 0 och hoppar varje sekund ett steg, i negativ eller positiv riktning med sannolikheten $0{,}5$ vardera. Vad är sannolikheten att partikeln efter 10 sekunder befinner sig på talet 2?

Lösning

Översätt till antal steg. Låt η = antal steg i positiv riktning av de 10. Positionen efter 10 steg är $\eta - (10 - \eta) = 2\eta - 10$. Att hamna på 2 betyder $2\eta - 10 = 2$, alltså $\eta = 6$.

Fördelning. Stegen är oberoende med samma sannolikhet $0{,}5$ för positivt steg, så $\eta \in \mathrm{Bin}(10,\ 0{,}5)$.

$$P(\eta = 6) = \binom{10}{6} 0{,}5^{10} = \frac{210}{1024} \approx \boxed{0{,}205}$$

Rimlighetskoll. Paritet: efter ett jämnt antal steg står partikeln alltid på ett jämnt tal, så talet 2 är nåbart. Ett udda tal hade gett sannolikheten 0.

Facit: 0,205.

Se även

- Stokastisk variabel
- Hypergeometrisk fördelning
- Poissonfördelning
- Normalfördelning
- Talföljder och summor
- Mängdlära

Resurser

- 3Blue1Brown: Binomial distributions — bygger upp fördelningen från grunden, med bra intuition för binomialkoefficienten
- Khan Academy: Binomial random variables
- Wikipedia: Binomialfördelning

Hypergeometrisk fördelning

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Binomialfördelning

Den hypergeometriska fördelningen räknar hur många markerade objekt vi får när vi drar ett stickprov **utan återläggning** ur en ändlig population. Den är binomialfördelningens motsvarighet när dragningarna påverkar varandra.

---

1. Sannolikhetsfunktionen

Ur en population på N objekt, varav K är markerade, drar vi n stycken **utan återläggning**. Låt ξ vara antalet markerade i dragningen. Då är

$$P(\xi = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Vännman skriver fördelningen med andelen markerade $p = K/N$ i stället för antalet, $\xi \in \mathrm{Hyp}(N,\ n,\ p)$, så att $K = Np$.

Kvoten är ren kombinatorik: nämnaren räknar alla möjliga dragningar, täljaren de gynnsamma — välj k av de markerade och resterande $n - k$ av de omarkerade.

> **Skillnaden mot binomialfördelningen**
>
> Utan återläggning ändras populationen för varje draget objekt, så försöken är varken oberoende eller likafördelade. Just därför duger inte $\mathrm{Bin}(n, p)$ här — samma räkning som dragning ur en urna.

---

2. Väntevärde och varians

$$\boxed{E(\xi) = n\frac{K}{N}, \qquad V(\xi) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}}$$

Väntevärdet är detsamma som för binomialfördelningen med $p = K/N$. Variansen skiljer sig med faktorn $\frac{N-n}{N-1}$, den **ändliga populationskorrektionen**: den är alltid ≤ 1, så utan återläggning sprider utfallet mindre. Är N stor jämfört med n går faktorn mot 1 och fördelningen mot $\mathrm{Bin}(n, K/N)$ — därför spelar det i praktiken ingen roll om man drar med eller utan återläggning ur en tillräckligt stor population.

Exempel — defekta komponenter i ett stickprov

En låda innehåller 20 komponenter, varav 5 är defekta. Vi drar 4 stycken utan återläggning. Vad är sannolikheten att exakt en är defekt?

Lösning

Setup. $N = 20$, $K = 5$ defekta, $n = 4$ dragna, och vi söker $k = 1$.

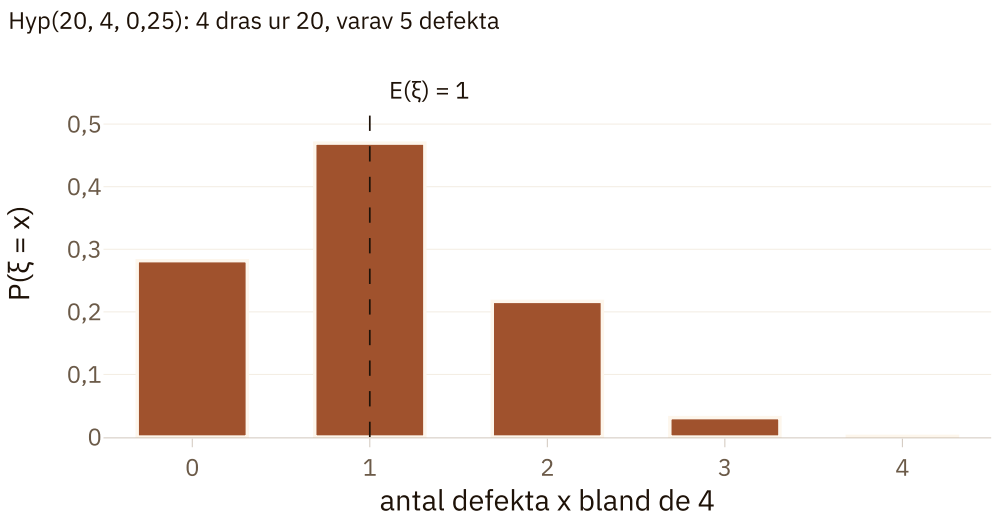
Räkningen.

$$P(\xi = 1) = \frac{\binom{5}{1}\binom{15}{3}}{\binom{20}{4}} = \frac{5 \cdot 455}{4845} = \boxed{\frac{455}{969}} \approx 0{,}470$$

Rimlighetskoll. Väntevärdet är $E(\xi) = 4 \cdot \frac{5}{20} = 1$, alltså är exakt en defekt det mest sannolika utfallet — att det svarar för knappt hälften av sannolikhetsmassan stämmer. Binomialapproximationen ger $0{,}422$; att den underskattar är väntat, eftersom den saknar populationskorrektionen och därför sprider ut massan för mycket.

På räknaren. Fördelningen saknas som funktion: `nCr(5,1)*nCr(15,3)/nCr(20,4)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.



Hela fördelningen i exemplet: masscentrum ligger vid väntevärdet 1, och sannolikheten att få alla fyra defekta är nästan noll.

---

3. Approximationer

Är urvalet litet jämfört med populationen spelar det nästan ingen roll att vi drar utan återläggning:

Villkor	Approximation
$n/N \leq 0{,}1$	$\text{Hyp}(N, n, p) \approx \text{Bin}(n, p)$
$p + n/N \leq 0{,}1$ och $n \geq 10$	$\text{Hyp}(N, n, p) \approx \text{Po}(np)$

Poissonapproximationen är ofta enklast att räkna med eftersom den bara har en parameter och finns i tabell.

⚠ **Vanliga fel**

- **Använda binomialfördelningen för små populationer.** Utan återläggning och med $n/N > 0{,}1$ ändras sannolikheten märkbart mellan dragningarna.
- **Tappa korrektionsfaktorn i variansen.** $V(\xi) = \dfrac{N-n}{N-1}\, np(1-p)$; utan faktorn blir variansen för stor.
- **Blanda ihop n och N.** N är populationen, n antalet dragna.

---

4. Lösta lektionsuppgifter

Lektion 4, övning 4 — Parametrarna i urnan

I en urna ligger 25 kulor, 10 röda och resten svarta. Utan återläggning och på måfå väljs 5 kulor. Låt ξ = antalet röda kulor i urvalet. Då är ξ hypergeometriskt fördelad. Vilka är parametrarna?

Lösning

Populationens storlek är $N = 25$, urvalets storlek $n = 5$, och andelen röda är $p = 10/25 = 0{,}4$:

$$\xi \in \boxed{\mathrm{Hyp}(25,\ 5,\ 0{,}4)}$$

Sannolikheterna är desamma som räknades i lektion 2, övning 1, till exempel $P(\xi = 2) \approx 0{,}385$.

Facit: N = 25, n = 5, p = 0,4.

Lektion 5, övning 9 — Stickprov ur ett stort parti

En tillverkningsprocess har felkvoten $0{,}05$. Ur ett parti om 3000 enheter väljs 100 på måfå. Låt ξ = antal defekta bland de 100. Bestäm det minsta tröskelvärdet k sådant att $P(\xi > k) < 0{,}05$.

Lösning

Modell. Utan återläggning ur ett parti med $Np = 3000 \cdot 0{,}05 = 150$ defekta: $\xi \in \mathrm{Hyp}(3000,\ 100,\ 0{,}05)$.

Approximation 1. $n/N = 100/3000 < 0{,}1$, så $\xi \approx \mathrm{Bin}(100,\ 0{,}05)$ — men den finns inte i tabellen.

Approximation 2. $p + n/N = 0{,}05 + 0{,}033 \approx 0{,}083 \le 0{,}1$ och $n > 10$, så $\xi \approx \mathrm{Po}(np) = \mathrm{Po}(5)$. Den är lättare att räkna med.

Tröskeln. Vi söker minsta k med

$$P(\xi > k) < 0{,}05 \quad \Longleftrightarrow \quad P(\xi \leq k) > 0{,}95$$

Poissontabellen för $\lambda = 5$ ger $F(8) \approx 0{,}932$ och $F(9) \approx 0{,}968$, alltså $\boxed{k = 9}$.

Tolkning. Det är mindre än $5\ \%$ risk för fler än 9 defekta. Justerar vi processen när 10 eller fler av 100 är defekta sker det i genomsnitt vid mindre än vart tjugonde stickprov, om processen fungerar som den ska.

Svar i boken (2:a uppl., 3.20): 9. Uppgiften är återgiven ur det handskrivna lösningsförslaget, och svaret stämmer också med det.

Se även

- Binomialfördelning
- Stokastisk variabel
- Mängdlära

Resurser

- Wikipedia: Hypergeometrisk fördelning
- Wikipedia: Hypergeometric distribution — utförligare, med härledningar

---

Poissonfördelning

19 okt. 2026, 5 min läsning

UTKAST

#matematik #statistik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Binomialfördelning

Poissonfördelningen beskriver antalet sällsynta, oberoende händelser under en viss tid eller i ett visst område, till exempel samtal till en växel per minut eller fel per meter kabel. Den har bara en parameter, väntevärdet λ.

---

1. Sannolikhetsfunktionen

Poissonfördelningen räknar **antalet händelser i ett fixt intervall** — per sekund, per kvadratmeter, per produktionsparti. Låt ξ vara det antalet och λ den genomsnittliga intensiteten. Då är $\xi \in \mathrm{Po}(\lambda)$ med

$$P(\xi = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$$

Till skillnad från binomialfördelningen finns ingen övre gräns för k: vi räknar händelser, inte lyckade försök av ett fixt antal.

Tre antaganden

1. **Konstant intensitet** λ över hela intervallet.
2. **Oberoende** — en händelse påverkar inte sannolikheten för nästa.
3. **Inga sammanfallande händelser** — två inträffar aldrig exakt samtidigt.

---

2. Väntevärde, varians och kopplingen till binomialfördelningen

$$E(\xi) = V(\xi) = \lambda$$

Att väntevärde och varians sammanfaller är fördelningens signum: har ett datamaterial ungefär lika stort medelvärde som varians är Poisson en rimlig första modell.

Fördelningen uppstår som **gränsfall av binomialfördelningen** när $n \to \infty$ och $p \to 0$ med produkten $np = \lambda$ konstant (härledningen nedan). Det förklarar dess räckvidd: många möjliga tillfällen, försumbar sannolikhet var för sig, måttligt förväntat antal. Summan av oberoende Poissonvariabler är dessutom själv Poissonfördelad, med λ adderade — därför spelar det ingen roll om vi räknar per sekund eller per minut, bara λ skalas med.

Härledning: Poisson som gränsfall av binomialfördelningen

Sätt $p = \lambda/n$ i binomialfördelningen och låt $n \to \infty$ med k fixt:

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\cdot\underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\to\,1}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{\to\,e^{-\lambda}}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to\,1} \longrightarrow \frac{\lambda^k}{k!}e^{-\lambda}$$

Mittfaktorn går mot $e^{-\lambda}$ enligt standardgränsvärdet $(1 + x/n)^n \to e^x$, se Gränsvärden. Samma gränsövergång ger $E = np = \lambda$ och $V = np(1-p) \to \lambda$.

Exempel — sönderfall i en detektor

Ett preparat ger i genomsnitt 4 registrerade sönderfall per sekund. Vad är sannolikheten att exakt 2 registreras under en given sekund?

Lösning

Setup. Sönderfall är oberoende och intensiteten konstant, så $\xi \in \mathrm{Po}(4)$ med $\lambda = 4$ och vi söker $k = 2$.

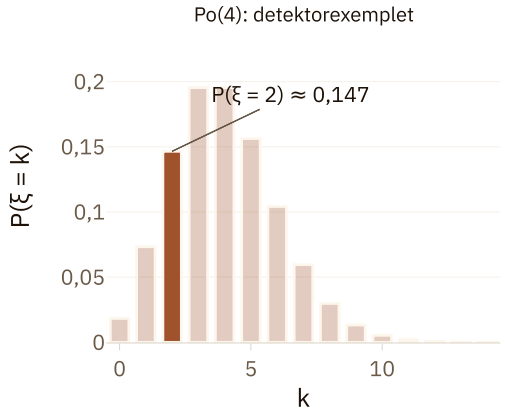
Räkningen.

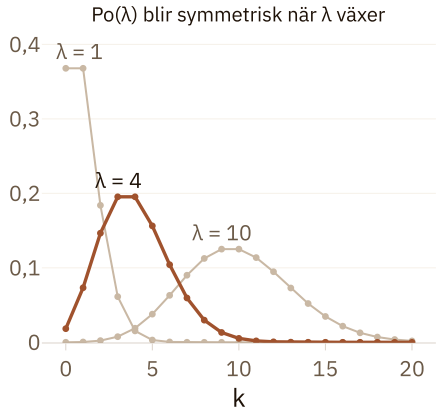
$$P(\xi = 2) = \frac{4^2}{2!}e^{-4} = 8e^{-4} \approx \boxed{0{,}147}$$

Rimlighetskoll. Väntevärdet är 4 och $\sigma = \sqrt{4} = 2$, så $k = 2$ ligger en standardavvikelse under medelvärdet — en dryg tiondel av massan är rimligt där. Binomialmodellen $\mathrm{Bin}(1000,\ 0{,}004)$, alltså många små sönderfallschanser med samma $np = 4$, ger $0{,}146$. Gränsövergången stämmer.

På räknaren. `poissPdf(4,2)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.





Till vänster detektorexemplet, med $k = 2$ markerat strax under toppen vid $\lambda = 4$. Till höger syns hur fördelningen ändrar karaktär med λ: vid $\lambda = 1$ är den kraftigt högersned med sin massa vid noll, medan den vid $\lambda = 10$ har blivit nästan symmetrisk — det är normalapproximationen som börjar gälla.

Radioaktivt sönderfall är standardexemplet just för att antagandena håller exakt — se Kärnfysik.

---

3. Approximationer

Villkor	Approximation
$\xi \in \text{Bin}(n, p)$ med $p \leq 0{,}1$ och $n \geq 10$	$\xi \approx \text{Po}(np)$
$\xi \in \text{Hyp}(N, n, p)$ med $p + n/N \leq 0{,}1$ och $n \geq 10$	$\xi \approx \text{Po}(np)$

Poissontabellen i kursboken ger fördelningsfunktionen $P(\xi \leq x)$, precis som `poissCdf(λ,0,x)` på räknaren.

⚠ **Vanliga fel**

- **Glömma att skala λ.** Är intensiteten 4 per timme blir antalet under 30 minuter $\text{Po}(2)$, inte $\text{Po}(4)$.
- **Ta roten ur fel storhet.** $V(\xi) = \lambda$, så standardavvikelsen är $\sqrt{\lambda}$, inte λ.
- **"Minst en" utan komplement.** $P(\xi \geq 1) = 1 - e^{-\lambda}$ är enklast; att summera uppåt tar aldrig slut.

4. Lösta lektionsuppgifter

Lektion 5, övning 1 — Kunder till ett systembolag

Antalet kunder som anländer till ett systembolag under en 3-minutersperiod antas vara $\xi \in \mathrm{Po}(4)$. Beräkna sannolikheten att det under en 3-minutersperiod anländer

a) högst 1 kund,

b) minst 1 kund,

c) mer än 5 kunder,

d) exakt 4 kunder.

Lösning

$$P(\xi = k) = \frac{4^k}{k!}e^{-4}.$$

a)

$$P(\xi \le 1) = e^{-4}(1 + 4) = 5e^{-4} \approx \boxed{0{,}092}$$

b) Komplementet till "ingen":

$$P(\xi \ge 1) = 1 - e^{-4} \approx \boxed{0{,}982}$$

c) Komplementet till "högst 5". Tabellen (eller `poissCdf(4,0,5)`) ger $F(5) \approx 0{,}785$:

$$P(\xi > 5) = 1 - F(5) \approx \boxed{0{,}215}$$

d)

$$P(\xi = 4) = \frac{4^4}{4!}e^{-4} = \frac{256}{24}e^{-4} \approx \boxed{0{,}195}$$

Facit: a) 0,092, b) 0,982, c) 0,215, d) 0,195.

Lektion 4, övning 8 — Fartyg till en hamn

Antalet fartyg som anländer till en hamn under en dag är Poissonfördelat med $\lambda = 2$.

a) Vad är sannolikheten att fler än 3 fartyg anländer en viss dag?

b) Hur många fartyg måste hamnen ha kapacitet för att med minst $95\ \%$ sannolikhet kunna betjäna alla fartyg som anländer en bestämd dag?

Lösning

ξ = antal fartyg en dag, $\xi \in \text{Po}(2)$, $P(\xi = x) = \dfrac{2^x}{x!}e^{-2}$.

a)

$$P(\xi > 3) = 1 - P(\xi \leq 3) = 1 - e^{-2}\left(1 + 2 + \frac{4}{2} + \frac{8}{6}\right) \approx 1 - 0{,}857 = \boxed{0{,}143}$$

b) Vi söker minsta kapacitet c med $P(\xi \leq c) \geq 0{,}95$:

$$P(\xi \leq 3) \approx 0{,}857, \qquad P(\xi \leq 4) \approx 0{,}947, \qquad P(\xi \leq 5) \approx 0{,}983$$

Fyra räcker inte, fem gör det: kapacitet för $\boxed{5 \text{ fartyg}}$.

Svar i boken (2:a uppl., 3.19): a) 0,143, b) 5 fartyg. Uppgiften är återgiven ur det handskrivna lösningsförslaget, och svaret stämmer också med det.

Se även

- Binomialfördelning
- Stokastisk variabel
- Hypergeometrisk fördelning

- Kärnfysik

Resurser

- Khan Academy: Poisson distribution
- Wikipedia: Poissonfördelning

Likformig fördelning

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Sannolikhetsfunktion

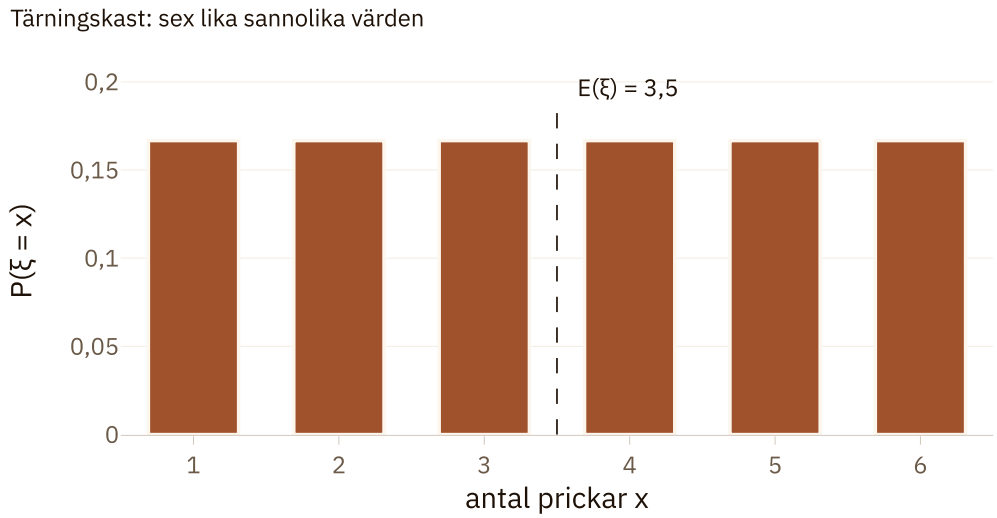
En stokastisk variabel är likformigt fördelad när inget värde är mer troligt än något annat. I det diskreta fallet är det den klassiska sannolikhetsdefinitionen översatt till stokastiska variabler. I det kontinuerliga fallet, rektangelfördelningen $R(a, b)$, är det modellen för “någon gång i intervallet, helt slumpmässigt”: en ankomsttid, ett avrundningsfel, en brottpunkt.

1. Sannolikhetsfunktionen

Om ξ kan anta de m värdena $x_1, \ldots, x_m$ och alla är lika sannolika är

$$P(\xi = x_i) = \frac{1}{m}, \qquad i = 1, \ldots, m$$

Standardexemplet är antalet prickar vid kast med en symmetrisk tärning: $m = 6$ och $P(\xi = x) = 1/6$ för $x = 1, \ldots, 6$.



Fördelningen är samma sak som den klassiska sannolikhetsdefinitionen: m lika sannolika utfall, där varje värde är ett gynnsamt utfall.

2. Väntevärde och varians för $1, 2, \ldots, m$

När värdena är heltalen $1, 2, \ldots, m$ ligger masscentrum mitt i:

$$E(\xi) = \frac{m+1}{2}, \qquad V(\xi) = \frac{m^2-1}{12}$$

För tärningen blir $E(\xi) = 3{,}5$ och $V(\xi) = 35/12 \approx 2{,}92$.

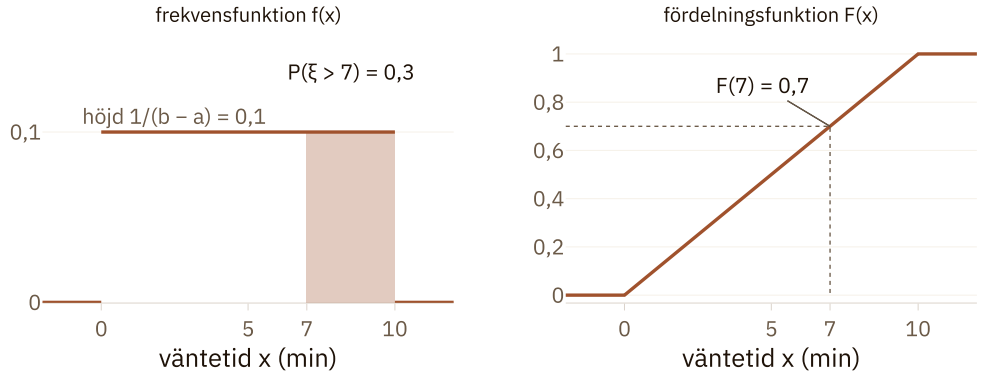
3. Kontinuerlig likformig fördelning: $R(a, b)$

En kontinuerlig variabel som är lika sannolik överallt i intervallet $[a, b]$ är **rektangelfördelad**, $\xi \in R(a, b)$. Frekvensfunktionen är konstant, och konstanten måste vara $1/(b-a)$ för att arean ska bli 1:

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{annars} \end{cases} \qquad F_\xi(x) = \frac{x-a}{b-a}, \quad a \le x \le b$$

Sannolikheten för ett delintervall är alltså dess längd delad med hela intervallets längd, samma idé som den klassiska definitionen men med längder i stället för antal.

$$E(\xi) = \frac{a+b}{2}, \qquad V(\xi) = \frac{(b-a)^2}{12}$$



Härledning av $E(\xi)$ och $V(\xi)$

$$E(\xi) = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}, \qquad E(\xi^2) = \int_a^b \frac{x^2}{b-a}\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b}{3}$$

$$V(\xi) = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{(b-a)^2}{12}$$

Väntevärdet är mittpunkten, vilket symmetrin kräver. Standardavvikelsen $(b-a)/\sqrt{12} \approx 0{,}29\,(b-a)$ beror bara på intervallets längd.

Exempel — Väntan på bussen

En buss går var tionde minut, och vi kommer till hållplatsen utan att titta på tidtabellen. Väntetiden är då $\xi \in R(0, 10)$ minuter.

a) Vad är sannolikheten att vi får vänta mer än 7 minuter?

b) Bestäm väntevärde och standardavvikelse för väntetiden.

Lösning

a) Intervallet $(7, 10]$ har längden 3 av totalt 10:

$$P(\xi > 7) = 1 - F_\xi(7) = 1 - \frac{7}{10} = \boxed{0{,}3}$$

b)

$$E(\xi) = \frac{0+10}{2} = \boxed{5 \text{ min}}, \qquad D(\xi) = \frac{10-0}{\sqrt{12}} \approx \boxed{2{,}89 \text{ min}}$$

Samma väntetid återkommer i lektion 19, där två sådana väntetider adderas och summan *inte* blir likformig.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **Glömma nämnaren.** $f_\xi(x) = 1/(b-a)$, inte 1. Bara på intervall med längden 1, som $R(0, 1)$, är höjden 1.
- **Tolv, inte två.** Variansen är $(b-a)^2/12$. Standardavvikelsen är $(b-a)/\sqrt{12}$, inte $(b-a)/2$.
- **Diskret eller kontinuerlig.** En tärning har $V = (m^2-1)/12 = 35/12$, rektangelfördelningen på $[1, 6]$ har $25/12$. Formlerna är inte utbytbara.

Se även

- Klassiska sannolikhetsdefinitionen
- Icke-linjära funktioner av stokastiska variabler
- Sannolikhetsfunktion
- Väntevärde
- Kontinuerliga stokastiska variabler

Resurser

- Wikipedia: Discrete uniform distribution

ffg-fördelning

19 okt. 2026, 2 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Oberoende upprepningar

---

ffg-fördelningen, för-första-gången-fördelningen, beskriver antalet oberoende försök som krävs tills en händelse inträffar för första gången. Till skillnad från binomialfördelningen är antalet försök inte bestämt i förväg.

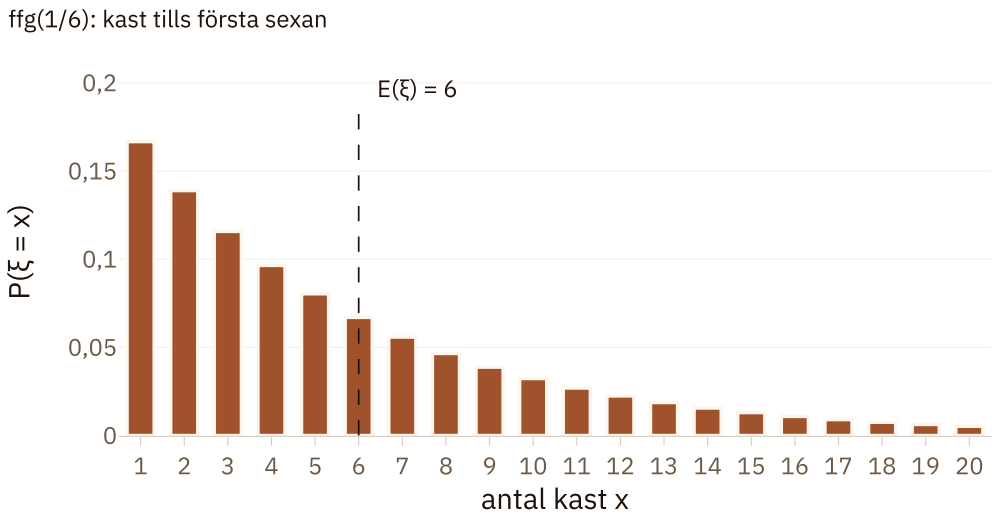
---

1. Sannolikhetsfunktionen

Upprepa ett delförsök oberoende tills händelsen A, med sannolikheten p i varje försök, inträffar för första gången. Låt ξ = antalet försök som krävs, det sista inräknat. Att $\xi = k$ betyder $k-1$ misslyckanden följda av ett lyckat:

$$P(\xi = k) = (1-p)^{k-1}\,p, \qquad k = 1, 2, 3, \ldots$$

Vi skriver $\xi \in \text{ffg}(p)$. Sannolikheterna avtar geometriskt: varje steg multiplicerar med $1-p$.



På räknaren heter fördelningen `geomPdf(p,k)` och `geomCdf(p,a,b)` — räknarens geometriska fördelning är samma som ffg, den börjar på 1.

2. Väntevärde och varians

$$E(\xi) = \frac{1}{p}, \qquad V(\xi) = \frac{1-p}{p^2}$$

Väntevärdet är intuitivt: med sannolikheten $1/6$ för en sexa behövs i genomsnitt 6 kast för den första.

Härledning av $E(\xi) = 1/p$

Med $q = 1 - p$ och derivatan av den geometriska serien, $\sum_{k\geq 1} k\,q^{k-1} = \frac{d}{dq}\frac{1}{1-q} = \frac{1}{(1-q)^2}$:

$$E(\xi) = \sum_{k=1}^{\infty} k\,q^{k-1}p = p \cdot \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Variansen följer på samma sätt ur andraderivatan av serien.

Fördelningsfunktionen har en enkel sluten form. “Mer än k försök behövs” betyder att de k första alla misslyckas:

$$P(\xi > k) = (1-p)^k, \qquad F_\xi(k) = 1 - (1-p)^k$$

Exempel — Kast tills första sexan

Vi kastar en tärning tills den första sexan kommer.

a) Vad är sannolikheten att det behövs mer än 6 kast?

b) Vad är sannolikheten att det räcker med högst 3 kast?

Lösning

Modell. ξ = antal kast, $\xi \in \text{ffg}(1/6)$.

a) De sex första kasten ska alla sakna sexa:

$$P(\xi > 6) = \left(\frac{5}{6}\right)^6 \approx \boxed{0{,}335}$$

Trots att väntevärdet är 6 kast behövs fler än sex i en tredjedel av fallen. Fördelningen har en lång svans åt höger.

b)

$$P(\xi \leq 3) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216} \approx \boxed{0{,}421}$$

På räknaren. b) `geomCdf(1/6,1,3)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

⚠ **Två varianter i litteraturen**

Ibland räknas i stället antalet *misslyckanden* före det första lyckade, som då börjar på 0. Den varianten kallas ofta geometrisk fördelning och har väntevärdet $(1-p)/p$. Kontrollera alltid vilken som avses.

Se även

- Oberoende upprepningar
- Binomialfördelning
- Stokastisk variabel
- Exponentialfördelning

Resurser

- Wikipedia: Geometric distribution

Väntevärde

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #storhet

> **Kurs:** S0008M **Förkunskaper:** Sannolikhetsfunktion, Frekvensfunktion

---

Väntevärdet $E(\xi)$ är fördelningens tyngdpunkt — det värde vi i genomsnitt får i det långa loppet om försöket upprepas många gånger. Vi beräknar det som summan, eller integralen, av varje värde vägt med sin sannolikhet.

---

1. Definitionen

Väntevärdet betecknas $E(\xi)$ eller kort μ.

Diskret stokastisk variabel, med sannolikhetsfunktionen $p_\xi(x)$:

$$\boxed{E(\xi) = \sum_x x\, p_\xi(x)}$$

Kontinuerlig stokastisk variabel, med frekvensfunktionen $f_\xi(x)$:

$$\boxed{E(\xi) = \int_{-\infty}^{\infty} x\, f_\xi(x)\, dx}$$

Summan i det diskreta fallet blir en integral i det kontinuerliga — samma mönster som för sannolikheterna själva.

---

2. Två tolkningar

Långsiktigt medelvärde

Väntevärdet ska tolkas som det medelvärde vi får efter ett stort antal oberoende observationer av ξ. Gör vi n observationer hamnar värdet x ungefär $n\, p_\xi(x)$ gånger, så medelvärdet blir ungefär

$$\frac{1}{n}\sum_x x \cdot n\, p_\xi(x) = \sum_x x\, p_\xi(x) = E(\xi)$$

Masscentrum

Tänk på sannolikhetsfördelningen som en massfördelning på reella axeln, med massan $p_\xi(x)$ i punkten x. Då är $E(\xi)$ masscentrum, eller tyngdpunkten: punkten där stolpdiagrammet skulle balansera.

Masstolkningen gör det lätt att uppskatta väntevärdet direkt ur en skiss, och för symmetriska fördelningar ligger det i symmetripunkten.

Väntevärdet behöver inte vara ett möjligt värde

En tärning har $E(\xi) = 3{,}5$, men ger aldrig $3{,}5$ prickar. Väntevärdet är inte heller samma sak som det mest sannolika värdet.

---

3. Väntevärdet av en funktion

Ska vi räkna väntevärdet av $g(\xi)$ behöver vi inte först ta fram fördelningen för $g(\xi)$. Det räcker att vikta $g(x)$ med ξ:s sannolikheter:

$$E(g(\xi)) = \sum_x g(x)\, p_\xi(x) \qquad \text{respektive} \qquad E(g(\xi)) = \int_{-\infty}^{\infty} g(x)\, f_\xi(x)\, dx$$

Med $g(x) = x^2$ ger det $E(\xi^2)$, som behövs för variansen. Linjära funktioner, $E(a\xi + b) = aE(\xi) + b$, behandlas i Linjära funktioner av stokastiska variabler.

---

4. Kända fördelningar

Väntevärdena för de vanliga fördelningarna står i formelbladet och behöver inte räknas fram varje gång.

Fördelning	$E(\xi)$
$\mathrm{Bin}(n, p)$	np

Fördelning	$E(\xi)$
$\mathrm{Hyp}(N, n, p)$	np
$\mathrm{Po}(\lambda)$	λ
$\mathrm{ffg}(p)$	$1/p$
rektangelfördelning på $[a, b]$	$(a+b)/2$
$\mathrm{Exp}(\lambda)$	$1/\lambda$
$N(\mu, \sigma)$	μ

5. Lösta lektionsuppgifter

Lektion 5, övning 3 — Vad betyder $\lambda = 4$?

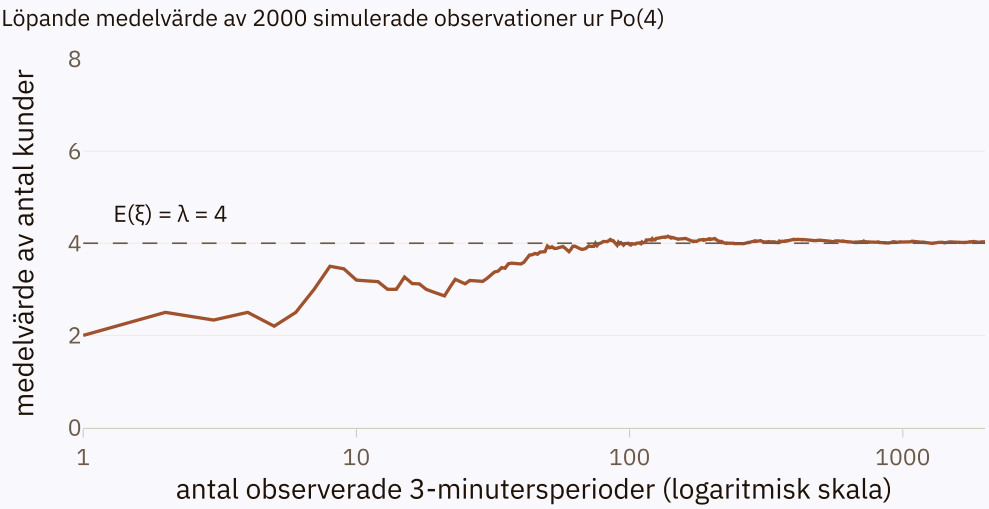
Antalet kunder ξ som anländer till ett systembolag under en 3-minutersperiod är Poissonfördelat med $\lambda = 4$. Parametern anger det *förväntade* antalet kunder per period: $E(\xi) = \lambda = 4$. Hur kan tolkningen "medelvärdet efter ett stort antal oberoende observationer av ξ" motiveras utifrån definitionen av väntevärde? Tolka också väntevärdet som masscentrum.

Lösning

Långsiktigt medelvärde. Observerar vi n oberoende 3-minutersperioder blir värdet x ungefär $n\,p_\xi(x)$ gånger, eftersom den relativa frekvensen närmar sig sannolikheten. Medelvärdet av observationerna blir då ungefär

$$\frac{1}{n}\sum_x x \cdot n\,p_\xi(x) = \sum_x x\,p_\xi(x) = E(\xi) = 4$$

Figuren visar det i en simulering: medelvärdet svänger kraftigt i början men lägger sig vid 4 när antalet perioder växer.



Masscentrum. Lägg massan $p_\xi(x)$ i varje punkt x på tallinjen. Stolpdiagrammet för $\mathrm{Po}(4)$ balanserar i $x = 4$ — till vänster ligger få värden nära, till höger fler värden men längre ut (se figuren i övning 1 under lektion 6 nedan).

Resonemangsuppgift — det finns inget facit att jämföra med.

Lektion 6, övning 1 — Kunder till systembolaget, igen

Antalet kunder som anländer till ett systembolag under en 3-minutersperiod är $\xi \in \mathrm{Po}(4)$.

a) Skissa fördelningens utseende med ett stolpdiagram.

b) Ställ upp ett uttryck för det förväntade antalet kunder under en 3-minutersperiod, $E(\xi)$. Hur ska väntevärdet tolkas?

c) Använd formelsamlingen för att bestämma väntevärdet och variansen.

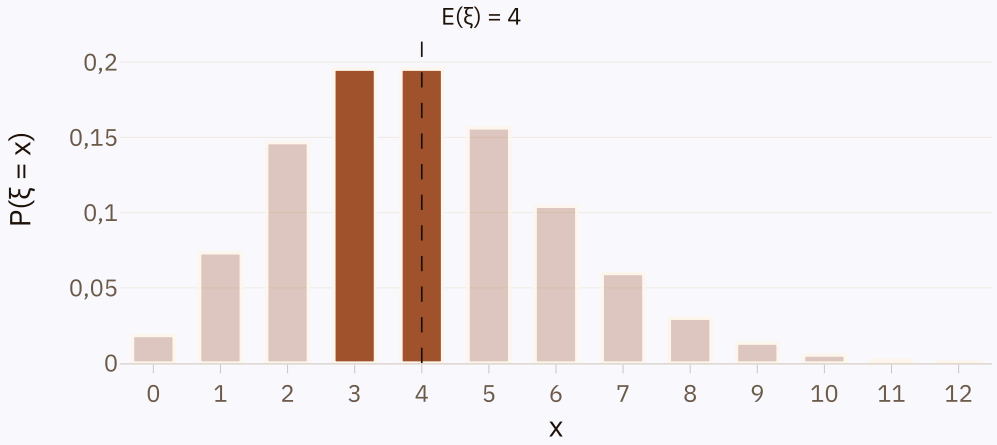
d) Vad innebär varians och standardavvikelse?

e) Bestäm det sannolikaste värdet för ξ och jämför med väntevärdet.

Lösning

a)

Po(4): mest sannolika värden 3 och 4 (markerade)



b) Definitionen med Poissons sannolikhetsfunktion:

$$E(\xi) = \sum_{x=0}^{\infty} x \cdot \frac{4^x}{x!} e^{-4}$$

Tolkning: räknar vi kunder under väldigt många 3-minutersperioder blir medelantalet per period ungefär $E(\xi)$.

c) För $\mathrm{Po}(\lambda)$ är både väntevärde och varians λ:

$$E(\xi) = \boxed{4}, \qquad V(\xi) = \boxed{4}, \qquad \sigma = \sqrt{4} = 2$$

d) Variansen är den förväntade kvadratiska avvikelsen från väntevärdet, $V(\xi) = E((\xi - \mu)^2)$. Standardavvikelsen $\sigma = \sqrt{V(\xi)}$ har samma enhet som ξ (här antal kunder) och mäter hur mycket ξ i genomsnitt avviker från μ.

e) Kvoten mellan två grannar är $\dfrac{p(x)}{p(x-1)} = \dfrac{4}{x}$. Den är större än 1 för $x < 4$ och lika med 1 för $x = 4$, så $p(3) = p(4) \approx 0{,}195$ är de största. De mest sannolika värdena är $\boxed{3 \text{ och } 4}$, och väntevärdet 4 sammanfaller med det ena.

Facit: b) $E(\xi) = \sum_{x=0}^{\infty} x \cdot 4^x e^{-4}/x!$, c) $E(\xi) = 4$, $V(\xi) = 4$, $\sigma = 2$, e) de mest sannolika värdena är 3 och 4.

Se även

- Varians och standardavvikelse
- Sannolikhetsfunktion
- Frekvensfunktion
- Mest sannolika värde
- Linjära funktioner av stokastiska variabler

Resurser

- Khan Academy: Expected value
- Wikipedia: Väntevärde

Varians och standardavvikelse

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #storhet

> **Kurs:** S0008M **Förkunskaper:** Väntevärde

Variansen $V(\xi) = E((\xi - \mu)^2)$ mäter hur mycket en stokastisk variabel i genomsnitt avviker kvadratiskt från sitt väntevärde. Standardavvikelsen $\sigma = \sqrt{V(\xi)}$ är roten ur variansen och har samma enhet som variabeln själv, vilket gör den lättare att tolka.

---

1. Definitionen

Låt $\mu = E(\xi)$ vara väntevärdet. **Variansen** är den förväntade kvadratiska avvikelsen från μ:

$$\boxed{V(\xi) = E((\xi - \mu)^2)}$$

vilket i det diskreta respektive kontinuerliga fallet blir

$$V(\xi) = \sum_x (x-\mu)^2 \, p_\xi(x) \qquad \text{respektive} \qquad V(\xi) = \int_{-\infty}^{\infty} (x-\mu)^2 \, f_\xi(x)\, dx$$

Standardavvikelsen är roten ur variansen:

$$\boxed{\sigma = \sqrt{V(\xi)}}$$

> **Varför roten?**
>
> Kvadreringen gör att positiva och negativa avvikelser inte tar ut varandra, men den ger också variansen fel enhet — kvadratmeter om ξ är en längd i meter. Roten ger tillbaka samma enhet som ξ och μ.

---

2. Räkneregeln

Utvecklar vi kvadraten och använder att väntevärdet är linjärt får vi ett ofta bekvämare uttryck:

$$\boxed{V(\xi) = E(\xi^2) - (E(\xi))^2}$$

där $E(\xi^2) = \sum_x x^2 p_\xi(x)$ eller $\int x^2 f_\xi(x)\,dx$. Den är särskilt praktisk i det kontinuerliga fallet, eftersom vi slipper integrera $(x-\mu)^2$.

---

3. Tolkning

σ som typisk avvikelse

Standardavvikelsen kan tolkas som ett mått på hur mycket ξ i genomsnitt avviker från väntevärdet. För symmetriska fördelningar ligger större delen av sannolikhetsmassan inom $[\mu - 2\sigma,\ \mu + 2\sigma]$.

Två fördelningar kan ha samma väntevärde men helt olika spridning. Variansen säger ingenting om var fördelningen ligger, bara hur utsmetad den är.

---

4. Kända fördelningar

Fördelning	$V(\xi)$
$\mathrm{Bin}(n,p)$	$np(1-p)$
$\mathrm{Hyp}(N,n,p)$	$np(1-p)\dfrac{N-n}{N-1}$
$\mathrm{Po}(\lambda)$	λ
$\mathrm{ffg}(p)$	$(1-p)/p^2$
rektangelfördelning på $[a,b]$	$(b-a)^2/12$
$\mathrm{Exp}(\lambda)$	$1/\lambda^2$
$N(\mu,\sigma)$	σ^2

⚠ $N(\mu, \sigma)$ **har variansen** σ^2

Kursen skriver normalfördelningen med **standardavvikelsen** som andra parameter. Variansen är kvadraten på den.

---

5. Lösta lektionsuppgifter

Lektion 5, övning 6 — Fördelningar med givet μ och σ

a) Låt ξ ha sannolikhetsfördelningen

x	-1	1
$P(\xi = x)$	$0{,}5$	$0{,}5$

Beräkna väntevärde, varians och standardavvikelse.

b) Ge ett exempel på en fördelning med $\mu = 0$ och $\sigma = 10$. Verifiera med definitionerna.

c) Ge ett exempel med $\mu = 0$ och $\sigma = 3{,}7$.

d) Ge ett exempel med $\mu = 1{,}5$ och $\sigma = 3{,}7$.

e) Gör om b) med kravet att minst 3 värden antas med positiv sannolikhet.

f) Ge ett exempel där det mest sannolika värdet är 1 och väntevärdet är större än 100.

Lösning

a)

$$\mu = (-1)\cdot 0{,}5 + 1\cdot 0{,}5 = \boxed{0}, \qquad V(\xi) = (-1-0)^2\cdot 0{,}5 + (1-0)^2\cdot 0{,}5 = \boxed{1}, \qquad \sigma =$$

Idén till b)–e). I a) ligger massan symmetriskt på avståndet 1 från μ, och σ blev just det avståndet. Placerar vi massan $0{,}5$ på avståndet σ på var sin sida om μ får vi väntevärdet μ och standardavvikelsen σ.

b) $P(\xi = -10) = P(\xi = 10) = 0{,}5$. Då är $\mu = 0$ och $V(\xi) = 100 \cdot 0{,}5 + 100 \cdot 0{,}5 = 100$, så $\sigma = 10$.

c) $P(\xi = -3{,}7) = P(\xi = 3{,}7) = 0{,}5$.

d) Flytta c) till $\mu = 1{,}5$: $P(\xi = 1{,}5 - 3{,}7) = P(\xi = -2{,}2) = 0{,}5$ och $P(\xi = 5{,}2) = 0{,}5$. Förflyttningen ändrar väntevärdet men inte spridningen.

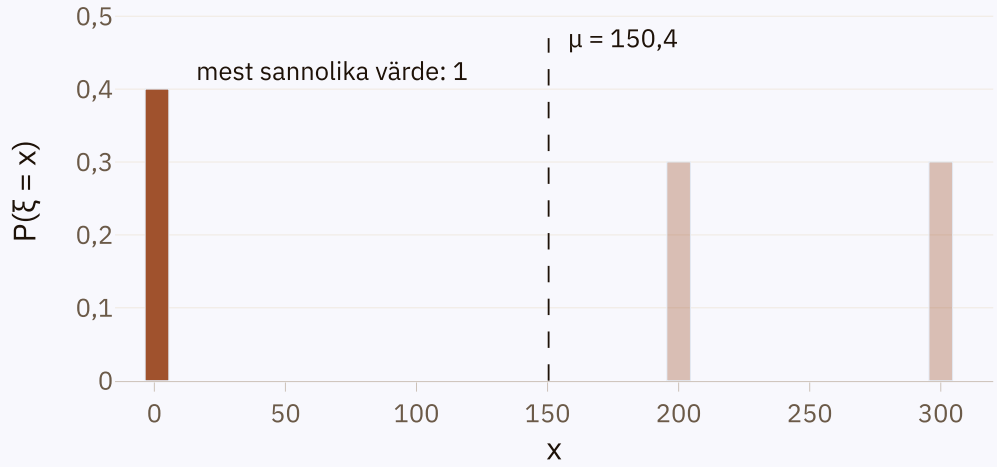
e) Tre värden med sannolikheten $1/3$ var: $-a$, 0 och a. Då är $\mu = 0$ och $V(\xi) = \frac{a^2+0+a^2}{3} = \frac{2a^2}{3}$. Kravet $V(\xi) = 100$ ger $a = \sqrt{150} \approx 12{,}2$.

x	$-\sqrt{150}$	0	$\sqrt{150}$
$P(\xi = x)$	$1/3$	$1/3$	$1/3$

f) Lägg den största enskilda massan i 1, men låt resten ligga långt ut:

x	1	200	300
$P(\xi = x)$	$0{,}4$	$0{,}3$	$0{,}3$

Det mest sannolika värdet är 1 (högsta stolpen), men $\mu = 1 \cdot 0{,}4 + 200 \cdot 0{,}3 + 300 \cdot 0{,}3 = 150{,}4 > 100$.



Facit: a) $\mu = 0$, $\sigma = 1$; b)–f) samma exempel som facit anger. Andra svar på b)–f) är också rätt.

Inför lektion 7, punkt D — Hur mycket ligger inom μ ± 2σ?

För symmetriska fördelningar är en bra tolkning av σ att större delen av sannolikhetsmassan ligger inom $[\mu - 2\sigma,\ \mu + 2\sigma]$. Undersök hur det blir när

a) $\xi \in \text{Bin}(6,\ 0{,}5)$,

b) $\xi \in N(0, 1)$.

Bestäm först sannolikheten att hamna utanför intervallet och utnyttja symmetrin.

Lösning

a) $\mu = np = 3$ och $\sigma = \sqrt{np(1-p)} = \sqrt{1{,}5} \approx 1{,}22$. Intervallet är $[0{,}55,\ 5{,}45]$, så bara $\xi = 0$ och $\xi = 6$ ligger utanför. Med symmetrin:

$$P(\text{utanför}) = P(\xi = 0) + P(\xi = 6) = 2 \cdot 0{,}5^6 = \frac{2}{64} \approx 0{,}031$$

Innanför ligger alltså $\boxed{0{,}969}$ av massan.

b) $\mu = 0$ och $\sigma = 1$. Med normalfördelningens tabell, $\Phi(2) \approx 0{,}9772$:

$$P(\xi < -2) = P(\xi > 2) = 1 - \Phi(2) \approx 0{,}0228$$

Innanför ligger $1 - 2 \cdot 0{,}0228 \approx \boxed{0{,}954}$.

Slutsats. I båda fallen ligger ungefär $95\ \%$ eller mer av massan inom två standardavvikelser från väntevärdet.

På räknaren. a) `binomCdf(6,0.5,1,5)`, b) `normCdf(-2,2)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Lektion 7, övning 6 (Vännman 4.20 b och 4.25 b) — Väntevärde och varians ur en frekvensfunktion

Den stokastiska variabeln ξ har frekvensfunktionen

$$f_\xi(x) = \begin{cases} 0{,}02\,x, & 0 < x < 10 \\ 0 & \text{annars} \end{cases}$$

Bestäm $E(\xi)$, $V(\xi)$ och standardavvikelsen σ.

Lösning

Kontroll av f. $\int_0^{10} 0{,}02\,x\,dx = 0{,}01 \cdot 10^2 = 1$ och $f \geq 0$, så det är en frekvensfunktion.

Väntevärdet — summan från det diskreta fallet blir en integral:

$$E(\xi) = \int_0^{10} x \cdot 0{,}02\,x\,dx = 0{,}02\left[\frac{x^3}{3}\right]_0^{10} = \frac{20}{3} \approx \boxed{6{,}67}$$

Variansen med räkneregeln $V(\xi) = E(\xi^2) - E(\xi)^2$:

$$E(\xi^2) = \int_0^{10} x^2 \cdot 0{,}02\,x\,dx = 0{,}02\left[\frac{x^4}{4}\right]_0^{10} = 50$$

$$V(\xi) = 50 - \left(\frac{20}{3}\right)^2 = \frac{50}{9} \approx \boxed{5{,}56}, \qquad \sigma = \sqrt{\frac{50}{9}} \approx \boxed{2{,}36}$$

Rimlighetskoll. Frekvensfunktionen växer åt höger, så massan ligger tyngre mot 10 och väntevärdet ska ligga över mitten 5. Det gör det.

På räknaren. `∫(x*0.02*x,x,0,10)` och `∫(x^2*0.02*x,x,0,10)-(20/3)^2`.

Uppgiften (Vännman 4.20 b och 4.25 b) är återgiven ur det handskrivna lösningsförslaget ($E = 20/3$, $V = 50/9$, $\sigma \approx 2{,}36$*), och svaren stämmer med det.*

Se även

- Väntevärde
- Linjära funktioner av stokastiska variabler
- Normalfördelning
- Median och kvantiler

Resurser

- Khan Academy: Variance and standard deviation of a discrete random variable
- Wikipedia: Varians

Kontinuerliga stokastiska variabler

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #grundbegrepp

> **Kurs:** S0008M **Förkunskaper:** Stokastisk variabel, Integraler

---

En kontinuerlig stokastisk variabel kan anta alla värden i ett intervall, och sannolikheten för ett enskilt värde är därför noll. I stället beskrivs fördelningen av en frekvensfunktion, där sannolikheten för ett intervall ges av arean under kurvan.

Exempel 1 — väntetid på tunnelbanan

En person anländer till en tunnelbanestation. Tågen går var tionde minut, men personen har varken tidtabell eller klocka. Låt ξ vara väntetiden tills tåget avgår.

Då gäller att $0 \leq \xi < 10$, och alla värden i intervallet $[0, 10)$ kan antas.

---

1. Sannolikheten för ett enskilt värde är noll

För en kontinuerlig stokastisk variabel gäller

$$\boxed{P(\xi = x) = 0 \quad \text{för alla } x}$$

Vi kan alltså inte använda sannolikhetsfunktionen $P(\xi = x)$, som vi använde för att beskriva diskreta fördelningar, se Sannolikhetsfunktion. Det finns oändligt många utfall, och därför går det inte att tilldela varje enskilt utfall ett sannolikhetstal skilt från noll.

---

2. Frekvensfunktionen

I stället integrerar vi en frekvensfunktion f över det intervall vi är intresserade av:

$$P(a \leq \xi \leq b) = \int_a^b f(x)\,dx$$

Följande gäller:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Definitionen, egenskaperna och sambandet med fördelningsfunktionen finns samlade i noten om frekvensfunktionen — där bor också den inramade formeln.

Olikheterna spelar ingen roll

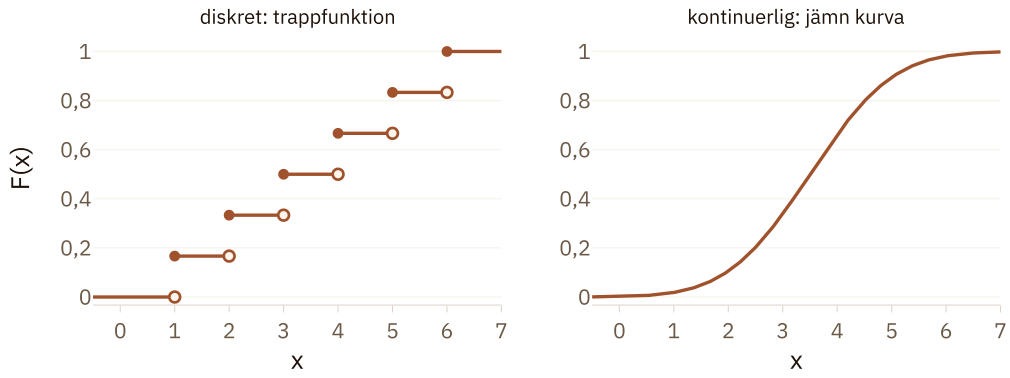
För en kontinuerlig stokastisk variabel kan man slarva med $<$ och $\leq$, eftersom ändpunkterna var för sig har sannolikhet noll:

$$P(a < \xi < b) = P(a \leq \xi < b) = P(a < \xi \leq b) = P(a \leq \xi \leq b)$$

---

3. Fördelningsfunktionen

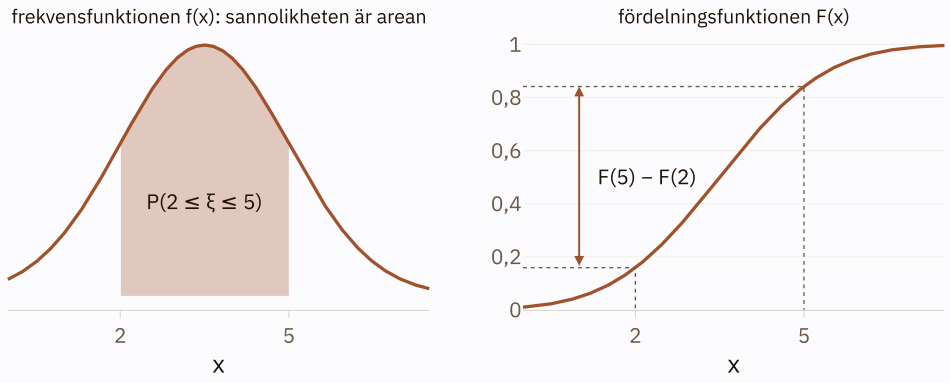
Diskreta fördelningar ger en trappfunktion, medan $F(x)$ för en kontinuerlig fördelning är en jämn kurva.



> “Om man har fördelningsfunktionen är livet underbart” — Adam

Exempel 2 — sannolikhet ur fördelningsfunktionen

Vi har en kontinuerlig fördelning och söker sannolikheten att ξ hamnar mellan 2 och 5.



Lösning

Arean under frekvensfunktionen mellan 2 och 5 är precis det som F hunnit växa på samma sträcka:

$$P(2 \le \xi \le 5) = P(\xi \le 5) - P(\xi \le 2) = \boxed{F(5) - F(2)}$$

4. Tre vanliga kontinuerliga fördelningar

Rektangelfördelning

En fördelning som används när alla värden i ett intervall är lika sannolika.

$$f(x) = \frac{1}{b-a} \quad \text{för } a \le x \le b, \qquad f(x) = 0 \ \text{ för övriga } x$$

Alla delintervall med samma längd är lika sannolika: $P(c \le \xi \le d) = \dfrac{d-c}{b-a}$ för $a \le c \le d \le b$.
Fördelningen kallas också likformig fördelning, och är den kontinuerliga motsvarigheten till den diskreta likformiga fördelningen.

$$E(\xi) = \frac{a+b}{2}, \qquad V(\xi) = \frac{(b-a)^2}{12}$$

Exponentialfördelning

Det som redan har hänt ändrar inte fördelningen för den återstående väntetiden — till exempel vid radioaktivt sönderfall.

$$\xi \in \text{Exp}(\lambda), \qquad \lambda > 0$$

$$P(\xi > t + h \mid \xi > t) = P(\xi > h)$$

$$f(x) = \lambda e^{-\lambda x} \quad \text{för } x \geq 0, \qquad f(x) = 0 \ \text{ för } x < 0$$

$$E(\xi) = \frac{1}{\lambda}, \qquad V(\xi) = \frac{1}{\lambda^2}$$

Mer i Exponentialfördelning.

Normalfördelning

Den mest kända av alla fördelningar, klockformad och symmetrisk kring μ:

$$\xi \in N(\mu, \sigma), \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty$$

Andra parametern är standardavvikelsen σ, inte variansen. Fördelningsfunktionen saknar sluten form och slås upp i tabell; se Normalfördelning.

---

5. Lösta lektionsuppgifter

Lektion 6, övning 8 — Plats för en bil till?

På en gatstump av längden 13 m parkeras slumpmässigt en bil av längden 5 m. Vad är sannolikheten att ytterligare en bil av samma längd får plats? Vilka antaganden gör du?

Lösning

Antaganden. Båda bilarna är exakt 5 m långa, och den första bilen ställs “på måfå”: alla lägen där den ryms är lika sannolika.

Modell. Låt ξ = avståndet från gatstumpens början till bilens bakre ände. Bilen upptar $[\xi,\ \xi + 5]$ och ska rymmas inom $[0,\ 13]$, så $0 \leq \xi \leq 8$. Antagandet ger en rektangelfördelning på $[0,\ 8]$:

$$f(x) = \frac{1}{8}, \quad 0 \leq x \leq 8, \qquad f(x) = 0 \text{ annars}$$

När får en bil till plats? Antingen före den första bilen, vilket kräver $\xi \geq 5$, eller efter den, vilket kräver $13 - (\xi + 5) \geq 5$, alltså $\xi \leq 3$. De två fallen är disjunkta:

$$P(\text{plats}) = P(0 \le \xi \le 3) + P(5 \le \xi \le 8) = \int_0^3 \frac{1}{8}\,dx + \int_5^8 \frac{1}{8}\,dx = \frac{3}{8} + \frac{3}{8} = \boxed{0{,}75}$$

Rimlighetskoll. Den första bilen blockerar bara när den står i mitten, $3 < \xi < 5$, alltså en fjärdedel av det möjliga intervallet.

Uppgiften (Vännman 4.12) är återgiven ur det handskrivna lösningsförslaget, och svaret stämmer med det.

Se även

- Stokastisk variabel
- Frekvensfunktion
- Fördelningsfunktion
- Exponentialfördelning
- Median och kvantiler
- Väntevärde
- Varians och standardavvikelse
- Integraler

Resurser

- Khan Academy: Continuous random variables
- Wikipedia: Kontinuerlig sannolikhetsfördelning

Frekvensfunktion

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Kontinuerliga stokastiska variabler, Fördelningsfunktion

Frekvensfunktionen $f_\xi(x)$ beskriver fördelningen för en kontinuerlig stokastisk variabel. Sannolikheten att variabeln hamnar i ett intervall är arean under frekvensfunktionen över intervallet.

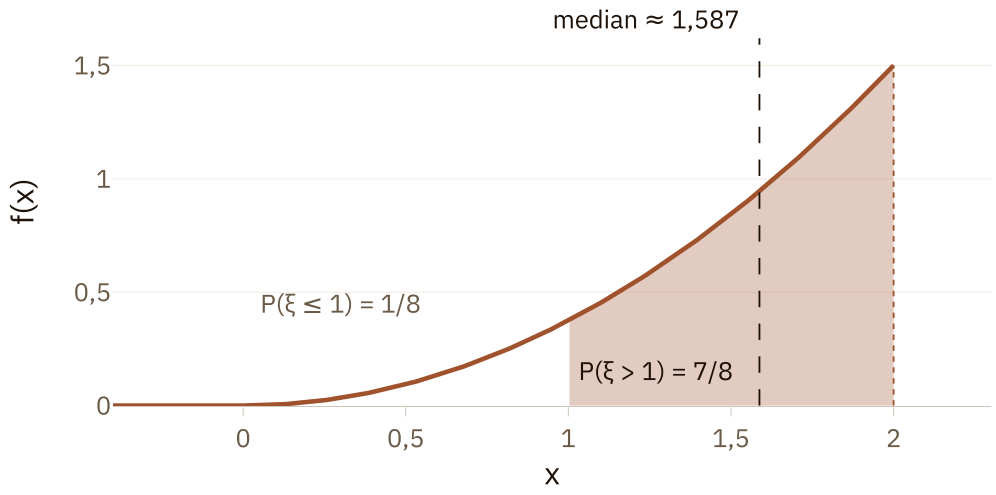
1. Definitionen

En kontinuerlig stokastisk variabel ξ har **frekvensfunktionen** f_ξ om

$$\boxed{P(a \le \xi \le b) = \int_a^b f_\xi(x)\,dx}$$

för alla $a \le b$. I annan litteratur kallas samma funktion **täthetsfunktion**; kursen säger frekvensfunktion.

f(x) = 3x²/8 på [0, 2]: sannolikheten är arean



2. Egenskaper

En funktion f är en frekvensfunktion om och endast om

$$f(x) \geq 0 \quad \text{för alla } x, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Det räcker alltså att kontrollera de två villkoren för att visa att en given funktion är frekvensfunktion för någon stokastisk variabel. De motsvarar exakt villkoren på en sannolikhetsfunktion, med summan utbytt mot en integral.

> ⚠ **$f(x)$ är ingen sannolikhet**
>
> $f(x)$ får vara större än 1 — en rektangelfördelning på $[0,\ 0{,}5]$ har $f(x) = 2$. Det är arean som är en sannolikhet, inte höjden. Däremot är $f(x)\,dx \approx P\,(x < \xi \leq x + dx)$ för en liten intervallängd dx.

---

3. Samband med fördelningsfunktionen

Fördelningsfunktionen är arean till vänster om x, och frekvensfunktionen är dess derivata:

$$\boxed{F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,dt, \qquad f_\xi(x) = F_\xi'(x)}$$

(derivatan gäller i alla punkter där F_ξ är deriverbar). Ur det följer

$$P\,(a < \xi \leq b) = F_\xi(b) - F_\xi(a)$$

och att $P\,(\xi = x) = 0$ för varje enskilt x: ett intervall med längd noll har area noll.

> **Summor blir integraler**
>
> | Diskret | Kontinuerlig |
> |---|---|
> | $p_\xi(x) = P\,(\xi = x)$ | $f_\xi(x)$ |
> | $\sum_x p_\xi(x) = 1$ | $\int f_\xi(x)\,dx = 1$ |
> | $P\,(\xi \in A) = \sum_{x \in A} p_\xi(x)$ | $P\,(\xi \in A) = \int_A f_\xi(x)\,dx$ |
> | $E(\xi) = \sum_x x\,p_\xi(x)$ | $E(\xi) = \int x\,f_\xi(x)\,dx$ |

Exempel — Kontrollera och använd en frekvensfunktion

En stokastisk variabel ξ har frekvensfunktionen $f_\xi(x) = \frac{3}{8}x^2$ för $0 \le x \le 2$ och 0 annars.

a) Visa att f_ξ är en frekvensfunktion.

b) Beräkna $P(\xi > 1)$.

c) Bestäm medianen och väntevärdet.

Lösning

a) $f_\xi(x) \ge 0$ överallt, och

$$\int_0^2 \frac{3}{8}x^2\,dx = \left[\frac{x^3}{8}\right]_0^2 = 1$$

b) Fördelningsfunktionen är $F_\xi(x) = x^3/8$ på $[0, 2]$, så

$$P(\xi > 1) = 1 - F_\xi(1) = 1 - \frac{1}{8} = \boxed{\frac{7}{8}}$$

Intervallet $(1, 2]$ är halva definitionsmängden men bär $7/8$ av sannolikheten, eftersom f_ξ är störst där (figuren ovan).

c) Medianen löser $F_\xi(m) = 0{,}5$:

$$\frac{m^3}{8} = \frac{1}{2} \iff m = \sqrt[3]{4} \approx \boxed{1{,}587}$$

Väntevärdet är

$$E(\xi) = \int_0^2 x \cdot \frac{3}{8}x^2\,dx = \frac{3}{8} \cdot \frac{2^4}{4} = \boxed{1{,}5}$$

Väntevärdet ligger till vänster om medianen: svansen åt vänster drar det nedåt.

På räknaren. Integralmallen i `menu` → Calculus → Integral, t.ex. $\int_1^2 \frac{3}{8}x^2\,dx$ för b). Medianen med `solve(x^3/8=0.5,x)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

⚠ **Vanliga fel**

- **Läsa av höjden som en sannolikhet.** $f_\xi(1{,}5) \approx 0{,}84$ är ingen sannolikhet; bara arean är det.
- **Glömma kontrollen av arean.** Står det "bestäm c så att f blir en frekvensfunktion" är det villkoret $\int f = 1$ som ger c.
- **Integrera över fel område.** Utanför $[0, 2]$ är $f_\xi = 0$; gränserna ska klippas till där f_ξ är positiv.

Se även

- Kontinuerliga stokastiska variabler
- Fördelningsfunktion
- Sannolikhetsfunktion
- Väntevärde
- Median och kvantiler

Resurser

- Khan Academy: Probability density functions
- Wikipedia: Probability density function

Mest sannolika värde

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #storhet

> **Kurs:** S0008M **Förkunskaper:** Sannolikhetsfunktion, Frekvensfunktion

Det mest sannolika värdet är det värde där sannolikhetsfunktionen respektive frekvensfunktionen antar sitt största värde. Det sammanfaller inte nödvändigtvis med väntevärdet — för en sned fördelning skiljer de sig åt.

1. Definitionen

Diskret variabel: det mest sannolika värdet är det x där sannolikhetsfunktionen $p_\xi(x)$ är störst — den **högsta stolpen** i stolpdiagrammet.

Kontinuerlig variabel: det x där frekvensfunktionen $f_\xi(x)$ har sitt maximum — toppen på kurvan.

Det kan finnas flera: $\mathrm{Po}(4)$ har två lika höga stolpar, i 3 och 4.

2. Att hitta det för en diskret fördelning

När sannolikhetsfunktionen är given som formel är det enklast att jämföra grannar. Kvoten

$$\frac{p_\xi(x)}{p_\xi(x-1)}$$

är större än 1 så länge stolparna växer och mindre än 1 när de avtar. Det mest sannolika värdet ligger där kvoten passerar 1. Alternativt räknar vi ut stolparna runt väntevärdet och ser vilken som är högst — det mest sannolika värdet ligger ofta nära väntevärdet för de vanliga fördelningarna.

3. Mest sannolika värde, väntevärde och median

Mått	Är	Syns i figuren som
mest sannolika värde	där sannolikheten är störst	högsta stolpen eller toppen
väntevärde	masscentrum	balanspunkten
median	mittpunkt i sannolikhetsmassa	punkten som delar massan i två lika delar

⚠ Sneda fördelningar drar isär måtten

För en symmetrisk fördelning med en topp sammanfaller alla tre. En lång svans åt höger drar väntevärdet åt höger men flyttar inte toppen. En enstaka tung stolpe långt bort kan till och med göra väntevärdet hur stort som helst medan det mest sannolika värdet står kvar.

---

4. Lösta lektionsuppgifter

Lektion 5, övning 7 — Tretton arbetare med handborrmaskin

Tretton arbetare använder, statistiskt oberoende av varandra, en handborrmaskin $20\ \%$ av arbetstiden. Låt ξ = antalet arbetare som i ett visst ögonblick använder sin handborrmaskin.

a) Bestäm fördelningen för ξ.

b) Bestäm $E(\xi)$ med hjälp av formelbladet.

c) Vilket är det mest sannolika antalet arbetare som använder sin handborrmaskin?

Lösning

a) Tretton oberoende arbetare, var och en med sannolikheten $0{,}2$:

$$p(x) = P(\xi = x) = \binom{13}{x} 0{,}2^{x}\, 0{,}8^{13-x}, \quad x = 0, 1, \ldots, 13, \qquad \text{dvs.} \quad \xi \in \text{Bin}(13,\ 0{,}2)$$

Binomialkoefficienten räknar sätten att välja *vilka* arbetare som använder sin maskin, och potenserna ger sannolikheten för varje sådant val.

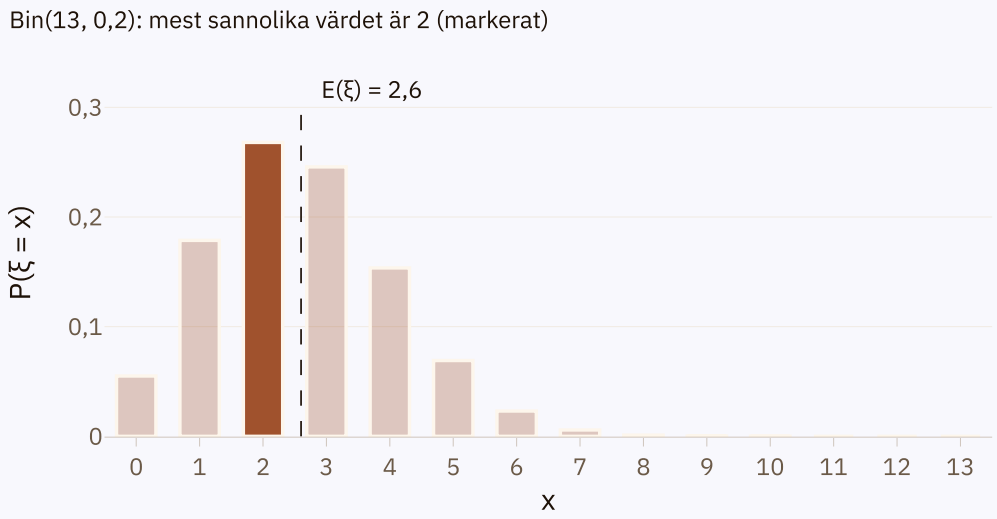
b)

$$E(\xi) = np = 13 \cdot 0{,}2 = \boxed{2{,}6}$$

c) Ur tabellen över fördelningsfunktionen, $p(x) = F(x) - F(x-1)$:

$$p(0) \approx 0{,}0550, \quad p(1) \approx 0{,}1787, \quad p(2) \approx 0{,}2680, \quad p(3) \approx 0{,}2457$$

och stolparna fortsätter att avta för $x \geq 3$. Det mest sannolika antalet är $\boxed{2}$.



Jämförelse. Väntevärdet $2{,}6$ ligger mellan de två högsta stolparna, lite förskjutet åt höger eftersom fördelningen har en svans åt höger.

Uppgiften (Vännman 3.31) är återgiven ur det handskrivna lösningsförslaget, och svaren stämmer med det.

Se även

- Väntevärde
- Median och kvantiler
- Sannolikhetsfunktion

- Frekvensfunktion

Resurser

- Wikipedia: Typvärde

Exponentialfördelning

19 okt. 2026, 6 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Kontinuerliga stokastiska variabler

---

Exponentialfördelningen beskriver väntetiden till nästa händelse i en process där händelser inträffar oberoende av varandra och med konstant intensitet. Den är minneslös: den återstående väntetiden beror inte på hur länge vi redan har väntat.

---

1. Frekvensfunktion och fördelningsfunktion

En kontinuerlig stokastisk variabel ξ är **exponentialfördelad** med parameter $\lambda > 0$, $\xi \in \text{Exp}(\lambda)$, om frekvensfunktionen är

$$\boxed{f_\xi(x) = \lambda e^{-\lambda x}, \qquad x \geq 0}$$

och $f_\xi(x) = 0$ för $x < 0$. Integrerar vi får vi fördelningsfunktionen

$$\boxed{F_\xi(x) = P(\xi \leq x) = 1 - e^{-\lambda x}, \qquad x \geq 0}$$

Den behövs nästan alltid, eftersom sannolikheter för intervall då räknas utan integraler:

$$P(\xi > x) = e^{-\lambda x}, \qquad P(a < \xi \leq b) = e^{-\lambda a} - e^{-\lambda b}$$

---

2. Väntevärde och varians

$$E(\xi) = \frac{1}{\lambda}, \qquad V(\xi) = \frac{1}{\lambda^2}, \qquad \sigma = \frac{1}{\lambda}$$

Härledning av $E(\xi)$ och $V(\xi)$

Partiell integration, där randtermerna försvinner eftersom $x^k e^{-\lambda x} \to 0$ när $x \to \infty$:

$$E(\xi) = \int_0^\infty x\,\lambda e^{-\lambda x}\,dx = [\,-xe^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

$$E(\xi^2) = \int_0^\infty x^2\,\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda}E(\xi) = \frac{2}{\lambda^2}, \qquad V(\xi) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Parametern λ är alltså en **intensitet**, antal händelser per tidsenhet, och $1/\lambda$ är den genomsnittliga väntetiden. En komponent med $\lambda = 0{,}005$ per timme håller i genomsnitt 200 timmar.

Kopplingen till Poissonfördelningen

Inträffar händelser med intensiteten λ per tidsenhet är antalet händelser under en tidsenhet Poissonfördelat, $\mathrm{Po}(\lambda)$, och tiden mellan två händelser $\mathrm{Exp}(\lambda)$. Samma process, två frågor: *hur många* respektive *hur länge*.

---

3. Minneslösheten

För $s, t \geq 0$ gäller

$$P(\xi > s + t \mid \xi > s) = \frac{P(\xi > s + t)}{P(\xi > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(\xi > t)$$

En komponent som redan har fungerat i s timmar har samma chans att klara t timmar till som en helt ny. Det är en rimlig modell för komponenter som inte slits, till exempel elektronik som går sönder av slumpmässiga spänningstoppar, men inte för delar som nöts.

Vanliga fel

- **Parameter eller väntevärde.** $\mathrm{Exp}(\lambda)$ har väntevärdet $1/\lambda$. Står det "medellivslängd 200 timmar" är $\lambda = 1/200$, inte 200.
- **Tro att äldre komponenter är sämre.** Minneslösheten gör $P(\xi > s + t \mid \xi > s) = P(\xi > t)$; ålder spelar ingen roll i modellen.
- **Integrera i onödan.** $P(\xi > x) = e^{-\lambda x}$ direkt ur fördelningsfunktionen.

---

4. Lösta lektionsuppgifter

Lektion 6, övning 6 — Sensorn i en digitalkamera (tenta 2005-10-24)

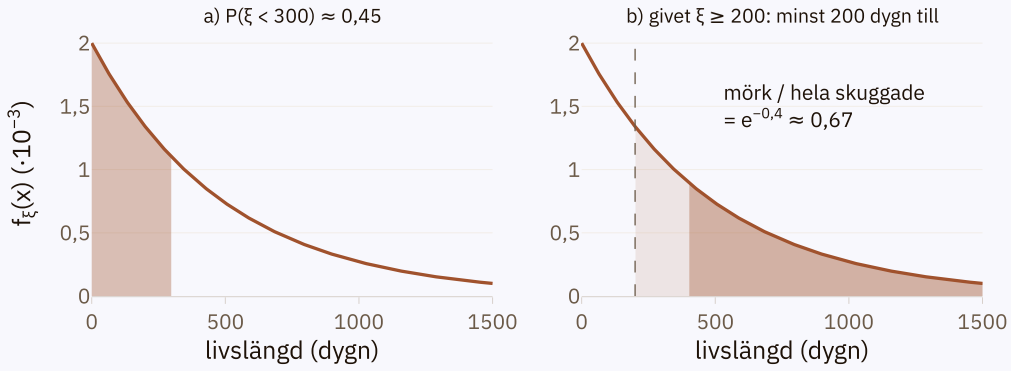
Livslängden ξ för en typ av sensor i en digitalkamera kan beskrivas av en exponentialfördelning med $\lambda = 0{,}002$, enhet dygn.

a) Vad är sannolikheten att livslängden understiger 300 dygn?

b) Givet att en sensor hållit i 200 dygn, vad är sannolikheten att den håller minst 200 dygn till?

Lösning

Modell. $\xi \in \text{Exp}(0{,}002)$, så $F_\xi(x) = 1 - e^{-0{,}002x}$ för $x \geq 0$ och $P(\xi > x) = e^{-0{,}002x}$.



a)

$$P(\xi < 300) = F_\xi(300) = 1 - e^{-0{,}002 \cdot 300} = 1 - e^{-0{,}6} \approx \boxed{0{,}45}$$

b) "Håller minst 200 dygn till" är $\xi \geq 400$, givet $\xi \geq 200$. Eftersom $\{\xi \geq 400\}$ ligger helt inom $\{\xi \geq 200\}$ är snittet bara $\{\xi \geq 400\}$:

$$P(\xi \geq 400 \mid \xi \geq 200) = \frac{P(\xi \geq 400)}{P(\xi \geq 200)} = \frac{e^{-0{,}8}}{e^{-0{,}4}} = e^{-0{,}4} \approx \boxed{0{,}67}$$

Det är samma som $P(\xi \geq 200)$ för en ny sensor — minneslösheten från avsnitt 3. Att sensorn redan hållit 200 dygn säger ingenting om hur länge till den håller.

På räknaren. a) `1-e^(-0.002*300)`, b) `e^(-0.002*200)`.

Facit: a) 0,45, b) 0,67.

Lektion 7, övning 1 — Programstart

Tiden ξ (sekunder) det tar för ett visst program att starta har fördelningsfunktionen

$$F(x) = \begin{cases} 0, & x < 3 \\ 1 - e^{-0{,}2(x-3)}, & x \geq 3 \end{cases}$$

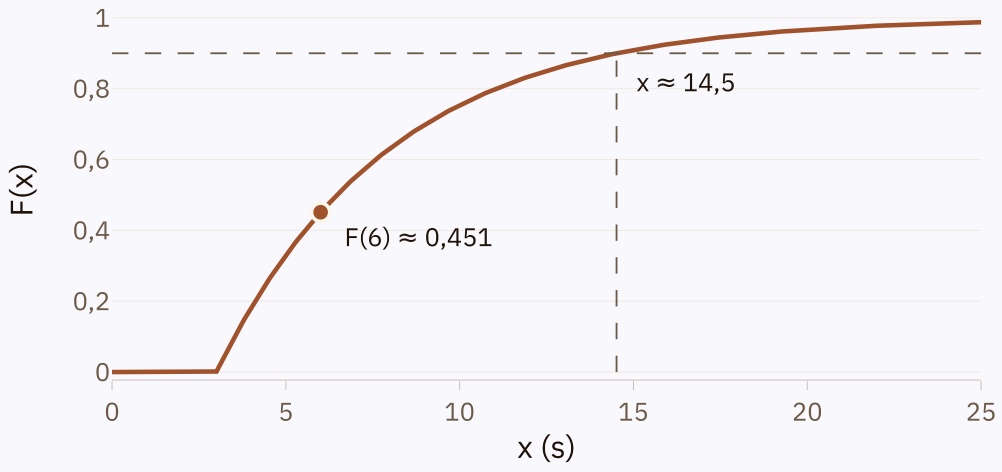
Skissa fördelningsfunktionen och bestäm

a) sannolikheten att det tar mer än 6 sekunder för programmet att starta,

b) den tid x som programmet startar inom med $90\ \%$ sannolikhet.

Lösning

Fördelningen är en exponentialfördelning med $\lambda = 0{,}2$, förskjuten 3 sekunder åt höger: programmet startar aldrig snabbare än på 3 sekunder.



a)

$$P(\xi > 6) = 1 - F(6) = e^{-0{,}2\cdot 3} = e^{-0{,}6} \approx \boxed{0{,}55}$$

b) Vi söker x med $F(x) = 0{,}9$:

$$1 - e^{-0{,}2(x-3)} = 0{,}9 \iff e^{-0{,}2(x-3)} = 0{,}1 \iff x = 3 + \frac{\ln 10}{0{,}2} \approx \boxed{14{,}5\ \text{s}}$$

På räknaren. b) `solve(1-e^(-0.2*(x-3))=0.9,x)`.

Facit: a) 0,55, b) 14,5 s.

Lektion 7, övning 2 — Livslängd hos komponenter

Livslängden ξ (timmar) för en typ av elektroniska komponenter är $\mathrm{Exp}(0{,}005)$. Skissa frekvensfunktionen, markera sannolikheterna nedan i den och beräkna dem genom att utgå från frekvensfunktionen:

a) $P(\xi \leq 150)$,

b) $P(\xi > 300)$,

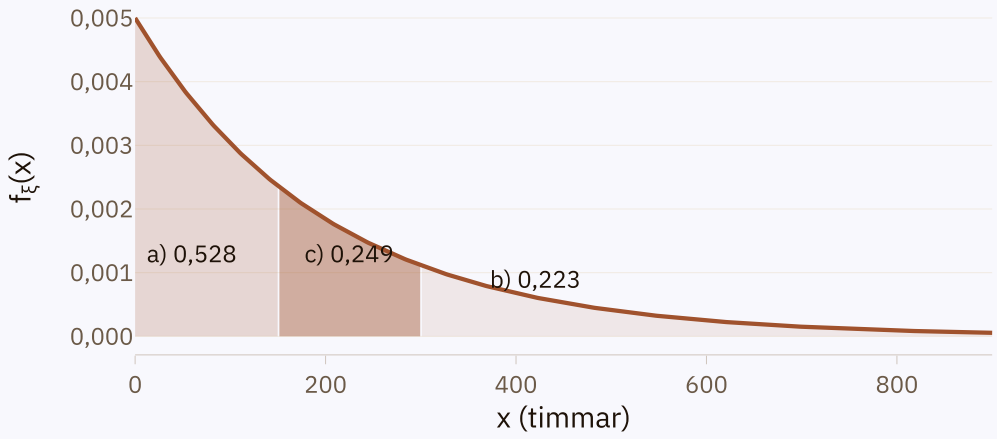
c) $P(150 < \xi \leq 300)$,

d) $P(\xi = 200)$.

Varför blir summan av a), b) och c) lika med 1?

Lösning

$f_\xi(x) = 0{,}005\, e^{-0{,}005x}$ för $x \geq 0$.



a)

$$P(\xi \le 150) = \int_0^{150} 0{,}005\, e^{-0{,}005x}\, dx = 1 - e^{-0{,}75} \approx \boxed{0{,}528}$$

b)

$$P(\xi > 300) = \int_{300}^{\infty} 0{,}005\, e^{-0{,}005x}\, dx = e^{-1{,}5} \approx \boxed{0{,}223}$$

c)

$$P(150 < \xi \le 300) = e^{-0{,}75} - e^{-1{,}5} \approx \boxed{0{,}249}$$

d) För en kontinuerlig variabel har varje enskild punkt sannolikheten noll — integralen över ett intervall av längd noll:

$$P(\xi = 200) = \boxed{0}$$

Summan. a), b) och c) är disjunkta och täcker tillsammans hela $[0, \infty)$, så de måste summera till 1: $0{,}528 + 0{,}223 + 0{,}249 = 1{,}000$.

På räknaren. a) `1-e^(-0.005*150)`; med CAS går det också direkt: `∫(0.005*e^(-0.005*x),x,0,150)`.

Facit: a) 0,528, b) 0,223, c) 0,249, d) 0.

Lektion 7, övning 3 — Fyra av tio komponenter håller

Livslängden för en komponent är som i övning 2, $\text{Exp}(0{,}005)$. Vad är sannolikheten att bland 10 komponenter exakt 4 har en livslängd som överstiger 300 timmar? Händelserna att livslängden överstiger 300 timmar för olika komponenter antas oberoende.

Lösning

Ny stokastisk variabel. Låt η = antal komponenter av de 10 som håller mer än 300 timmar. Varje komponent gör det med sannolikheten $p = P(\xi > 300) = e^{-1{,}5} \approx 0{,}223$, oberoende av de andra, så $\eta \in \text{Bin}(10,\ 0{,}223)$.

$$P(\eta = 4) = \binom{10}{4} 0{,}223^4 \, (1 - 0{,}223)^6 \approx \boxed{0{,}114}$$

Uppgiften kombinerar en kontinuerlig fördelning, som ger p, med en binomialfördelning för antalet.

På räknaren. `binomPdf(10,e^(-1.5),4)`.

Facit: 0,114.

Se även

- Kontinuerliga stokastiska variabler
- Poissonfördelning
- Fördelningsfunktion
- Normalfördelning
- ffg-fördelning

Resurser

- Khan Academy: Continuous random variables
- Wikipedia: Exponentialfördelning

Normalfördelning

19 okt. 2026, 6 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Kontinuerliga stokastiska variabler

---

Normalfördelningen är den klockformade fördelning som bestäms helt av sitt väntevärde μ och sin standardavvikelse σ. Den är den viktigaste fördelningen i kursen, eftersom summor av många oberoende bidrag blir approximativt normalfördelade.

---

1. Frekvensfunktionen

En kontinuerlig stokastisk variabel ξ är **normalfördelad**, $\xi \in N(\mu, \sigma)$, om dess frekvensfunktion är

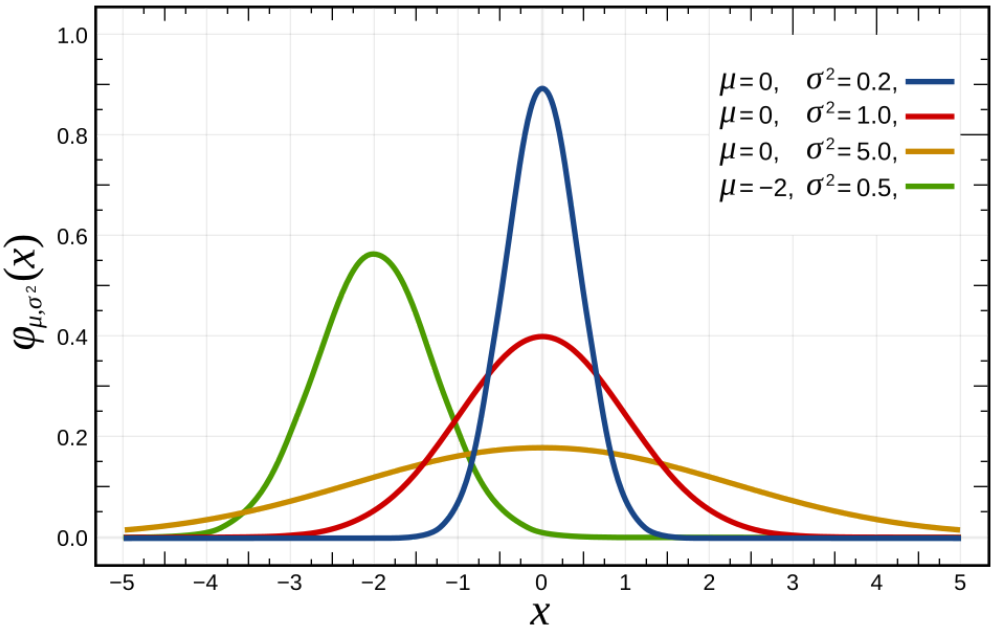
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

> ⚠ **$N(\mu, \sigma)$ hos Vännman, $N(\mu, \sigma^2)$ nästan överallt annars**
>
> Kursen följer Vännman och skriver $\xi \in N(\mu, \sigma)$ med **standardavvikelsen** som andra parameter. De flesta andra böcker, Wikipedia, Python och R skriver $\xi \sim N(\mu, \sigma^2)$, med tilde och **varians**. Det är inte bara skrivsättet som skiljer: $N(170,\ 10)$ här är samma fördelning som $N(170,\ 100)$ där. Kontrollera alltid vilken konvention en källa använder innan du stoppar in siffror.

Två parametrar räcker för att fastställa hela fördelningen, och de gör var sin sak:

- μ — väntevärdet. Flyttar klockan i sidled utan att ändra formen.
- σ — standardavvikelsen. Styr bredden; hög σ ger en låg och bred klocka, låg σ en smal och hög.

Faktorn $1/(\sigma\sqrt{2\pi})$ är bara normeringen som gör att arean under kurvan blir 1. Kurvan är symmetrisk kring μ, så μ är också median och mest sannolika värde.



Normaliserade gaussianer med väntevärde μ och varians σ^2. *Bild från Wikipedia, som använder variansen som parameter.*

---

2. Sannolikheter via standardisering

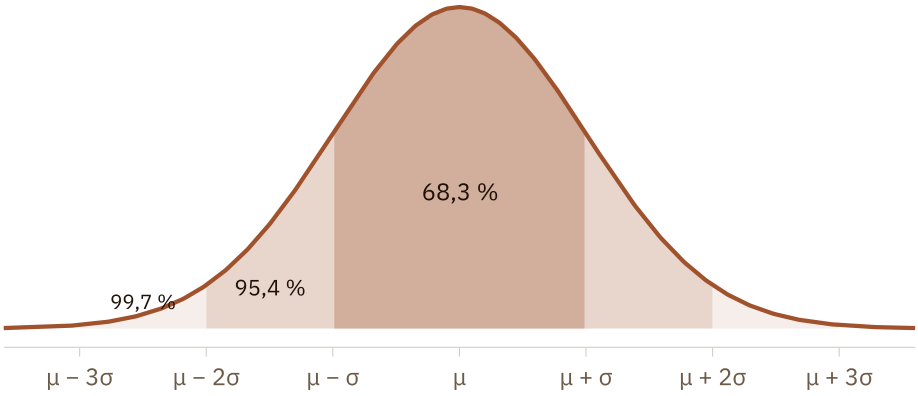
Frekvensfunktionen går inte att integrera i slutna funktioner, så alla sannolikheter slås upp i tabell över $N(0,1)$. **Sats 4.2** säger att om $\xi \in N(\mu,\sigma)$ så är

$$\boxed{P(\xi \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

där Φ är fördelningsfunktionen för den standardiserade normalfördelningen. Vi mäter alltså avståndet från väntevärdet i antal standardavvikelser. Samma sak på räknaren: `normCdf(-∞,x,μ,σ)`.

68–95–99,7

Andelen av massan som ligger inom 1, 2 respektive 3 standardavvikelser från väntevärdet är ungefär $68\ \%$, $95\ \%$ och $99{,}7\ \%$ — oavsett vilka μ och σ som gäller. Det är den snabbaste rimlighetskollen som finns när man räknat fram en normalsannolikhet.



3. Varför den dyker upp överallt

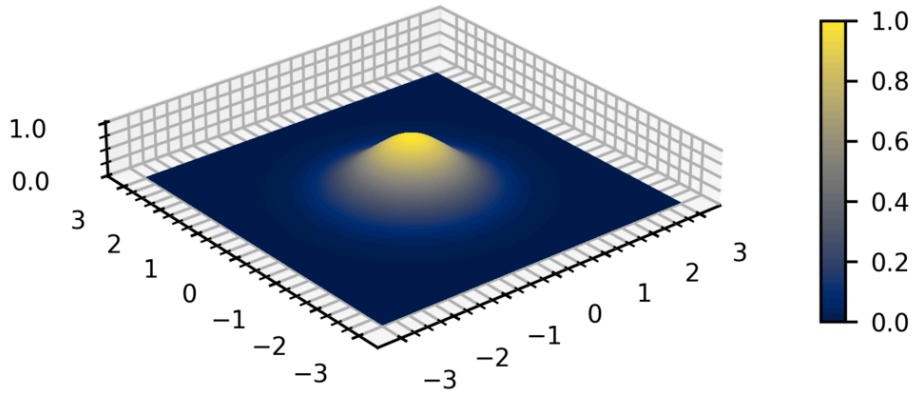
Normalfördelningen är inte vanlig för att naturen råkar tycka om klockkurvor, utan för att den är en **gränsfördelning**: summerar man många oberoende bidrag går resultatet mot en normalfördelning oavsett hur de enskilda bidragen är fördelade. Det är innehållet i centrala gränsvärdessatsen, och det är därför mätfel, längder och brus så ofta modelleras normalt.

Situation	Vad som händer
Summa av många oberoende termer	Går mot normalfördelning — Centrala gränsvärdessatsen
Binomial med $np(1-p) \geq 10$	Approximeras av $N(np,\ \sqrt{np(1-p)})$
Poissonfördelning med stort λ	Approximeras av $N(\lambda,\ \sqrt{\lambda})$
Linjärkombination av oberoende normalfördelade	Är exakt normalfördelad — se Summor av normalfördelade stokastiska variabler

Den bevarar sig själv

Att en summa av normalfördelade variabler *själv* är normalfördelad är ovanligt bland fördelningar och gör räkningarna lätta: det räcker att hålla reda på väntevärde och standardavvikelse, så är hela fördelningen känd.

4. Två dimensioner



Gaussisk funktion över en tvådimensionell domän. *Bild från Wikipedia.*

Med två variabler samtidigt bestäms fördelningen av två väntevärden, två standardavvikelser och korrelationen mellan variablerna — se Tvådimensionell normalfördelning.

---

5. Lösta lektionsuppgifter

Exempel — sannolikheten för en avvikelse uppåt

Längden hos en population är normalfördelad med $\mu = 170$ cm och $\sigma = 10$ cm. Hur stor andel är längre än 185 cm?

Lösning

Setup. Vi söker $P(\xi > 185)$ där $\xi \in N(170,\ 10)$. Standardisera först.

Räkningen.

$$P(\xi > 185) = 1 - \Phi(\frac{185 - 170}{10}) = 1 - \Phi(1{,}5) = 1 - 0{,}9332 = \boxed{0{,}0668}$$

Rimlighetskoll. 185 cm ligger 1,5 standardavvikelser över medelvärdet. Enligt 68–95-regeln ligger 16 % ovanför 1σ och 2,5 % ovanför 2σ, så svaret ska hamna mellan dessa. Knappt 7 % gör det.

På räknaren. `normCdf(185,∞,170,10)`.

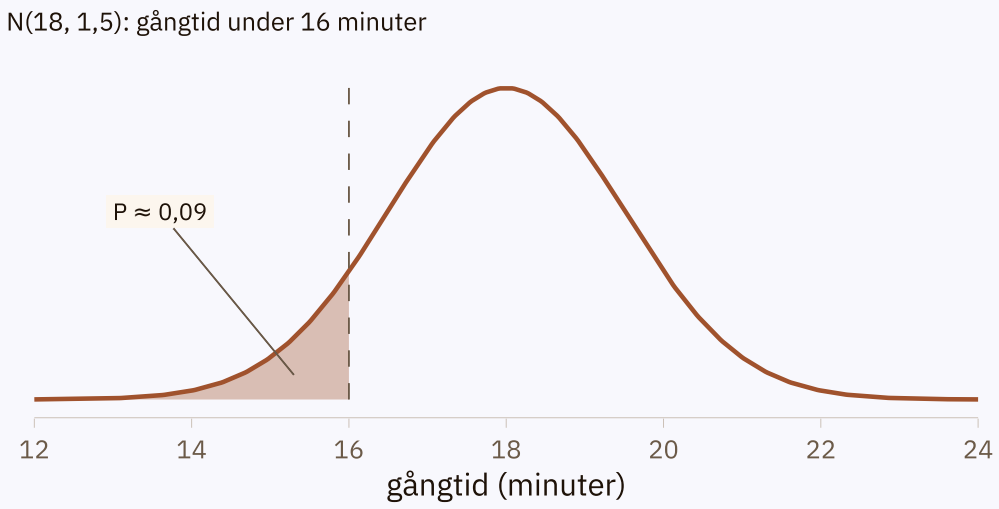
Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Lektion 7, övning 5 — Gångtid till universitetet

Gångtiden ξ (minuter) vintertid mellan hemmet och Luleå tekniska universitet för en viss statistiklärare är normalfördelad med $\mu = 18$ och $\sigma = 1{,}5$. Beräkna sannolikheten att gångtiden en slumpmässigt vald vinterdag understiger 16 minuter.

Lösning

$\xi \in N(18,\ 1{,}5)$. Den sökta sannolikheten är den skuggade svansen:



Sats 4.2:

$$P(\xi < 16) = \Phi\left(\frac{16-18}{1{,}5}\right) = \Phi(-1{,}33) = 1 - \Phi(1{,}33) \approx 1 - 0{,}9082 \approx \boxed{0{,}09}$$

Rimlighetskoll. 16 minuter ligger 1,33 standardavvikelser under medelvärdet — mellan 1σ (16 % under) och 2σ (2,5 % under).

På räknaren. `normCdf(-∞,16,18,1.5)` — bocka i **Draw** så ritas den skuggade svansen.

Facit: 0,09.

Lektion 7, övning 7 (Vännman 4.29) — Symmetri kring μ

Låt $\xi \in N(\mu, \sigma)$. Beräkna

a) $P(\xi < \mu)$,

b) $P(\xi < \mu - \sigma)$,

c) $P(\xi > \mu + \sigma)$,

d) $P(\xi < \mu - 2\sigma)$,

e) $P(\xi > \mu + 2\sigma)$.

Diskutera vad resultaten säger om μ och σ för en symmetrisk fördelning.

Lösning

Standardisera med $\eta = (\xi - \mu)/\sigma \in N(0, 1)$.

a) Kurvan är symmetrisk kring μ, så hälften av massan ligger till vänster:

$$P(\xi < \mu) = \Phi(0) = \boxed{0{,}5}$$

b)

$$P(\xi < \mu - \sigma) = P(\eta < -1) = 1 - \Phi(1) \approx 1 - 0{,}8413 = \boxed{0{,}1587}$$

c) Samma svans på andra sidan:

$$P(\xi > \mu + \sigma) = P(\eta > 1) = \boxed{0{,}1587}$$

d) och **e)** På samma sätt:

$$P(\xi < \mu - 2\sigma) = P(\xi > \mu + 2\sigma) = 1 - \Phi(2) \approx 1 - 0{,}9772 = \boxed{0{,}0228}$$

Tolkning. μ är symmetricentrum och delar massan mitt itu. σ är en naturlig längdskala: ungefär $16\ \%$ av massan ligger mer än en standardavvikelse ut åt varje håll, och drygt $2\ \%$ mer än två.

Uppgiften (Vännman 4.29) är återgiven ur det handskrivna lösningsförslaget, och svaren stämmer med det.

Lektion 8, övning 1 — Inom en och tre standardavvikelser

Antag att $\xi \in N(\mu, \sigma)$. Beräkna

a) $P(\mu - \sigma < \xi < \mu + \sigma)$,

b) $P(\mu - 3\sigma < \xi < \mu + 3\sigma)$.

Tolka i ord betydelsen av σ i normalfördelningen med hjälp av sannolikheterna.

Lösning

Efter standardisering beror svaren inte på μ och σ:

a)

$$P(-1 < \eta < 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 2 \cdot 0{,}8413 - 1 = \boxed{0{,}683}$$

b)

$$P(-3 < \eta < 3) = 2\Phi(3) - 1 \approx 2 \cdot 0{,}99865 - 1 = \boxed{0{,}997}$$

Tolkning. Ungefär två tredjedelar av alla observationer hamnar inom en standardavvikelse från väntevärdet, och praktiskt taget alla inom tre. Ett värde mer än 3σ från μ är så ovanligt att det bör väcka misstanke om att modellen inte stämmer.

På räknaren. `normCdf(-1,1)` och `normCdf(-3,3)`.

Facit: a) 0,683, b) 0,997.

Se även

- Standardiserad normalfördelning
- Kontinuerliga stokastiska variabler
- Linjära funktioner av stokastiska variabler
- Summor av normalfördelade stokastiska variabler
- Centrala gränsvärdessatsen
- Tvådimensionell normalfördelning

Resurser

- 3Blue1Brown: But what is the Central Limit Theorem? — varför just den här kurvan dyker upp överallt
- Khan Academy: Normal distributions
- Wikipedia: Normalfördelning

Standardiserad normalfördelning

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Normalfördelning

Den standardiserade normalfördelningen är $N(0,1)$, normalfördelningen med väntevärde 0 och standardavvikelse 1. Varje normalfördelning kan föras över till den genom att vi drar bort väntevärdet och delar med standardavvikelsen, och därför räcker en enda tabell: $\Phi(x)$ för sannolikheter och λ_α för kvantiler.

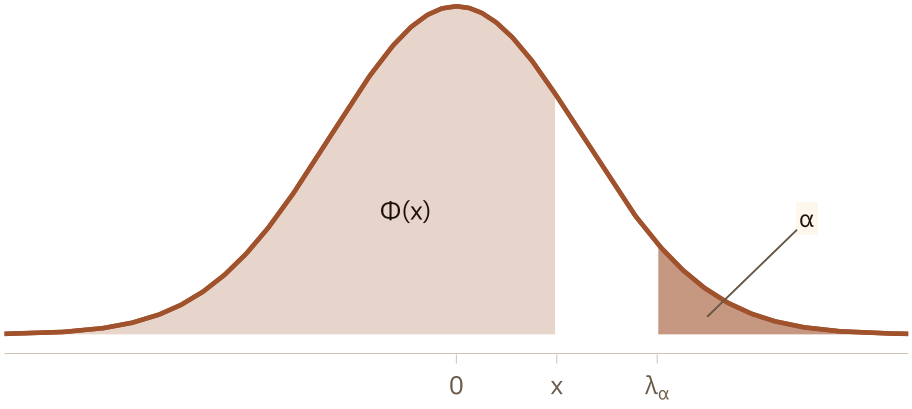
1. $N(0,1)$ och Φ

Är $\xi \in N(0,1)$ blir frekvensfunktionen

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

och fördelningsfunktionen betecknas med ett eget tecken:

$$\boxed{\Phi(x) = P(\xi \le x)}$$



$\Phi(x)$ är den ljusa arean till vänster om x. Kursbokens normalfördelningstabell ger $\Phi(x)$ för $x \ge 0$.

2. Symmetrin

Kurvan är symmetrisk kring 0: arean till vänster om $-x$ är lika stor som arean till höger om x. Alltså

$$\boxed{\Phi(-x) = 1 - \Phi(x)}$$

Därför behöver tabellen bara positiva x. Två följder som används hela tiden:

$$P(\xi > x) = 1 - \Phi(x), \qquad P(-x < \xi < x) = 2\Phi(x) - 1$$

> **Rita alltid figuren**
>
> Skissa klockan, markera gränserna och skugga den sökta arean innan du slår i tabellen. Då syns direkt om svaret ska vara större eller mindre än $0{,}5$, och om symmetrin behövs.

---

3. Standardisering

Sats 6.1. Om $\xi \in N(\mu, \sigma)$ så är

$$\boxed{\frac{\xi - \mu}{\sigma} \in N(0, 1)}$$

Satsen är en omformulering av Sats 4.2 i normalfördelningen: $P(\xi \le x) = P((\xi - \mu)/\sigma \le (x - \mu)/\sigma) = \Phi((x - \mu)/\sigma)$. Att bilda $(\xi - \mu)/\sigma$ kallas att **standardisera** eller normera ξ. Det fungerar eftersom en linjär funktion av en normalfördelad variabel är normalfördelad, och räknereglerna för väntevärde och varians ger väntevärde 0 och standardavvikelse 1.

---

4. Kvantiler: λ_α

Ofta frågar vi baklänges: vilket x har en given sannolikhet ovanför sig? För $\xi \in N(0, 1)$ definieras λ_α av

$$\boxed{P(\xi > \lambda_\alpha) = \alpha}, \qquad \text{alltså} \qquad \Phi(\lambda_\alpha) = 1 - \alpha$$

λ_α är den mörka svansen i figuren ovan. Kursboken har en egen tabell för de vanligaste α, och på miniräknaren ger `invNorm(1-α)` samma sak. I annan litteratur skrivs λ_α ofta z_α.

α	0,10	0,05	0,025	0,01	0,005	0,001
λ_α	1,2816	1,6449	1,9600	2,3263	2,5758	3,0902

På grund av symmetrin är $P(\xi < -\lambda_\alpha) = \alpha$, och $P(-\lambda_{\alpha/2} < \xi < \lambda_{\alpha/2}) = 1 - \alpha$. Det sista är grunden för konfidensintervall.

Vanliga fel

- **Dela med variansen.** Standardiseringen är $(\xi - \mu)/\sigma$ med standardavvikelsen, inte σ^2.
- **Negativa argument.** Tabellen saknar negativa värden; använd $\Phi(-x) = 1 - \Phi(x)$.
- **Blanda ihop Φ och λ_α.** $\Phi(x)$ är sannolikheten *till vänster*; λ_α har sannolikheten α *till höger*, så $\lambda_{0{,}025} = 1{,}96$.
- `normCdf` **med varians.** Räknaren vill ha σ som sista argument, inte σ^2.

---

5. Lösta lektionsuppgifter

Lektion 7, övning 4 — Mätfel

Mätfelet ξ är normalfördelat $N(0, 1)$. Skissa frekvensfunktionen och markera de sökta sannolikheterna. Beräkna

a) $P(\xi > 1{,}75)$,

b) $P(\xi < -1{,}75)$,

c) $P(-0{,}78 < \xi < 2{,}13)$.

Bestäm talet a så att

d) $P(\xi > a) = 0{,}025$,

e) $P(\xi < a) = 0{,}001$.

Lösning

a)

$$P(\xi > 1{,}75) = 1 - \Phi(1{,}75) \approx 1 - 0{,}9599 = \boxed{0{,}040}$$

b) Symmetrin: vänstersvansen under $-1{,}75$ är lika stor som högersvansen över $1{,}75$:

$$P(\xi < -1{,}75) = \Phi(-1{,}75) = 1 - \Phi(1{,}75) \approx \boxed{0{,}040}$$

c)

$$P(-0{,}78 < \xi < 2{,}13) = \Phi(2{,}13) - \Phi(-0{,}78) = \Phi(2{,}13) - (1 - \Phi(0{,}78)) \approx 0{,}9834 - 1 +$$

d) Per definition är $a = \lambda_{0{,}025}$:

$$a = \lambda_{0{,}025} \approx \boxed{1{,}96}$$

e) Svansen till vänster har sannolikheten $0{,}001$, så a är negativt. Symmetrin ger $P(\xi < -\lambda_{0{,}001}) = 0{,}001$:

$$a = -\lambda_{0{,}001} \approx \boxed{-3{,}09}$$

På räknaren. a) `normCdf(1.75,∞)`, c) `normCdf(-0.78,2.13)`, d) `invNorm(0.975)`, e) `invNorm(0.001)`.

Facit: a) 0,040, b) 0,040, c) 0,766, d) 1,96, e) −3,09.

Se även

- Normalfördelning
- Fördelningsfunktion
- Linjära funktioner av stokastiska variabler
- Konfidensintervall
- Hypotesprövning

Resurser

- Khan Academy: Normal distributions
- Wikipedia: Standard normal table

Median och kvantiler

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #storhet

> **Kurs:** S0008M **Förkunskaper:** Fördelningsfunktion, Frekvensfunktion

Medianen är det värde som delar sannolikhetsmassan i två lika stora delar. Mer allmänt anger en kvantil var en viss andel av massan har passerats, och kvartilavståndet mäter hur utspridd den mittersta hälften av fördelningen är.

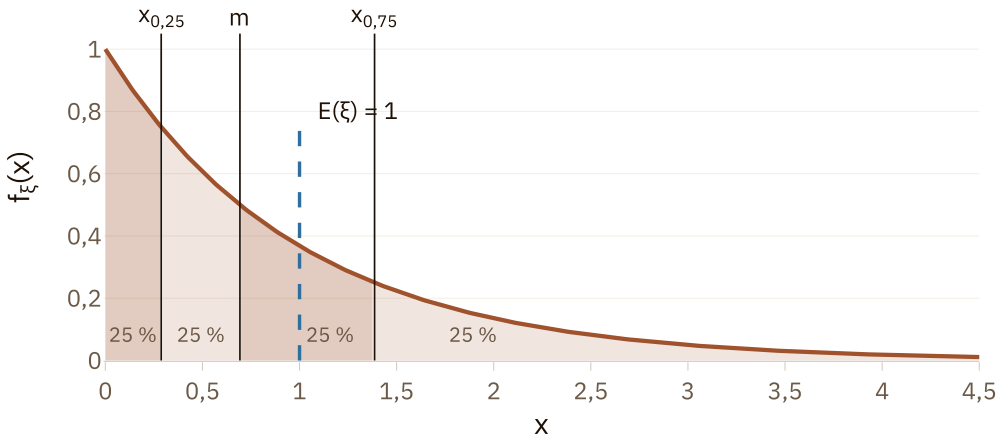
1. Medianen

För en kontinuerlig stokastisk variabel ξ är **medianen** m det värde som uppfyller

$$\boxed{F_\xi(m) = P(\xi \le m) = 0{,}5}$$

Hälften av sannolikhetsmassan ligger till vänster om m och hälften till höger, så $P(\xi < m) = P(\xi > m) = 0{,}5$. I en figur delar medianen arean under frekvensfunktionen i två lika stora delar.

Exp(1): kvartilerna delar arean i fyra delar om 25 % var



För $\mathrm{Exp}(1)$ är medianen $m = \ln 2 \approx 0{,}69$ men väntevärdet 1: den långa svansen åt höger drar väntevärdet med sig.

Medianen är ett **lägesmått**, precis som väntevärdet. För en symmetrisk fördelning sammanfaller de; för en sned fördelning drar en lång svans väntevärdet med sig, men inte medianen.

---

2. Kvantiler

Medianen är ett specialfall av en **kvantil**: det värde x där fördelningsfunktionen har nått en viss nivå. De vanligaste är

Kvantil	Uppfyller
nedre kvartilen	$F_\xi(x) = 0{,}25$
medianen	$F_\xi(x) = 0{,}50$
övre kvartilen	$F_\xi(x) = 0{,}75$

För att hitta en kvantil löser vi ekvationen $F_\xi(x) = q$ för x. På räknaren gör `invNorm(q,μ,σ)` just det för normalfördelningen.

> **Övre kvantiler i inferensdelen**
>
> För $N(0,1)$ och t-fördelningen använder kursen i stället **övre** kvantiler: λ_α är det värde som har sannolikheten α *till höger* om sig, $P(\xi > \lambda_\alpha) = \alpha$. Det är samma sak som att $F(\lambda_\alpha) = 1 - \alpha$.

---

3. Kvartilavståndet

Skillnaden mellan övre och nedre kvartilen kallas **kvartilavståndet**:

$$\boxed{\text{kvartilavstånd} = x_{0{,}75} - x_{0{,}25}}$$

där x_q betecknar lösningen till $F_\xi(x) = q$. Det är ett **spridningsmått**, precis som standardavvikelsen: det mellersta halvan av massan ligger inom ett intervall av den längden.

> **Samma mått som i beskrivande statistik**
>
> Median, kvartiler och kvartilavstånd för en stokastisk variabel är fördelningens motsvarighet till de mått vi räknar på ett datamaterial i laboration 1, med Tukeys definitioner; se Beskrivande statistik. Där räknar vi på observerade värden; här är det egenskaper hos modellen.

Exempel — Median och kvartiler för $f(x) = 2x$

ξ har frekvensfunktionen $f_\xi(x) = 2x$ för $0 \le x \le 1$ och 0 annars. Bestäm medianen, kvartilerna och kvartilavståndet, och jämför medianen med väntevärdet.

Lösning

Fördelningsfunktionen på $[0, 1]$ är

$$F_\xi(x) = \int_0^x 2t\,dt = x^2$$

Varje kvantil löser $x^2 = q$, alltså $x_q = \sqrt{q}$:

$$x_{0,25} = \boxed{0{,}5}, \qquad m = x_{0,5} = \frac{1}{\sqrt{2}} \approx \boxed{0{,}707}, \qquad x_{0,75} = \frac{\sqrt{3}}{2} \approx \boxed{0{,}866}$$

$$\text{kvartilavstånd} = 0{,}866 - 0{,}5 \approx \boxed{0{,}366}$$

Väntevärdet är $E(\xi) = \int_0^1 2x^2\,dx = 2/3 \approx 0{,}667$, alltså lite mindre än medianen. Massan ligger tätast nära 1, och den tunnare svansen mot 0 drar väntevärdet åt vänster.

På räknaren. `solve(x^2=0.75,x)` ger båda rötterna $\pm 0{,}866$; bara den i $[0, 1]$ gäller.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **Lösa $f(x) = 0{,}5$ i stället för $F(x) = 0{,}5$.** Medianen hittas med fördelningsfunktionen, inte frekvensfunktionen.
- **Behålla fel rot.** Ekvationen $F(x) = q$ kan ha lösningar utanför definitionsmängden; välj den där F faktiskt gäller.
- **Övre och nedre kvantil.** $\lambda_{0,05}$ i inferensdelen har $0{,}05$ *till höger*, alltså $F(\lambda_{0,05}) = 0{,}95$.

Se även

- Fördelningsfunktion
- Frekvensfunktion
- Väntevärde
- Varians och standardavvikelse
- Mest sannolika värde
- Teckenintervall

Resurser

- Wikipedia: Kvantil
- Wikipedia: Interquartile range

Linjära funktioner av stokastiska variabler

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Väntevärde, Varians och standardavvikelse, Oberoende stokastiska variabler

---

Väntevärdet är linjärt, så $E(a\xi + b\eta + c) = aE(\xi) + bE(\eta) + c$ gäller alltid. Variansen är det inte — för oberoende variabler gäller $V(a\xi + b\eta) = a^2V(\xi) + b^2V(\eta)$, alltså adderas varianserna men med kvadrerade koefficienter.

---

1. En variabel: $a\xi + b$

För konstanter a och b gäller

$$\boxed{E(a\xi + b) = aE(\xi) + b, \qquad V(a\xi + b) = a^2\,V(\xi)}$$

och därmed $\sigma_{a\xi+b} = |a|\,\sigma_\xi$.

Att lägga till b flyttar hela fördelningen utan att ändra spridningen. Att multiplicera med a skalar avstånden, och eftersom variansen är ett kvadratiskt mått skalas den med a^2.

---

2. Summor: Sats 5.1

För stokastiska variabler $\xi_1, \ldots, \xi_n$ och konstanter $c_1, \ldots, c_n$ gäller

$$\boxed{E\Big(\sum_{i=1}^{n} c_i\xi_i\Big) = \sum_{i=1}^{n} c_i\,E(\xi_i)}$$

alltid, och om variablerna är oberoende

$$\boxed{V\left(\sum_{i=1}^{n} c_i \xi_i\right) = \sum_{i=1}^{n} c_i^2 \, V(\xi_i)}$$

⚠ Varianser adderas — standardavvikelser gör det inte

$\sigma_{\xi_1+\xi_2} = \sqrt{\sigma_1^2 + \sigma_2^2}$, inte $\sigma_1 + \sigma_2$. Och vid en *differens* adderas varianserna också, eftersom $c = -1$ ger $c^2 = 1$:

$$V(\xi_1 - \xi_2) = V(\xi_1) + V(\xi_2)$$

Är variablerna beroende gäller väntevärdesregeln fortfarande, men variansen får en extra term med kovariansen.

---

3. Medelvärdet av oberoende likafördelade

Ett viktigt specialfall: $\xi_1, \ldots, \xi_n$ oberoende och likafördelade med väntevärde μ och standardavvikelse σ. För summan och medelvärdet $\bar{\xi} = \frac{1}{n}\sum \xi_i$ ger Sats 5.1

$$E(\textstyle\sum \xi_i) = n\mu, \quad V(\textstyle\sum \xi_i) = n\sigma^2, \qquad E(\bar{\xi}) = \mu, \quad V(\bar{\xi}) = \dfrac{\sigma^2}{n}$$

Summans standardavvikelse växer bara som $\sqrt{n}$, och medelvärdets krymper som $\sigma/\sqrt{n}$. Det är skälet till att upprepade mätningar lönar sig — se Medelvärdets fördelning.

---

4. Lösta lektionsuppgifter

Lektion 8, övning 4 — Försök i två etapper

Ett vetenskapligt försök består av två etapper, där den andra inte kan påbörjas förrän den första är klar. Låt ξ_1 och ξ_2 vara tiden (minuter) för respektive etapp, och anta att de är oberoende med $E(\xi_1) = 120$, $V(\xi_1) = 36$, $E(\xi_2) = 250$ och $V(\xi_2) = 121$. Bestäm väntevärde och standardavvikelse för den totala tiden $\xi = \xi_1 + \xi_2$. Diskutera hur resultatet kan förändras om variablerna är beroende.

Lösning

Väntevärde. Gäller alltid:

$$E(\xi) = E(\xi_1) + E(\xi_2) = 120 + 250 = \boxed{370 \text{ min}}$$

Varians och standardavvikelse. Oberoendet ger

$$V(\xi) = V(\xi_1) + V(\xi_2) = 36 + 121 = 157, \qquad \sigma_\xi = \sqrt{157} \approx \boxed{12{,}53 \text{ min}}$$

Om de är beroende. Väntevärdet ändras inte. Variansen kan däremot ändras mycket. Två extremfall: om ξ_2 alltid är en exakt kopia av ξ_1 blir $V(\xi_1 + \xi_1) = V(2\xi_1) = 4V(\xi_1)$, och om de tar ut varandra helt, som $\xi_1 - \xi_1 = 0$, blir variansen 0.

Facit: E(ξ) = 370, σ = 12,53. Stämmer med det handskrivna lösningsförslaget.

Lektion 8, övning 5 — Fem byggelement

Vid tillverkning av byggelement varierar längden, och längderna kan betraktas som oberoende likafördelade stokastiska variabler med väntevärde 1 m och standardavvikelse 0,005 m. Vi vill ta ut fem element som ska bli 5 meter sammanlagt.

Metod 1: välj ett element på måfå och slipa till fyra andra så att de blir exakt lika långa som det första.

Metod 2: ta fem element på måfå.

Vilken metod är bäst, i meningen att standardavvikelsen för den totala längden är minst? Slipningen antas exakt. Varför vill vi ha liten standardavvikelse, och är resultatet intuitivt rimligt?

Lösning

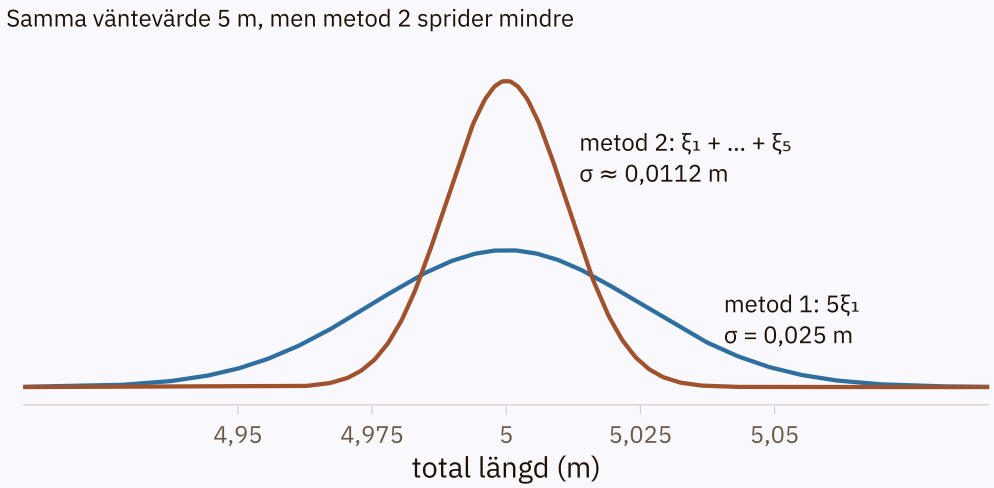
Beteckningar. ξ_k = längden (m) av element nummer k, $k = 1, \dots, 5$, oberoende med $E(\xi_k) = 1$ och $\sigma = 0{,}005$. Låt ξ vara den totala längden.

Metod 1. Alla fem blir lika långa som det första, så $\xi = 5\xi_1$:

$$E(\xi) = 5, \qquad V(\xi) = 25\,V(\xi_1) = 25 \cdot 0{,}005^2 = 0{,}000625, \qquad \sigma_\xi = 5\sigma = 0{,}025 \text{ m}$$

Metod 2. $\xi = \xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5$ med oberoende termer:

$$E(\xi) = 5, \qquad V(\xi) = 5\,V(\xi_1) = 0{,}000125, \qquad \sigma_\xi = \sqrt{5}\,\sigma \approx 0{,}011 \text{ m}$$



Svar. Metod 2 ger minst standardavvikelse. Båda metoderna träffar 5 m i genomsnitt, men liten standardavvikelse betyder att den verkliga längden sällan hamnar långt ifrån.

Rimlighet. Med metod 1 upprepas det första elementets fel fem gånger. Med metod 2 är felen oberoende, så de tar delvis ut varandra: ett för långt element kompenseras ofta av ett för kort.

Facit: metod 2, eftersom $V(\xi_1 + \ldots + \xi_5) = 0{,}000125 < 0{,}000625 = V(5\xi_1)$. Stämmer med det handskrivna lösningsförslaget.

Se även

- Väntevärde
- Varians och standardavvikelse
- Oberoende stokastiska variabler
- Summor av normalfördelade stokastiska variabler
- Medelvärdets fördelning
- Kovarians

Resurser

- Khan Academy: Combining random variables
- Wikipedia: Variance — sum of uncorrelated variables

Oberoende stokastiska variabler

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Oberoende händelser, Stokastisk variabel

---

Stokastiska variabler är oberoende om vetskap om värdet på den ena inte ändrar fördelningen för den andra. Begreppet är samma som för oberoende händelser, men krävs för alla händelser som variablerna kan ge upphov till.

---

1. Definitionen

Två stokastiska variabler ξ och η är **oberoende** om händelserna $\{\xi \le x\}$ och $\{\eta \le y\}$ är oberoende för alla x och y:

$$\boxed{P(\xi \le x,\ \eta \le y) = P(\xi \le x)\,P(\eta \le y)}$$

Kommatecknet betyder "och", alltså snitt. Definitionen utvidgas på samma sätt till $\xi_1, \dots, \xi_n$: sannolikheten att alla ligger under sina gränser är produkten av de enskilda sannolikheterna.

För diskreta variabler är det samma sak som att sannolikhetsfunktionerna multipliceras,

$$P(\xi = x,\ \eta = y) = P(\xi = x)\,P(\eta = y) \quad \text{för alla } x, y$$

och för kontinuerliga att frekvensfunktionerna gör det. Den simultana fördelningen behandlas utförligt i Flerdimensionella fördelningar.

---

2. När antar vi oberoende?

Oberoende är nästan alltid ett **modellantagande**, inte något vi räknar fram. Det är rimligt när variablerna kommer från fysiskt åtskilda försök: två olika tärningskast, mätningar på olika enheter, tiderna för två etapper som inte påverkar varandra. Det är inte rimligt när en gemensam orsak styr båda, till exempel längd och vikt hos samma person.

Oberoende likafördelade

När $\xi_1, \ldots, \xi_n$ är oberoende och har **samma fördelning** kallas de *oberoende likafördelade*. Det är standardmodellen för ett stickprov: n upprepade mätningar av samma storhet under samma förhållanden. Engelska texter förkortar det i.i.d.

---

3. Vad oberoendet ger

Oberoende behövs framför allt för variansen. Väntevärdet av en summa är alltid summan av väntevärdena, men

$$V(\xi_1 + \xi_2) = V(\xi_1) + V(\xi_2) \qquad \text{om } \xi_1, \xi_2 \text{ är oberoende}$$

Utan oberoende tillkommer en term som mäter samvariationen, kovariansen. Se Linjära funktioner av stokastiska variabler.

Samma variabel två gånger är inte två oberoende variabler

$\xi_1 + \xi_1 = 2\xi_1$ har variansen $4V(\xi_1)$, medan $\xi_1 + \xi_2$ med två oberoende kopior har $2V(\xi_1)$. Att skriva $\xi + \xi$ när man menar två oberoende mätningar är ett vanligt fel.

Exempel — Två växlar

Antalet samtal per minut till två växlar är $\xi \in \text{Po}(2)$ och $\eta \in \text{Po}(3)$, oberoende av varandra.

a) Vad är sannolikheten att ingen av växlarna får något samtal under en minut?

b) Beräkna $E(\xi - \eta)$ och $V(\xi - \eta)$.

Lösning

a) Oberoendet låter oss multiplicera:

$$P(\xi = 0,\ \eta = 0) = P(\xi = 0)\,P(\eta = 0) = e^{-2}\,e^{-3} = e^{-5} \approx \boxed{0{,}0067}$$

Det stämmer med att $\xi + \eta \in \mathrm{Po}(5)$, där $P(\xi + \eta = 0) = e^{-5}$ (summan av oberoende Poissonvariabler).

b) Väntevärdet kräver inget oberoende:

$$E(\xi - \eta) = 2 - 3 = \boxed{-1}$$

Variansen gör det, och minustecknet försvinner eftersom $V(-\eta) = (-1)^2 V(\eta)$:

$$V(\xi - \eta) = V(\xi) + V(\eta) = 2 + 3 = \boxed{5}$$

På räknaren. a) `poissPdf(2,0)*poissPdf(3,0)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

⚠ **Vanliga fel**

- **Subtrahera varianser.** $V(\xi - \eta) = V(\xi) + V(\eta)$, aldrig $V(\xi) - V(\eta)$.
- **Addera standardavvikelser.** $D(\xi + \eta) = \sqrt{V(\xi) + V(\eta)}$, inte $D(\xi) + D(\eta)$.
- **Multiplicera utan oberoende.** $P(\xi = x,\ \eta = y) = P(\xi = x)P(\eta = y)$ gäller bara om variablerna är oberoende.

Se även

- Oberoende händelser
- Linjära funktioner av stokastiska variabler
- Flerdimensionella fördelningar
- Kovarians
- Stokastisk variabel

Resurser

- Wikipedia: Independence (probability theory)

Summor av normalfördelade stokastiska variabler

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Linjära funktioner av stokastiska variabler, Normalfördelning

---

En linjärkombination av oberoende normalfördelade stokastiska variabler är själv normalfördelad. Fördelningsformen bevaras alltså, och det räcker att beräkna summans väntevärde och varians för att känna hela dess fördelning.

---

1. Satsen

Låt $\xi_1, \dots, \xi_n$ vara oberoende med $\xi_i \in N(\mu_i, \sigma_i)$, och låt $a_1, \dots, a_n$ och b vara konstanter. Då är

$$a_1\xi_1 + \cdots + a_n\xi_n + b \in N\left(\sum_{i=1}^{n} a_i\mu_i + b,\ \sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}\right)$$

Väntevärdet och variansen är inget nytt — de följer av räknereglerna för linjära funktioner, som gäller för alla fördelningar. Det nya är att summan är **exakt** normalfördelad.

> ⚠ **Varianserna adderas, inte standardavvikelserna**
>
> För $\xi_1 + \xi_2$ är standardavvikelsen $\sqrt{\sigma_1^2 + \sigma_2^2}$, inte $\sigma_1 + \sigma_2$. Och för en **differens** $\xi_1 - \xi_2$ adderas varianserna också, eftersom $a_2^2 = (-1)^2 = 1$:
>
> $$\xi_1 - \xi_2 \in N\left(\mu_1 - \mu_2,\ \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

---

2. Vad som krävs

Antagande som faller	Vad som händer
ξ_i inte normalfördelade	Väntevärde och varians gäller fortfarande, men fördelningen är okänd. Är antalet termer stort ger Centrala gränsvärdessatsen ändå en approximativ normalfördelning.
ξ_i inte oberoende	Väntevärdet gäller fortfarande, men variansen får en extra term med kovariansen.

3. Att räkna sannolikheter

När summans fördelning $N(\mu, \sigma)$ är bestämd räknar vi som vanligt genom att standardisera:

$$P(\xi \leq x) = \Phi(\frac{x - \mu}{\sigma})$$

Söker vi i stället en gräns x med en given sannolikhet går vi baklänges med en kvantil: $P(\xi \leq x) = 1 - \alpha$ ger $x = \mu + \lambda_\alpha \, \sigma$.

Rita figur

Skissa normalkurvan och skugga den sannolikhet som efterfrågas. Det avgör direkt om det ska vara $\Phi(z)$ eller $1 - \Phi(z)$, och om kvantilen ska läggas på ena eller båda sidorna.

Vanliga fel

- **2ξ i stället för $\xi_1 + \xi_2$.** Två *oberoende* exemplar ger $V(\xi_1 + \xi_2) = 2\sigma^2$, men samma variabel två gånger ger $V(2\xi) = 4\sigma^2$. Läs uppgiften: är det samma mätning eller två?
- **Varians som andra parameter.** I kursen betyder $N(\mu, \sigma)$ standardavvikelsen, och `normCdf` på räknaren tar också σ — ta roten ur variansen innan du sätter in den.
- **Oberoende bortglömt.** Variansformeln kräver oberoende; annars tillkommer kovarianstermer, se Kovarians.

4. Lösta lektionsuppgifter

Lektion 9, övning 1–3 — Försök i två etapper

Ett vetenskapligt försök består av två etapper, där den andra inte kan påbörjas förrän den första är avslutad. Låt ξ_1 och ξ_2 vara tiden i minuter för respektive etapp. Antag att ξ_1 och ξ_2 är oberoende och normalfördelade med $E(\xi_1) = 120$, $V(\xi_1) = 36$, $E(\xi_2) = 250$ och $V(\xi_2) = 121$.

Övning 1. Beräkna sannolikheten att den totala tiden $\xi_1 + \xi_2$ överstiger 350 minuter.

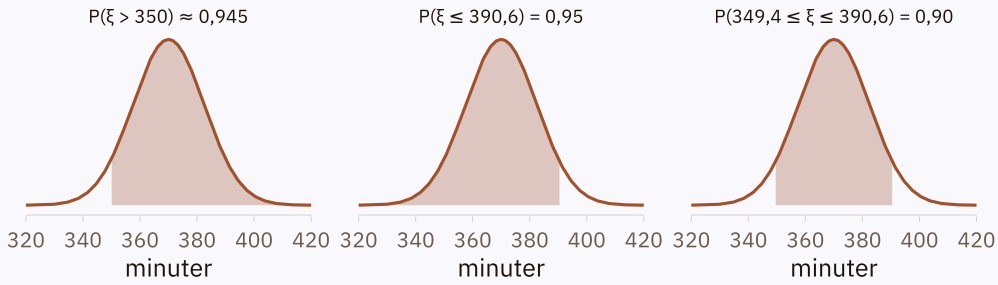
Övning 2. Bestäm en tid inom vilken försöket är avslutat med $95\ \%$ sannolikhet.

Övning 3. Bestäm gränser, symmetriskt kring väntevärdet, mellan vilka den totala tiden ligger med $90\ \%$ sannolikhet.

Lösning

Summans fördelning. Oberoende och normalfördelade, så

$$\xi = \xi_1 + \xi_2 \in N(120 + 250,\ \sqrt{36 + 121}) = N(370,\ \sqrt{157}) \approx N(370;\ 12{,}53)$$



Övning 1.

$$P(\xi > 350) = 1 - \Phi\left(\frac{350 - 370}{12{,}53}\right) = 1 - \Phi(-1{,}60) = \Phi(1{,}60) \approx \boxed{0{,}945}$$

Övning 2. Vi söker t med $P(\xi \le t) = 0{,}95$, alltså $t = 370 + \lambda_{0{,}05} \cdot 12{,}53$ med $\lambda_{0{,}05} = 1{,}6449$:

$$t = 370 + 1{,}6449 \cdot 12{,}53 \approx \boxed{390{,}6 \text{ minuter}}$$

Övning 3. $90\ \%$ i mitten lämnar $5\ \%$ i varje svans, så samma kvantil $\lambda_{0{,}05}$ på båda sidor:

$$370 \pm 1{,}6449 \cdot 12{,}53 \quad \Longrightarrow \quad \boxed{[349{,}4;\ 390{,}6]}$$

Rimlighetskoll. Den övre gränsen i övning 3 är samma som i övning 2: båda lämnar $5\ \%$ ovanför. Och 350 minuter ligger strax över den undre gränsen $349{,}4$, så sannolikheten i övning 1 ska vara strax under $0{,}95$. Den är $0{,}945$.

På räknaren. 1) `normCdf(350,∞,370,√(157))`, 2) `invNorm(0.95,370,√(157))`, 3) `invNorm(0.05,370,√(157))` och samma med `0.95`.

Facit: 1) 0,945, 2) 390,6, 3) [349,4; 390,6].

Lektion 9, övning 4 (Vännman 6.11) — Skruv och mutter

Diametern hos en skruv är $\xi_1 \in N(4{,}10;\ 0{,}16)$ mm och diametern hos hålet i en mutter är $\xi_2 \in N(4{,}35;\ 0{,}12)$ mm, oberoende av varandra. Muttern passar skruven om hålet är större än skruven men högst $0{,}6$ mm större. Vad är sannolikheten att en slumpvis vald mutter passar en slumpvis vald skruv?

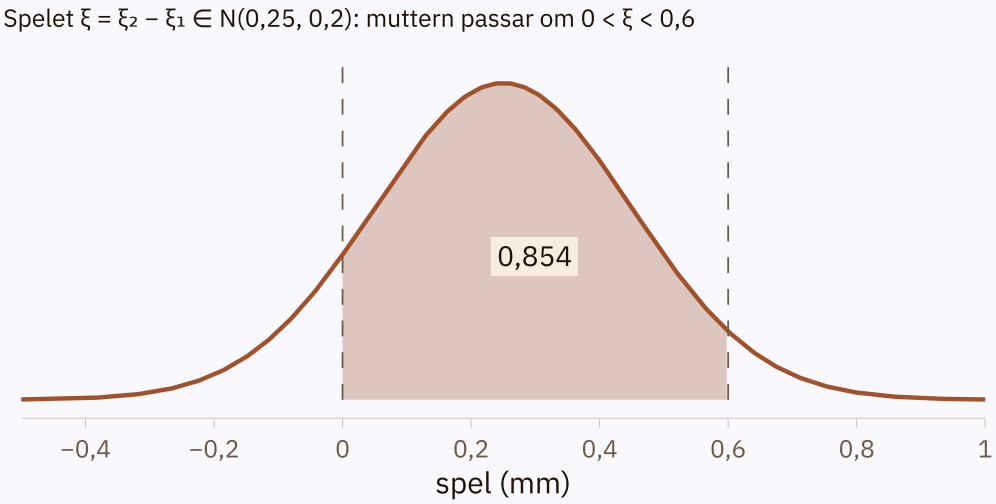
Lösning

Villkoret som olikhet. Med $\xi = \xi_2 - \xi_1$ passar muttern om $0 < \xi < 0{,}6$.

Differensens fördelning. Varianserna adderas:

$$\xi \in N(4{,}35 - 4{,}10,\ \sqrt{0{,}16^2 + 0{,}12^2}) = N(0{,}25;\ 0{,}2)$$

Sannolikheten är arean mellan 0 och $0{,}6$:



$$P(0 < \xi < 0{,}6) = \Phi\left(\frac{0{,}6 - 0{,}25}{0{,}2}\right) - \Phi\left(\frac{0 - 0{,}25}{0{,}2}\right) = \Phi(1{,}75) - \Phi(-1{,}25)$$

$$= \Phi(1{,}75) - (1 - \Phi(1{,}25)) \approx 0{,}9599 - 1 + 0{,}8944 \approx \boxed{0{,}854}$$

På räknaren. `normCdf(0,0.6,0.25,0.2)`.

Uppgiften (Vännman 6.11) är återgiven ur det handskrivna lösningsförslaget, och svaret stämmer med det.

Lektion 10, övning 1 — Majas och Joels utgifter

Majas och Joels utgifter för kursmaterial (kr) under en månad är oberoende stokastiska variabler. Majas utgifter är normalfördelade med väntevärde 420 kr och standardavvikelse 30 kr, Joels med väntevärde 340 kr och standardavvikelse 25 kr.

a) Vad är sannolikheten att deras sammanlagda utgifter en månad överstiger 800 kr?

b) Vad är sannolikheten att Joels utgifter överstiger Majas en slumpmässigt vald månad?

Vilket resultat används, och vilka förutsättningar krävs?

Lösning

Låt $\xi \in N(420,\ 30)$ vara Majas och $\eta \in N(340,\ 25)$ Joels utgifter. Satsen ovan kräver att de är **oberoende** och **normalfördelade** — båda är givna.

a)

$$\xi + \eta \in N(760,\ \sqrt{30^2 + 25^2}) = N(760;\ 39{,}05)$$

$$P(\xi + \eta > 800) = 1 - \Phi\left(\frac{800 - 760}{39{,}05}\right) = 1 - \Phi(1{,}02) \approx \boxed{0{,}154}$$

b) "Joel överstiger Maja" är $\eta - \xi > 0$. Differensen:

$$\eta - \xi \in N(340 - 420,\ \sqrt{25^2 + 30^2}) = N(-80;\ 39{,}05)$$

$$P(\eta - \xi > 0) = 1 - \Phi\left(\frac{0 - (-80)}{39{,}05}\right) = 1 - \Phi(2{,}05) \approx \boxed{0{,}020}$$

Facit: a) 0,1539, b) 0,0202. Facit avrundar z till två decimaler före tabellslagning; utan avrundning blir a) 0,1528.

Se även

- Linjära funktioner av stokastiska variabler
- Normalfördelning
- Standardiserad normalfördelning
- Medelvärdets fördelning
- Centrala gränsvärdessatsen

Resurser

- Wikipedia: Sum of normally distributed random variables

Medelvärdets fördelning

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Summor av normalfördelade stokastiska variabler

Medelvärdet av n oberoende observationer från samma normalfördelning är självt normalfördelat, med samma väntevärde men med standardavvikelsen delad med $\sqrt{n}$. Resultatet är grunden för nästan all inferens om väntevärden.

1. Summa och medelvärde av likafördelade variabler

Låt $\xi_1, \ldots, \xi_n$ vara oberoende och alla $N(\mu, \sigma)$. Med $a_i = 1$ respektive $a_i = 1/n$ i satsen om summor av normalfördelade variabler får vi

$$\sum_{i=1}^{n} \xi_i \in N(n\mu,\ \sigma\sqrt{n}), \qquad \bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i \in N(\mu,\ \frac{\sigma}{\sqrt{n}})$$

2. Kvadratroten

Spridningen växer som $\sqrt{n}$, inte som n

Summan av n oberoende variabler har standardavvikelsen $\sigma\sqrt{n}$, medan n gånger **samma** variabel, $n\xi_1$, har standardavvikelsen $n\sigma$. Oberoende fel tar delvis ut varandra, identiska fel gör det inte.

För medelvärdet betyder det att standardavvikelsen $\sigma/\sqrt{n}$ krymper när n växer — men långsamt: fyra gånger så många mätningar halverar spridningen.

Samma jämförelse gjordes med byggelementen i Linjära funktioner av stokastiska variabler: fem element valda på måfå gav mindre spridning i totallängden än ett element kopierat fem gånger.

3. När ξ_i inte är normalfördelade

Väntevärdet och standardavvikelsen

$$E(\bar{\xi}) = \mu, \qquad V(\bar{\xi}) = \frac{\sigma^2}{n}$$

gäller för **alla** fördelningar, så länge variablerna är oberoende och har samma μ och σ. Normalfördelningen är däremot exakt bara om ξ_i är normalfördelade. Annars ger Centrala gränsvärdessatsen att $\bar{\xi}$ är approximativt normalfördelad när n är stort.

Varför det spelar roll för inferens

$\bar{\xi}$ är den naturliga punktskattningen av μ. Att vi känner dess fördelning, $N(\mu,\ \sigma/\sqrt{n})$, är det som gör det möjligt att bygga konfidensintervall och test för μ.

Vanliga fel

- **Glömma roten.** $D(\bar{\xi}) = \sigma/\sqrt{n}$, inte σ/n.
- **Summa eller medelvärde.** Summan har $D = \sigma\sqrt{n}$, medelvärdet $\sigma/\sqrt{n}$. Läs vilken storhet frågan gäller.
- $\xi_1 + \cdots + \xi_n$ **är inte** $n\xi_1$. n oberoende mätningar har variansen $n\sigma^2$; en mätning gånger n har $n^2\sigma^2$.

4. Lösta lektionsuppgifter

Lektion 9, övning 6 — Störda kunder under 14 dagar

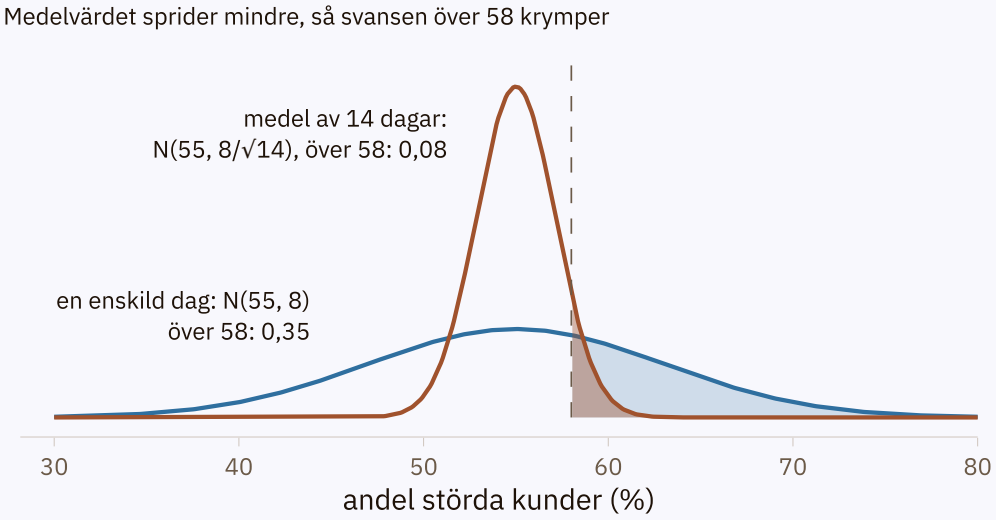
Ett telemarketingföretag har kommit fram till att andelen störda kunder en godtyckligt vald dag kan beskrivas av en normalfördelning med väntevärde 55 procent och standardavvikelse 8 procent. Vad är sannolikheten att den genomsnittliga andelen störda kunder under en 14-dagarsperiod överstiger 58 procent?

Lösning

Låt ξ_i = andelen störda dag i, $i = 1, \ldots, 14$, oberoende och $N(55,\ 8)$. Medelvärdet:

$$\bar{\xi} \in N(55,\ \frac{8}{\sqrt{14}}) = N(55;\ 2{,}138)$$

$$P(\bar{\xi} > 58) = 1 - \Phi(\frac{58-55}{2{,}138}) = 1 - \Phi(1{,}40) \approx \boxed{0{,}081}$$



Jämförelse. En enskild dag är $P(\xi_i > 58) = 1 - \Phi(3/8) \approx 0{,}35$. Genomsnittet över 14 dagar sprider mycket mindre, så det är betydligt ovanligare att det hamnar över 58.

På räknaren. `normCdf(58,∞,55,8/√(14))`.

Facit: 0,081 (tabell med z avrundat; utan avrundning 0,080). Antagandet om oberoende mellan dagarna står inte i uppgiften men krävs.

Lektion 9, övning 9 (Vännman 6.14) — Kex i en förpackning

Tjockleken hos ett kex är $N(0{,}4;\ 0{,}05)$ cm, oberoende mellan kexen. En förpackning rymmer en stapel på 12 cm.

a) Vad är sannolikheten att minst 28 kex får plats?

b) Vad är sannolikheten att minst 30 kex får plats?

Lösning

Låt ξ_k = tjockleken hos kex nummer k, $\xi_k \in N(0{,}4;\ 0{,}05)$, oberoende. Minst n kex får plats om de n första tillsammans är högst 12 cm.

a) $\xi = \xi_1 + \cdots + \xi_{28}$:

$$\xi \in N(28 \cdot 0{,}4,\ 0{,}05\sqrt{28}) = N(11{,}2;\ 0{,}265)$$

$$P(\xi \le 12) = \Phi\left(\frac{12 - 11{,}2}{0{,}265}\right) = \Phi(3{,}02) \approx \boxed{0{,}9987}$$

b) $\xi = \xi_1 + \cdots + \xi_{30} \in N(12;\ 0{,}05\sqrt{30})$, och 12 cm är precis väntevärdet:

$$P(\xi \le 12) = \Phi(0) = \boxed{0{,}5}$$

Uppgiften (Vännman 6.14) är återgiven ur det handskrivna lösningsförslaget, och svaren stämmer med det.

Se även

- Summor av normalfördelade stokastiska variabler
- Centrala gränsvärdessatsen
- Linjära funktioner av stokastiska variabler
- Punktskattning
- Konfidensintervall

Resurser

- Khan Academy: Sampling distribution of the sample mean
- Wikipedia: Standard error

Centrala gränsvärdessatsen

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #sats

> **Kurs:** S0008M **Förkunskaper:** Medelvärdets fördelning

Centrala gränsvärdessatsen säger att summan av många oberoende, likafördelade stokastiska variabler är approximativt normalfördelad, oavsett hur de enskilda variablerna är fördelade. Det förklarar varför normalfördelningen dyker upp så ofta, och gör den användbar även när grundfördelningen är okänd.

---

1. Satsen

Låt $\xi_1, \xi_2, \ldots$ vara oberoende och likafördelade med väntevärde μ och standardavvikelse σ. Då gäller för stora n approximativt

$$\boxed{\sum_{i=1}^{n} \xi_i \approx N(n\mu,\ \sigma\sqrt{n}), \qquad \bar{\xi} \approx N(\mu,\ \frac{\sigma}{\sqrt{n}})}$$

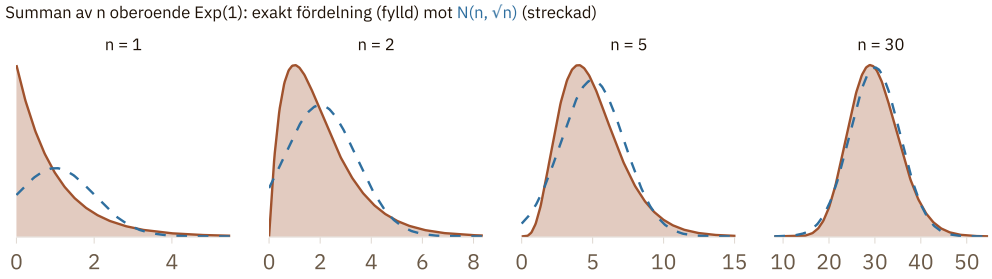
Väntevärdet och standardavvikelsen är exakt desamma som i medelvärdets fördelning — det enda satsen tillför är **formen**: den blir normal, vilken fördelning ξ_i än har.

> **Exakt eller approximativt**
>
> Är ξ_i själva normalfördelade är summan **exakt** normalfördelad för varje n, se Summor av normalfördelade stokastiska variabler. Centrala gränsvärdessatsen behövs när de inte är det, och ger då en approximation som blir bättre ju större n är.

---

2. Hur stort måste n vara?

Det beror på hur olik normalfördelningen grundfördelningen är. Summerar vi exponentialfördelade variabler, som är kraftigt skeva, syns fortfarande skevheten vid $n = 5$, men vid $n = 30$ är summan nästan omöjlig att skilja från normalkurvan:



För en symmetrisk grundfördelning räcker betydligt mindre n.

Känn igen situationen

Signalen i en uppgift är en **summa** eller ett **medelvärde** av **många oberoende** bidrag — antal bilar i 1000 hushåll, totalvikten av 100 förpackningar, positionen efter 600 slumpsteg. Skriv då storheten som $\sum \xi_i$, bestäm $E(\xi_i)$ och $V(\xi_i)$, och använd satsen.

Ett viktigt specialfall är när ξ_i bara antar värdena 0 och 1: då är summan binomialfördelad, och satsen ger normalapproximationen av binomialfördelningen.

Vanliga fel

- **Satsen gäller summan, inte varje term.** De enskilda ξ_i blir inte normalfördelade; det är $\sum \xi_i$ och $\bar{\xi}$ som blir approximativt normalfördelade.
- **Glömma kraven.** Termerna ska vara oberoende och likafördelade med ändlig varians, och n tillräckligt stort.
- **Halvkorrektion för heltalsvariabler.** Approximeras en diskret summa med $N(\mu, \sigma)$ räknas $P(\xi \leq k)$ som $\Phi((k + 0{,}5 - \mu)/\sigma)$; se Normalapproximation.

3. Lösta lektionsuppgifter

Lektion 10, övning 5 — Slumpvandring i 10 minuter

En partikel startar på talet 0 och hoppar varje sekund ett steg, i negativ eller positiv riktning med sannolikheten $0{,}5$ vardera. Beräkna sannolikheten att partikeln efter 10 minuter befinner sig i spannet från -50 till 50.

Lösning

Som summa av steg. Låt ξ_k = hoppet i sekund k, $k = 1, \dots, 600$. Hoppen är oberoende och

$$E(\xi_k) = 1 \cdot 0{,}5 + (-1) \cdot 0{,}5 = 0, \qquad V(\xi_k) = 1^2 \cdot 0{,}5 + (-1)^2 \cdot 0{,}5 - 0^2 = 1$$

Positionen efter 600 sekunder är $\xi = \xi_1 + \cdots + \xi_{600}$. Enligt centrala gränsvärdessatsen är approximativt

$$\xi \approx N(0,\ \sqrt{600}) = N(0;\ 24{,}49)$$

Sannolikheten.

$$P(-50 < \xi < 50) \approx \Phi\left(\frac{50}{24{,}49}\right) - \Phi\left(-\frac{50}{24{,}49}\right) = 2\Phi(2{,}04) - 1 \approx \boxed{0{,}959}$$

Alternativ väg. Låt η = antal steg i positiv riktning, $\eta \in \text{Bin}(600,\ 0{,}5)$, så att $\xi = \eta - (600 - \eta) = 2\eta - 600$. Då är $E(\xi) = 2 \cdot 300 - 600 = 0$ och $V(\xi) = 4 \cdot 150 = 600$, samma fördelning som ovan.

På räknaren. `normCdf(-50,50,0,√(600))`. Den exakta sannolikheten går också att få, eftersom $-50 < \xi < 50$ betyder $276 \le \eta \le 324$: `binomCdf(600,0.5,276,324)` ger $0{,}954$ — approximationen är god.

Facit: 0,9586. Stämmer med det handskrivna lösningsförslaget.

Se även

- Medelvärdets fördelning
- Summor av normalfördelade stokastiska variabler
- Normalapproximation
- Normalfördelning
- Statistisk inferens

Resurser

- 3Blue1Brown: But what is the Central Limit Theorem? — visar hur summor av vilken fördelning som helst blir klockformade
- Wikipedia: Centrala gränsvärdessatsen

Normalapproximation

19 okt. 2026, 3 min läsning

UTKAST

#matematik #sannolikhetslära #approximation

> **Kurs:** S0008M **Förkunskaper:** Centrala gränsvärdessatsen, Binomialfördelning

Diskreta fördelningar som binomial- och Poissonfördelningen kan approximeras med normalfördelningen när väntevärdet och spridningen är tillräckligt stora. Approximationen följer av centrala gränsvärdessatsen och kan förbättras med halvkorrektion.

1. Binomialfördelningen

En binomialfördelad variabel $\xi \in \mathrm{Bin}(n, p)$ är en summa av n oberoende variabler som var och en är 1 med sannolikheten p och 0 annars. Centrala gränsvärdessatsen ger därför, med samma väntevärde och standardavvikelse,

$$\boxed{\xi \approx N(np,\ \sqrt{np(1-p)}) \quad \text{om } np(1-p) \geq 10}$$

Villkoret säger att variansen ska vara stor nog. Är p nära 0 eller 1 är fördelningen skev, och då krävs ett större n.

2. Poissonfördelningen

En Poissonfördelad variabel $\xi \in \mathrm{Po}(\lambda)$ har $E(\xi) = V(\xi) = \lambda$, och för stora λ gäller

$$\xi \approx N(\lambda,\ \sqrt{\lambda})$$

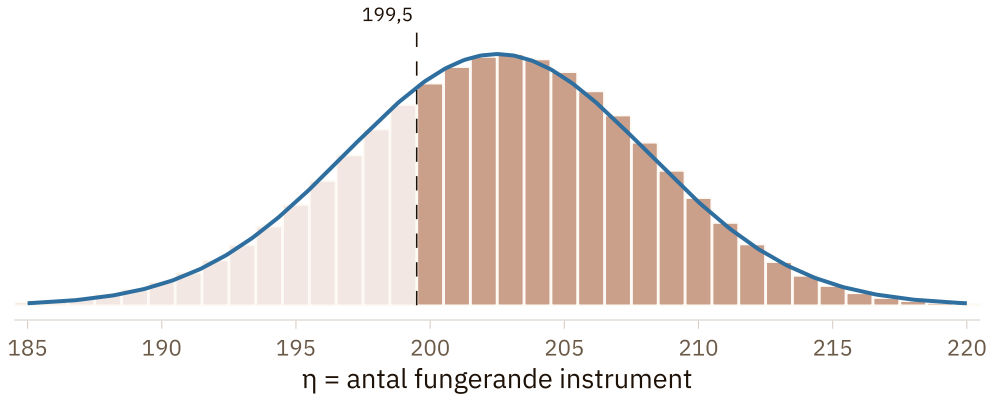
Som tumregel används approximationen när $\lambda \geq 15$; kontrollera gränsen mot kursens formelblad.

3. Halvkorrektion

En diskret variabel lägger all sin massa på heltalen, medan normalkurvan sprider den kontinuerligt. Stapeln för värdet k motsvaras av ytan under kurvan från $k - 0{,}5$ till $k + 0{,}5$. Därför blir approximationen bättre om vi flyttar gränsen ett halvt steg:

$$P(\xi \leq k) \approx \Phi\left(\frac{k + 0{,}5 - \mu}{\sigma}\right)$$

Bin(240, 0,84375) (staplar) och N(202,5, 5,625): η ≥ 200 räknas från 199,5



⚠ Samma händelse, olika svar utan halvkorrektion

För heltalsvärda ξ är $P(\xi \geq 200)$ och $1 - P(\xi \leq 199)$ samma sak. Utan halvkorrektion ger normalapproximationen ändå olika svar beroende på om gränsen sätts vid 200 eller 199. Halvkorrektionen lägger gränsen mitt emellan, vid $199{,}5$.

I kursens facit används **inte** halvkorrektion, om inte annat anges.

---

4. Lösta lektionsuppgifter

Lektion 10, övning 7 (Vännman 6.37) — Instrument i en förpackning

Ett instrument består av tre komponenter som var och en är defekt med sannolikheten $0{,}25$, oberoende av varandra. Instrumentet fungerar om högst en av komponenterna är defekt. En förpackning innehåller 240 instrument, oberoende av varandra. Vad är sannolikheten att minst 200 av dem fungerar?

Lösning

Ett instrument. Låt ξ_k = antal defekta komponenter i instrument k, $\xi_k \in \text{Bin}(3;\ 0{,}25)$.

$$P(\text{instrument } k \text{ fungerar}) = P(\xi_k \le 1) = 0{,}75^3 + 3 \cdot 0{,}25 \cdot 0{,}75^2 = 0{,}84375$$

Förpackningen. Låt η = antal fungerande instrument, $\eta \in \text{Bin}(240;\ 0{,}84375)$, med

$$E(\eta) = 240 \cdot 0{,}84375 = 202{,}5, \qquad V(\eta) = 202{,}5 \cdot 0{,}15625 \approx 31{,}64, \qquad \sigma \approx 5{,}625$$

Eftersom $np(1-p) \approx 31{,}6 \ge 10$ är approximativt $\eta \approx N(202{,}5;\ 5{,}625)$.

Sannolikheten.

$$P(\eta \ge 200) \approx 1 - \Phi\left(\frac{200 - 202{,}5}{5{,}625}\right) = \Phi(0{,}444) \approx \boxed{0{,}67}$$

Känsligheten för gränsen. Skriver vi i stället $P(\eta \ge 200) = 1 - P(\eta \le 199)$ ger samma approximation $\Phi(0{,}622) \approx 0{,}73$. Med halvkorrektion, gränsen vid $199{,}5$, blir det $\Phi(0{,}533) \approx 0{,}703$. Den exakta binomialsannolikheten är $0{,}708$ — halvkorrektionen hamnar klart närmast.

På räknaren. Det exakta värdet: `binomCdf(240,0.84375,200,240)`.

Uppgiften (Vännman 6.37, i 2:a uppl. 6.36) är återgiven ur det handskrivna lösningsförslaget, och alla fyra värden stämmer med det.

Se även

- Centrala gränsvärdessatsen
- Binomialfördelning
- Poissonfördelning
- Normalfördelning
- Standardiserad normalfördelning

Resurser

- Khan Academy: Normal approximation to binomial
- Wikipedia: Continuity correction

Statistisk inferens

19 okt. 2026, 3 min läsning

UTKAST

#matematik #statistik #grundbegrepp

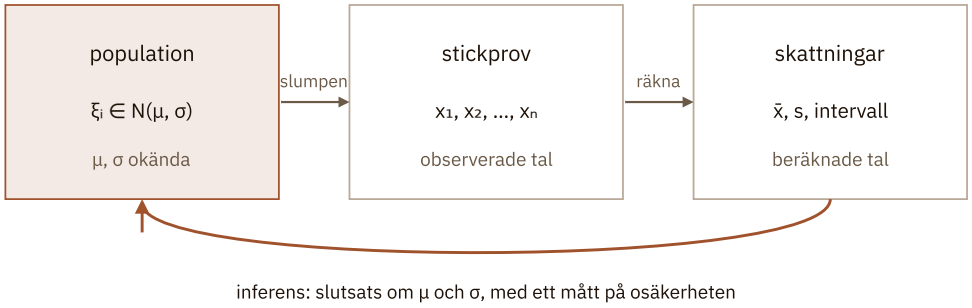
> **Kurs:** S0008M **Förkunskaper:** Centrala gränsvärdessatsen

Statistisk inferens är att dra slutsatser om en okänd population utifrån ett stickprov. Eftersom stickprovet är slumpmässigt blir slutsatsen aldrig säker, utan måste följas av ett mått på osäkerheten — ett konfidensintervall eller en signifikansnivå.

1. Från sannolikhetslära till statistik

I sannolikhetsläran känner vi fördelningen och räknar ut hur sannolika olika utfall är. I statistiken är det tvärtom: vi har **observerat** utfallen och vill säga något om den okända fördelningen som gav dem.

	Känt	**Okänt**
Sannolikhetslära	fördelningen, till exempel $N(\mu, \sigma)$ med givna μ, σ	vad nästa observation blir
Statistisk inferens	observationerna $x_1, \ldots, x_n$	parametrarna $\mu, \sigma, p, \ldots$



Sannolikhetsläran går åt höger i figuren: från en känd modell till vad stickprovet kan bli. Inferensen går tillbaka: från det observerade stickprovet till vad modellens parametrar kan vara.

2. Modell, stickprov och observationer

En inferensuppgift börjar alltid med en **stokastisk modell**:

- Mätvärdena $x_1, \ldots, x_n$ ses som observationer av stokastiska variabler $\xi_1, \ldots, \xi_n$.
- Variablerna antas **oberoende** och **likafördelade**, med en fördelning som beror på en eller flera okända **parametrar** — typiskt $\xi_i \in N(\mu, \sigma)$.

Samlingen $\xi_1, \ldots, \xi_n$ kallas ett **stickprov**, och $x_1, \ldots, x_n$ är det **observerade** stickprovet.

> **Skillnaden mellan ξ_i och x_i**
>
> ξ_i är slumpmässig och har en fördelning; x_i är ett tal. Alla resonemang om *varför* en metod fungerar görs med ξ_i (till exempel att $\bar{\xi} \in N(\mu, \sigma/\sqrt{n})$). När resonemanget är klart ersätter vi ξ_i med de observerade x_i och får ett tal eller ett intervall.

> **Ange modellantagandena först**
>
> Lektionsbladen och tentorna kräver att modellen skrivs ut innan något räknas: vad ξ_i betyder, vilken fördelning den har, vilka parametrar som är okända och att observationerna är oberoende. Det är en del av lösningen, inte en formalitet.

---

3. Tre sätt att dra slutsatser

Metod	Svarar på	Ger
Punktskattning	Vilket värde har parametern ungefär?	ett tal, till exempel $\bar{x}$
Konfidensintervall	Inom vilka gränser ligger parametern, med given säkerhet?	ett intervall med konfidensgrad $1 - \alpha$
Hypotesprövning	Kan ett påstående om parametern förkastas?	ett beslut med felrisk α

Alla tre bygger på samma sak: fördelningen för en stickprovsvariabel, oftast medelvärdet $\bar{\xi}$. Den fås ur satserna om summor eller, när observationerna inte är normalfördelade, ur centrala gränsvärdessatsen.

Exempel — Påsfyllning: modell och punktskattning

En maskin fyller påsar. Vikten varierar slumpmässigt med standardavvikelsen 4 g, känd sedan länge, men medelvikten kan ha drivit. Nio påsar vägs och medelvikten blir $\bar{x} = 502{,}9$ g. Ställ upp en modell och skatta medelvikten med medelfel.

Lösning

Modell. ξ_i = vikten hos påse i, $\xi_i \in N(\mu,\ 4)$ oberoende, $i = 1, \ldots, 9$. μ är okänd, $\sigma = 4$ g känd.

Punktskattning. $\mu^* = \bar{\xi} \in N(\mu,\ 4/\sqrt{9})$, så

$$\mu^*_{\text{obs}} = \boxed{502{,}9 \text{ g}}, \qquad D(\mu^*) = \frac{4}{\sqrt{9}} \approx \boxed{1{,}33 \text{ g}}$$

Samma påsar används i Hypotesprövning, Signifikansnivå och styrka och Konfidensintervall och hypotesprövning, där frågan är om μ fortfarande är 500 g.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **Hoppa över modellen.** Utan en utskriven modell går det inte att avgöra om metoden är rätt, och på tentan ger det avdrag.
- **Behandla $\bar{x}$ som μ.** $\bar{x}$ är en skattning med ett medelfel; μ är fortfarande okänd.
- **Anta oberoende utan att tänka.** Två mätningar på samma person är inte oberoende (se Stickprov i par).

Se även

- Punktskattning
- Konfidensintervall
- Teckenintervall
- Hypotesprövning
- Centrala gränsvärdessatsen

Resurser

- Khan Academy: Confidence intervals
- Wikipedia: Statistical inference

Punktskattning

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Statistisk inferens

En punktskattning är ett enda tal, beräknat ur stickprovet, som utgör vår gissning på en okänd parameter. Skattningen är väntevärdesriktig om dess väntevärde är lika med den sanna parametern, alltså om den varken systematiskt över- eller underskattar.

---

1. Skattning och skattningsvariabel

Låt θ vara en okänd parameter, till exempel μ. En **skattning** av θ är en funktion av stickprovet. Vännman markerar den med stjärna:

- $\theta^* = g(\xi_1, \ldots, \xi_n)$ är **stickprovsvariabeln** — slumpmässig, med en fördelning.
- $\theta^*_{\text{obs}} = g(x_1, \ldots, x_n)$ är den **observerade skattningen** — ett tal.

Eftersom θ^* är en stokastisk variabel kan vi fråga efter dess väntevärde och varians, och det är så vi bedömer hur bra skattningen är.

---

2. Två kvalitetskrav

Väntevärdesriktig och effektiv

- θ^* är **väntevärdesriktig** om $E(\theta^*) = \theta$. Skattningen träffar rätt i genomsnitt.
- Av två väntevärdesriktiga skattningar är den med **minst varians** mest **effektiv**. Den hamnar i genomsnitt närmast θ.

Spridningen hos skattningen kallas dess **medelfel**, $D(\theta^*) = \sqrt{V(\theta^*)}$.

---

3. Skattning av μ

Den naturliga skattningen av väntevärdet är medelvärdet:

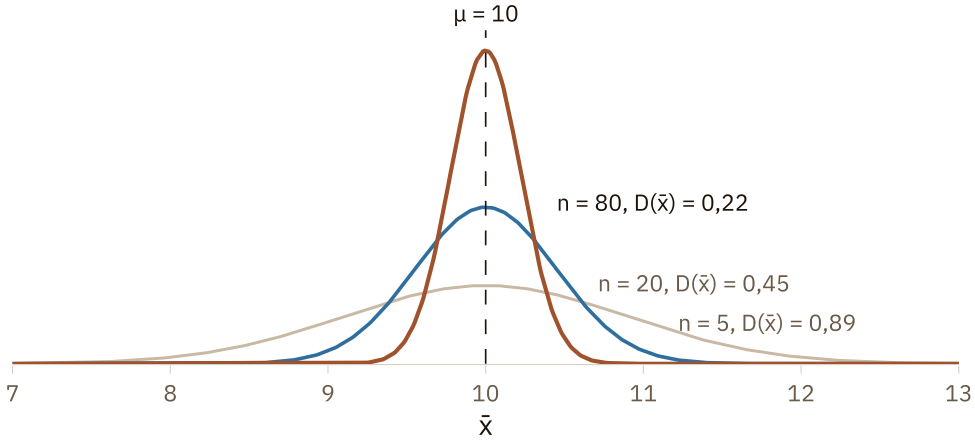
$$\boxed{\mu^* = \bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i}$$

Den är väntevärdesriktig, och för oberoende ξ_i med standardavvikelse σ är

$$E(\bar{\xi}) = \mu, \qquad D(\bar{\xi}) = \frac{\sigma}{\sqrt{n}}$$

Medelfelet minskar som $1/\sqrt{n}$: fyra gånger så många observationer halverar osäkerheten. För normalfördelade observationer är $\bar{\xi}$ den mest effektiva skattningen av μ, och $\bar{\xi} \in N(\mu,\ \sigma/\sqrt{n})$.

Fördelningen för x̄ när σ = 2: alla centrerade kring μ, smalare med större n



Figuren visar fördelningen för $\bar{\xi}$ vid tre stickprovsstorlekar. Alla tre är centrerade kring μ — skattningen är väntevärdesriktig oavsett n — men spridningen krymper, så en enskild observerad skattning hamnar närmare μ ju större stickprovet är.

---

4. Skattning av σ — s-metoden

Variansen skattas med stickprovsvariansen

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

och σ med $s = \sqrt{s^2}$. Motsvarande stickprovsvariabel skrivs σ^*, så s är en observation av σ^*.

Varför $n-1$?

Avvikelserna mäts från $\bar{x}$ och inte från det okända μ. Eftersom $\bar{x}$ ligger mitt i datamaterialet blir avvikelserna i genomsnitt lite för små. Att dela med $n-1$ i stället för n kompenserar exakt för det: $(\sigma^*)^2$ blir väntevärdesriktig för σ^2. På räknaren (One-Variable Statistics) är det `sx`, inte `σx`.

Härledning: $E((\sigma^*)^2) = \sigma^2$

Lägg till och dra ifrån μ i varje avvikelse. Korstermen summerar till $-2n(\bar{\xi}-\mu)^2$, och kvar blir

$$\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2 = \sum_{i=1}^{n}(\xi_i-\mu)^2 - n(\bar{\xi}-\mu)^2$$

Varje term i den första summan har väntevärdet $V(\xi_i)=\sigma^2$, och $E((\bar{\xi}-\mu)^2) = V(\bar{\xi}) = \sigma^2/n$. Alltså

$$E\left(\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2\right) = n\sigma^2 - n\cdot\frac{\sigma^2}{n} = (n-1)\,\sigma^2$$

Delar vi med $n-1$ blir väntevärdet exakt σ^2. Delar vi med n underskattar vi i genomsnitt variansen med faktorn $(n-1)/n$.

Räkneformen

Den högra formen, med $\sum x_i^2$, är den som används när uppgiften ger summor i stället för rådata. Räkna med många decimaler i $\bar{x}$ — differensen $\sum x_i^2 - n\bar{x}^2$ är ofta liten jämfört med termerna.

Ersätter vi σ med s i medelfelet får vi det **skattade medelfelet** $d = s/\sqrt{n}$, som används i t-intervallet.

Exempel — Fem mätningar

Fem mätningar av en längd gav (mm)

$$12{,}1 \quad 11{,}8 \quad 12{,}4 \quad 12{,}0 \quad 12{,}2$$

Mätvärdena betraktas som observationer av oberoende $\xi_i \in N(\mu, \sigma)$. Skatta μ och σ, och beräkna det skattade medelfelet för μ^*.

Lösning

$$\mu^*_{\text{obs}} = \bar{x} = \frac{60{,}5}{5} = \boxed{12{,}1 \text{ mm}}$$

Avvikelserna från $\bar{x}$ är 0, $-0{,}3$, $0{,}3$, $-0{,}1$ och $0{,}1$, så

$$s^2 = \frac{0 + 0{,}09 + 0{,}09 + 0{,}01 + 0{,}01}{5 - 1} = 0{,}05, \qquad s = \sqrt{0{,}05} \approx \boxed{0{,}224 \text{ mm}}$$

$$d = \frac{s}{\sqrt{5}} \approx \boxed{0{,}100 \text{ mm}}$$

En enskild mätning varierar alltså med omkring $0{,}22$ mm, men medelvärdet av fem bara med omkring $0{,}10$ mm.

På räknaren. `menu` → Statistics → Stat Calculations → One-Variable Statistics på listan; läs av `x̄` och `sx`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **`σx` i stället för `sx`.** Räknarens `σx` delar med n; stickprovsstandardavvikelsen s är `sx`.
- **s eller $s/\sqrt{n}$.** s skattar spridningen hos *en* observation, $d = s/\sqrt{n}$ spridningen hos *medelvärdet*.
- **Avrunda $\bar{x}$ för tidigt** i räkneformen $\sum x_i^2 - n\bar{x}^2$; differensen blir då fel.

Se även

- Statistisk inferens
- Konfidensintervall
- t-fördelningen
- Väntevärde

- Varians och standardavvikelse

Resurser

- Khan Academy: Sample variance
- Wikipedia: Bias of an estimator

Teckenintervall

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Statistisk inferens, Binomialfördelning

Teckenintervallet är ett konfidensintervall för medianen som bara bygger på observationernas tecken relativt ett värde. Metoden är fördelningsfri — den kräver inget normalfördelningsantagande — men ger till gengäld ett bredare intervall.

1. Förutsättningar

Observationerna $x_1, \ldots, x_n$ ses som oberoende observationer av en **kontinuerlig** stokastisk variabel ξ med okänd median m. Mer än så behöver vi inte veta — inte vilken fördelning det är.

Medianen uppfyller $P(\xi < m) = P(\xi > m) = 0{,}5$. Varje observation hamnar alltså under m med sannolikheten $0{,}5$, oberoende av de andra.

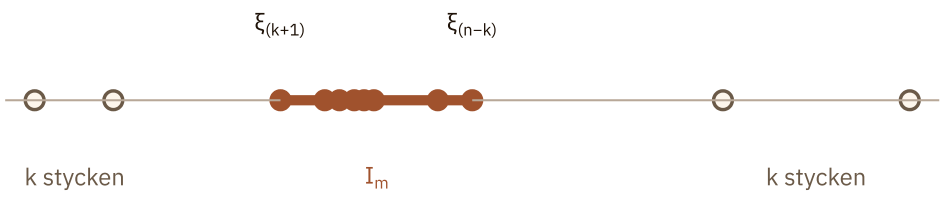
2. Ordnade observationer

Sortera observationerna i växande ordning och beteckna dem

$$\xi_{(1)} \leq \xi_{(2)} \leq \cdots \leq \xi_{(n)}$$

så att $\xi_{(1)}$ är den minsta och $\xi_{(n)}$ den största. Teckenintervallet stryker de k minsta och de k största:

$$\boxed{I_m = [\xi_{(k+1)},\ \xi_{(n-k)}]}$$



3. Konfidensgraden

Låt η = antalet observationer som är mindre än m. Varje observation är en oberoende upprepning med sannolikheten $0{,}5$, så

$$\eta \in \mathrm{Bin}(n,\ 0{,}5)$$

Intervallet missar m om m ligger till vänster om det — högst k observationer under m — eller till höger — högst k observationer över m. De två händelserna är disjunkta och lika sannolika av symmetriskäl:

$$\boxed{P\,(m \in I_m) = 1 - 2\,P\,(\eta \le k), \qquad \eta \in \mathrm{Bin}(n,\ 0{,}5)}$$

Konfidensgraden kan inte väljas fritt

Eftersom k är ett heltal kan vi bara få vissa konfidensgrader. Vi väljer det största k som ger minst den konfidensgrad vi vill ha, till exempel $95\ \%$.

Generell men inte effektiv

Teckenintervallet använder bara om varje observation ligger under eller över m, inte hur långt ifrån. Vet vi att observationerna är normalfördelade ger normalfördelningsintervallen kortare intervall med samma konfidensgrad.

Vanliga fel

- **Glömma faktorn 2.** Konfidensgraden är $1 - 2P\,(\eta \le k)$; intervallet kan missa på båda sidor.
- **Räkna från fel ände.** Intervallet går från den $(k+1)$:a minsta till den $(k+1)$:a största observationen.

4. Lösta lektionsuppgifter

Lektion 11, övning 1 och 2 — Konfidensgrad för teckenintervall

Låt ξ vara en kontinuerlig stokastisk variabel med median m, och gör 10 oberoende mätningar.

a) Bestäm $P(\xi < m)$.

b) Vilken konfidensgrad har teckenintervallet $[\xi_{(1)},\ \xi_{(10)}]$?

c) Låt η = antalet mätvärden som är $< m$. Vilka värden kan η anta, och vilken fördelning har η?

d) Bestäm konfidensgraden för intervallet $[\xi_{(3)},\ \xi_{(8)}]$.

e) Generalisera till ett stickprov av storlek n och intervallet $[\xi_{(k+1)},\ \xi_{(n-k)}]$.

Lösning

a) Medianen delar massan på mitten: $P(\xi < m) = \boxed{0{,}5}$.

c) η kan anta värdena $0, 1, \ldots, 10$. Varje mätning hamnar under m med sannolikheten $0{,}5$, oberoende av de andra, så $\eta \in \mathrm{Bin}(10,\ 0{,}5)$.

b) $[\xi_{(1)},\ \xi_{(10)}]$ är teckenintervallet med $k = 0$. Det missar bara om alla tio hamnar på samma sida om m:

$$P(m \in I) = 1 - 2\,P(\eta = 0) = 1 - 2 \cdot 0{,}5^{10} \approx \boxed{0{,}998}$$

d) $[\xi_{(3)},\ \xi_{(8)}]$ stryker två observationer i varje ände, $k = 2$:

$$P(m \in I) = 1 - 2\,P(\eta \le 2) = 1 - 2 \cdot \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2}}{2^{10}} = 1 - 2 \cdot \frac{56}{1024} \approx \boxed{0{,}891}$$

e) Samma resonemang med $\eta \in \mathrm{Bin}(n,\ 0{,}5)$ ger $P(m \in I) = 1 - 2P(\eta \le k)$, avsnitt 3 ovan.

På räknaren. d) `1-2*binomCdf(10,0.5,0,2)`.

Facit: b) konfidensgraden 99,8 % (med bokens data i övning 8.7 blir intervallet [140, 520]); c) $\eta \in \mathrm{Bin}(10,\ 0{,}5)$*; d) 0,891.*

Lektion 12, övning 10 (Vännman 8.19 b) — Densitet utan normalfördelningsantagande

Densiteten (g/cm³) hos ett material mäts 13 gånger. Vi antar ingenting om att mätvärdena är normalfördelade, bara att de är oberoende observationer av en kontinuerlig fördelning. Bilda ett konfidensintervall för fördelningens median m med konfidensgrad nära $98\ \%$.

Lösning

Metod. Utan normalfördelningsantagande är teckenintervallet det som går att använda. Förutsättningen — oberoende observationer från en kontinuerlig fördelning — är uppfylld.

Konfidensgrad per k. Med $n = 13$ och $\eta \in \mathrm{Bin}(13,\ 0{,}5)$:

k	0	1	2	3
$1 - 2P(\eta \le k)$	0,9998	0,9966	0,9775	0,9077

$k = 2$ ger $97{,}8\ \%$, det närmaste vi kommer $98\ \%$. Intervallet blir $[\xi_{(3)},\ \xi_{(11)}]$: stryk de två minsta och de två största mätvärdena.

Resultat. Med bokens mätvärden blir intervallet $\boxed{[0{,}501;\ 0{,}511]}$ med konfidensgrad $97{,}8\ \%$.

Jämförelse. Med normalfördelningsantagande blev $98\ \%$-intervallet $[0{,}502;\ 0{,}514]$, se t-fördelningen. Intervallen mäter olika saker — median respektive väntevärde — men för en normalfördelning sammanfaller de.

Uppgiften (Vännman 8.19 b, i 2:a uppl. 8.20 b) är återgiven ur det handskrivna lösningsförslaget, och intervallet och konfidensgraderna stämmer med det. Mätvärdena finns inte i valvet.

Se även

- Konfidensintervall
- Binomialfördelning
- Statistisk inferens

- Median och kvantiler

Resurser

- Wikipedia: Sign test
- Wikipedia: Order statistic

Konfidensintervall

19 okt. 2026, 6 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Punktskattning, Normalfördelning

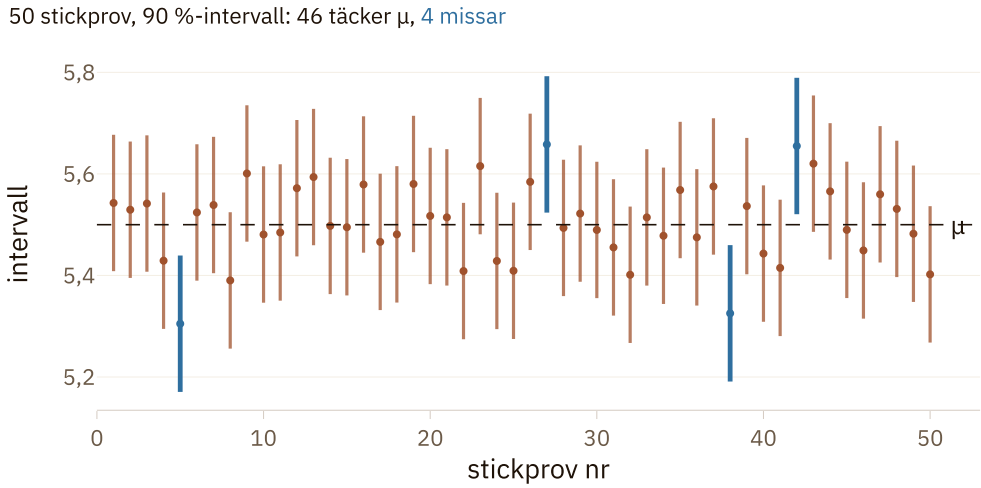
Ett konfidensintervall är ett intervall, beräknat ur stickprovet, som med en angiven konfidensgrad täcker den okända parametern. Konfidensgraden avser metoden och inte det enskilda intervallet: upprepade stickprov ger intervall som i exempelvis $95\ \%$ av fallen innehåller det sanna värdet.

1. Definitionen

Ett **konfidensintervall** I_θ för parametern θ med **konfidensgrad** $1-\alpha$ är ett intervall vars gränser är stickprovsvariabler, konstruerat så att

$$P(\theta \in I_\theta) = 1 - \alpha$$

Sannolikheten gäller *innan* vi har observerat: gränserna är slumpmässiga, θ är fix. När vi sätter in de observerade värdena får vi ett tal-intervall, och det antingen innehåller θ eller inte.



Figuren visar 50 stickprov från samma fördelning, vart och ett med sitt $90\ \%$-intervall. De flesta täcker det sanna μ; ungefär vart tionde missar. Vi vet aldrig om just vårt intervall är ett av de som missar — konfidensgraden säger hur ofta metoden lyckas.

2. Härledning: μ i $N(\mu, \sigma)$ med känt σ

Anta att $\xi_1, \ldots, \xi_n$ är oberoende och $N(\mu, \sigma)$, med σ känd. Den bästa punktskattningen är $\mu^* = \bar{\xi}$, och $\bar{\xi} \in N(\mu,\ \sigma/\sqrt{n})$. Standardisera:

$$\frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}} \in N(0, 1)$$

Med kvantilen $\lambda_{\alpha/2}$, där $P(\eta > \lambda_{\alpha/2}) = \alpha/2$ för $\eta \in N(0, 1)$, gäller

$$P\left(-\lambda_{\alpha/2} < \frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}} < \lambda_{\alpha/2}\right) = 1 - \alpha$$

Lös ut μ ur olikheterna:

$$P\left(\bar{\xi} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{\xi} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Ersätt $\bar{\xi}$ med det observerade $\bar{x}$:

$$\boxed{I_\mu = \left[\bar{x} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \ \bar{x} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Konfidensgrad	90 %	95 %	99 %
$\lambda_{\alpha/2}$	1,6449	1,9600	2,5758

3. Vad bestämmer intervallets bredd?

Intervallets bredd är $2\lambda_{\alpha/2}\,\sigma/\sqrt{n}$. Allt annat lika:

Ändring	**Effekt på bredden**
högre konfidensgrad	större $\lambda_{\alpha/2}$ — **bredare**
större spridning σ	**bredare**
fler observationer n	**smalare**, som $1/\sqrt{n}$

> ⚠ **Säkerhet kostar precision**
>
> Ett $100\ \%$-intervall finns alltid — hela reella axeln — men säger ingenting. Konfidensgraden och bredden står i motsats till varandra, och det enda sättet att få både hög säkerhet och ett smalt intervall är fler observationer.

---

4. Välj rätt metod

Tre metoder för läget i en fördelning, beroende på vad vi vet i förväg:

Förutsättning	Intervall	Not
Kontinuerlig fördelning, inget mer	$[\xi_{(k+1)},\ \xi_{(n-k)}]$ för medianen	Teckenintervall
Normalfördelning, σ **känd**	$\bar{x} \pm \lambda_{\alpha/2}\, \sigma/\sqrt{n}$	denna not
Normalfördelning, σ **okänd**	$\bar{x} \pm t_{\alpha/2}(n-1)\, s/\sqrt{n}$	t-fördelningen

Jämförelser mellan två metoder eller grupper leder vidare till Stickprov i par och Två stickprov.

> 🔥 **Tolka intervallet i ord**
>
> Ett svar är inte klart förrän intervallet är tolkat: “Med $95\ \%$ säkerhet ligger den förväntade halten mellan $0{,}11$ och $0{,}39$.” Vid jämförelser: ingår 0 i intervallet för skillnaden kan vi inte påvisa någon skillnad.

> ⚠ **Vanliga fel**
>
> - **Fel kvantil.** Ett tvåsidigt intervall med konfidensgrad $1-\alpha$ använder $\lambda_{\alpha/2}$; för $95\ \%$ är det $1{,}96$, inte $1{,}645$.
> - **Tolkning.** ”μ ligger i intervallet med sannolikhet $95\ \%$” är fel efter observationen; det är *metoden* som träffar i $95\ \%$ av fallen.
> - σ **känd eller skattad.** Har σ skattats ur samma data gäller [[t-fördelningen|t-intervallet]].

---

5. Lösta lektionsuppgifter

Lektion 11, övning 3 och 4 (Vännman 8.9) — Halten av ett ämne

Halten (%) av ett visst ämne i en råvara mäts sex gånger. Mätvärdena kan betraktas som observationer från $N(\mu, \sigma)$ med $\sigma = 0{,}2$ känd:

$$5{,}68 \quad 5{,}58 \quad 5{,}98 \quad 6{,}30 \quad 4{,}93 \quad 4{,}98$$

a) Härled och beräkna ett konfidensintervall för μ med konfidensgrad $0{,}90$. Gå igenom stegen: fördelningen för ξ_i, bästa punktskattningen av μ, intervallskattningen, och till sist insättning av observationerna.

b) (Övning 4.) Är det troligt att halten μ är $6{,}0$ %?

Lösning

Modell. ξ_i = mätvärde nr i, $\xi_i \in N(\mu,\ 0{,}2)$, $i = 1, \ldots, 6$, oberoende.

Punktskattning. $\mu^* = \bar{\xi} \in N(\mu,\ 0{,}2/\sqrt{6})$. Observerat: $\bar{x} = 5{,}575$.

Intervallskattning. Vi söker a så att $P(\bar{\xi} - a < \mu < \bar{\xi} + a) = 0{,}90$:

$$0{,}90 = P\left(-\frac{a}{\sigma/\sqrt{6}} < \frac{\bar{\xi} - \mu}{\sigma/\sqrt{6}} < \frac{a}{\sigma/\sqrt{6}}\right) = 2\,\Phi\left(\frac{a}{\sigma/\sqrt{6}}\right) - 1$$

Alltså $\Phi(a/(\sigma/\sqrt{6})) = 0{,}95$, så $a/(\sigma/\sqrt{6}) = \lambda_{0{,}05} = 1{,}6449$ och

$$a = 1{,}6449 \cdot \frac{0{,}2}{\sqrt{6}} \approx 0{,}134$$

a)

$$I_\mu = [5{,}575 - 0{,}134;\ \ 5{,}575 + 0{,}134] = \boxed{[5{,}44;\ 5{,}71]}$$

Med 90 % säkerhet ligger den förväntade halten mellan $5{,}44$ och $5{,}71$ %.

På räknaren. `menu` → Statistics → Confidence Intervals → z Interval med **Stats**: σ = `0.2`, $\bar{x}$ = `5.575`, n = `6`, `C Level` = `0.9`.

b) $6{,}0$ ingår inte i intervallet. Det är därför **inte troligt** att $\mu = 6{,}0$ % — metoden ger intervall som missar μ bara i 10 % av fallen.

Facit (övning 4): Nej, eftersom 6 inte ingår i intervallet. Intervallet stämmer med det handskrivna lösningsförslaget.

Lektion 11, övning 5 — Vad påverkar bredden?

Utgå från intervallet i övning 3 ($\sigma = 0{,}2$, $n = 6$, konfidensgrad 90 %). Hur påverkas intervallet av

a) konfidensgraden — vad händer om den blir 80 %? 99 %?

b) spridningen — vad händer om $\sigma = 0{,}4$? $\sigma = 0{,}1$?

c) antalet mätvärden — vad händer om $n = 10$? $n = 4$?

Lösning

Bredden är $2\lambda_{\alpha/2}\,\sigma/\sqrt{n}$; i utgångsläget $2 \cdot 0{,}134 = 0{,}269$.

Ändring	Bredd
80 % ($\lambda = 1{,}2816$)	0,209
99 % ($\lambda = 2{,}5758$)	0,421
$\sigma = 0{,}4$	0,537
$\sigma = 0{,}1$	0,134
$n = 10$	0,208
$n = 4$	0,329

a) Ökad konfidensgrad ger **bredare** intervall.

b) Ökad standardavvikelse ger **bredare** intervall — bredden är proportionell mot σ.

c) Fler mätvärden ger **smalare** intervall, men bara som $1/\sqrt{n}$.

Facit: a) ökad konfidensgrad ger bredare intervall, b) ökad standardavvikelse ger bredare intervall, c) ökad stickprovsstorlek ger smalare intervall — allt annat lika.

Lektion 12, övning 1 — Känd standardavvikelse

Observationerna nedan kan betraktas som observationer från $N(\mu, \sigma)$. Anta att $\sigma = 0{,}2$ är känd i förväg. Beräkna ett $95\ \%$ konfidensintervall för μ och tolka det i ord.

$0{,}5 \quad 0{,}2 \quad 0{,}3 \quad -0{,}2 \quad 0{,}3 \quad 0{,}3 \quad 0{,}4 \quad 0{,}2$

Lösning

Modell. $\xi_i \in N(\mu,\ 0{,}2)$, $i = 1, \ldots, 8$, oberoende.

$\bar{x} = 2{,}0/8 = 0{,}25$ och $\lambda_{0{,}025} = 1{,}96$:

$$I_\mu = [0{,}25 \mp 1{,}96 \cdot \frac{0{,}2}{\sqrt{8}}] = [0{,}25 \mp 0{,}139] = \boxed{[0{,}11;\ 0{,}39]}$$

Tolkning. Med $95\ \%$ säkerhet ligger det förväntade värdet μ mellan $0{,}11$ och $0{,}39$.

Samma data med σ okänd finns i t-fördelningen.

På räknaren. z Interval med **Data** (listan med de åtta värdena), σ = `0.2`, `C Level` = `0.95`.

Facit: [0,11; 0,39].

Se även

- Punktskattning
- t-fördelningen
- Teckenintervall

- Stickprov i par
- Två stickprov
- Konfidensintervall och hypotesprövning

Resurser

- Khan Academy: Confidence intervals
- Wikipedia: Konfidensintervall

t-fördelningen

19 okt. 2026, 5 min läsning

UTKAST

#matematik #statistik #fördelning

Kurs: S0008M **Förkunskaper:** Konfidensintervall, Punktskattning

---

När standardavvikelsen σ är okänd och ersätts med sin skattning s är den standardiserade medelvärdesvariabeln inte längre $N(0,1)$ utan t-fördelad med $n-1$ frihetsgrader. t-fördelningen har tjockare svansar, och konfidensintervallen blir därför bredare.

---

1. Två kvoter som liknar varandra

För oberoende $\xi_i \in N(\mu,\sigma)$ jämför vi

$$z = \frac{\bar{\xi}-\mu}{\sigma/\sqrt{n}} \qquad \text{och} \qquad t = \frac{\bar{\xi}-\mu}{\sigma^*/\sqrt{n}}$$

Skillnaden är nämnaren. I z står den fasta konstanten σ. I t står skattningen σ^*, som själv är slumpmässig och varierar från stickprov till stickprov.

Fördelningarna

$$\boxed{\frac{\bar{\xi}-\mu}{\sigma/\sqrt{n}} \in N(0,1), \qquad \frac{\bar{\xi}-\mu}{\sigma^*/\sqrt{n}} \in t(n-1)}$$

$t(f)$ är t-fördelningen med f **frihetsgrader**. För ett stickprov är $f = n-1$, samma tal som nämnaren i s^2.

Varför de måste skilja sig

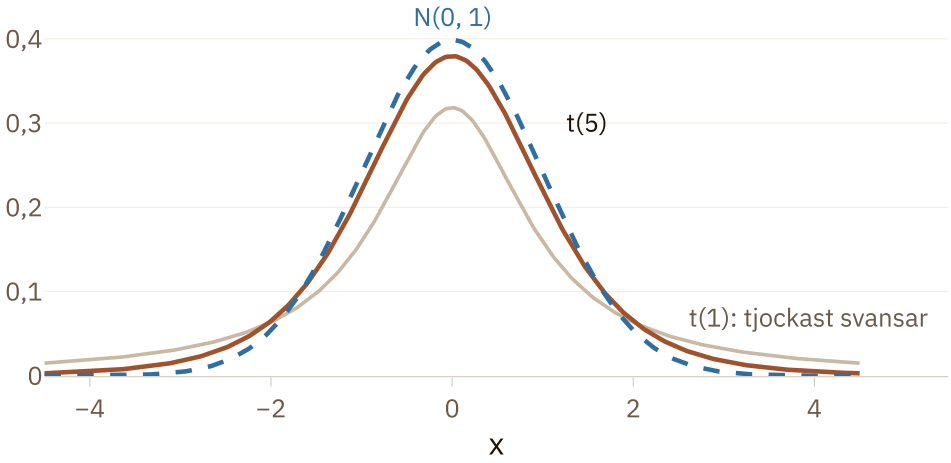
Tänk dig att vi simulerar många stickprov. Ibland blir s av en slump mycket mindre än σ, och då blir t stort även om $\bar{\xi}$ ligger nära μ. Den extra osäkerheten i nämnaren ger fler extrema värden — tjockare svansar — än $N(0,1)$. Med små stickprov är effekten stor; med stora är s nästan alltid

nära σ och skillnaden försvinner.

Fördelningen härleddes av kemisten W. S. Gosset, som arbetade med kvalitetskontroll på Guinness bryggeri och publicerade under pseudonymen *Student*. Därav namnet Students t-fördelning.

---

2. Egenskaper



- Symmetrisk kring 0, klockformad som $N(0, 1)$.
- **Tjockare svansar** — mer massa långt ut.
- När $f \to \infty$ går $t(f)$ mot $N(0, 1)$.

Kvantilen $t_\alpha(f)$ definieras som för normalfördelningen: $P\,(t > t_\alpha(f)) = \alpha$. Den läses ur t-tabellen och är alltid större än motsvarande λ_α:

f	3	7	12	24	∞
$t_{0{,}025}(f)$	3,182	2,365	2,179	2,064	$1{,}960 = \lambda_{0{,}025}$

---

3. Konfidensintervall för μ med okänt σ

Samma härledning som i Konfidensintervall, med $\lambda_{\alpha/2}$ utbytt mot $t_{\alpha/2}(n-1)$ och σ mot s:

$$I_\mu = [\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \ \bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}]$$

⚠ Vilket intervall när?

λ-intervallet kräver att σ är **känd i förväg**, innan vi ser mätvärdena — till exempel från lång erfarenhet av mätinstrumentet. Har vi bara räknat ut s ur samma data är σ okänd, och då gäller t-intervallet. Det vanliga fallet i praktiken är okänd σ.

På räknaren: `menu` → Statistics → Confidence Intervals → t Interval, eller kvantilen direkt med `invt(1-α/2,f)`.

⚠ Vanliga fel

- **Frihetsgraderna.** Ett stickprov har $f = n - 1$, inte n.
- **`invt` med fel svans.** `invt(p,f)` ger kvantilen med sannolikheten p *till vänster*; $t_{0{,}025}(8)$ är `invt(0.975,8)`.

---

4. Lösta lektionsuppgifter

Lektion 12, övning 2 — Samma data, okänd standardavvikelse

Observationerna $0{,}5 \quad 0{,}2 \quad 0{,}3 \quad -0{,}2 \quad 0{,}3 \quad 0{,}3 \quad 0{,}4 \quad 0{,}2$ kommer från $N(\mu, \sigma)$. Nu vet vi ingenting i förväg om σ. Beräkna ett $95\ \%$ konfidensintervall för μ och tolka det.

Lösning

Modell. $\xi_i \in N(\mu, \sigma)$, $i = 1, \ldots, 8$, oberoende, μ och σ okända.

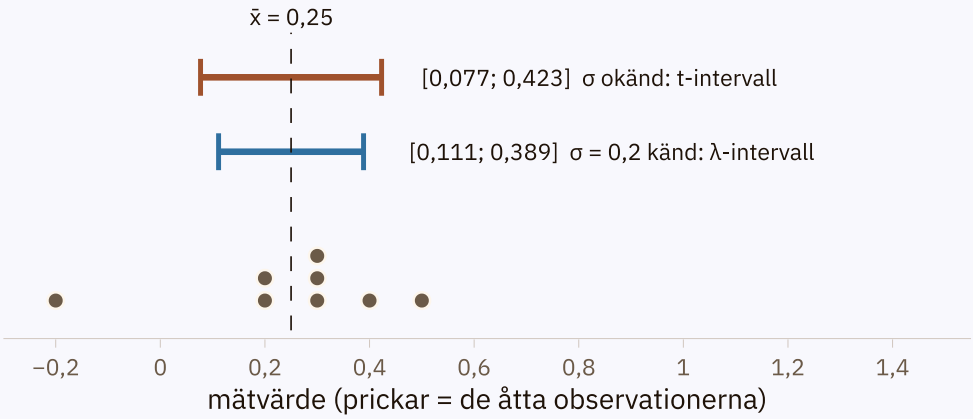
Skattningar. $\bar{x} = 0{,}25$ och

$$s = \sqrt{\frac{1}{7}\sum_{i=1}^{8}(x_i - 0{,}25)^2} \approx 0{,}207$$

Intervall. $t_{0{,}025}(7) = 2{,}365$:

$$I_\mu = [0{,}25 \mp 2{,}365 \cdot \frac{0{,}207}{\sqrt{8}}] = [0{,}25 \mp 0{,}173] \approx \boxed{[0{,}08;\ 0{,}42]}$$

Tolkning. Med $95\ \%$ säkerhet ligger μ mellan $0{,}08$ och $0{,}42$.



Jämförelse. Med $\sigma = 0{,}2$ känd blev intervallet $[0{,}11;\ 0{,}39]$, se Konfidensintervall. Här är $s \approx \sigma$, men $t_{0{,}025}(7) = 2{,}365$ är större än $1{,}96$ — priset för att inte känna σ.

På räknaren. Mata in de åtta värdena i en lista och kör t Interval med **Data** och `C Level 0.95`.

Facit: [0,07; 0,43] — facit avrundar gränserna utåt; de oavrundade gränserna är 0,077 och 0,423.

Lektion 12, övning 3 och 4 — z eller t?

Låt $\xi_1, \ldots, \xi_n$ vara oberoende och $N(\mu, \sigma)$.

a) På vilket sätt skiljer sig $\dfrac{\bar{\xi} - \mu}{\sigma^*/\sqrt{n}}$ och $\dfrac{\bar{\xi} - \mu}{\sigma/\sqrt{n}}$ åt, och vilka fördelningar har de?

b) Du ska beräkna ett konfidensintervall för μ med konfidensgrad $1 - \alpha$ och väljer mellan

$$[\bar{x} \mp \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}] \qquad \text{och} \qquad [\bar{x} \mp t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}]$$

I vilken situation ska respektive intervall användas?

Lösning

a) Den första har den slumpmässiga skattningen σ^* i nämnaren, den andra den fasta parametern σ. Därför

$$\frac{\bar{\xi}-\mu}{\sigma/\sqrt{n}} \in N(0,1), \qquad \frac{\bar{\xi}-\mu}{\sigma^*/\sqrt{n}} \in t(n-1)$$

Variationen i σ^* gör att t-kvoten oftare hamnar långt ut, se avsnitt 1.

b) λ-intervallet när σ är **känd i förväg**; t-intervallet när σ är **okänd** och skattas med s ur samma mätvärden. Båda förutsätter normalfördelade observationer.

Facit (övning 3): $N(0,1)$ *respektive* $t(n-1)$.

Lektion 12, övning 9 (Vännman 8.19 a) — Densitet, 98 %-intervall

Densiteten (g/cm³) hos ett material mäts 13 gånger. Mätvärdena antas vara oberoende observationer från $N(\mu,\sigma)$ med okänt σ. Ur mätvärdena fås $\bar{x} = 0{,}508$ och $\sum x_i^2 = 3{,}355708$. Beräkna ett $98\ \%$ konfidensintervall för μ.

Lösning

Modell. ξ_i = densitet vid mätning i, $\xi_i \in N(\mu,\sigma)$, $i = 1,\dots,13$, oberoende.

Skattning av σ med räkneformen:

$$s^2 = \frac{1}{12}\left(3{,}355708 - 13\cdot 0{,}508^2\right) = 7{,}3\cdot 10^{-5}, \qquad s \approx 0{,}00854$$

Intervall. $98\ \%$ ger $\alpha/2 = 0{,}01$ och $t_{0{,}01}(12) = 2{,}681$:

$$I_\mu = \left[0{,}508 \mp 2{,}681\cdot\frac{0{,}00854}{\sqrt{13}}\right] = [0{,}508 \mp 0{,}0064] \approx \boxed{[0{,}502;\ 0{,}514]}$$

Med $98\ \%$ säkerhet ligger den förväntade densiteten mellan $0{,}502$ och $0{,}514$ g/cm³. Utan normalfördelningsantagande, se Teckenintervall.

Uppgiften (Vännman 8.19 a, i 2:a uppl. 8.20 a) är återgiven ur det handskrivna lösningsförslaget, och intervallet stämmer med det: [0,501; 0,515], avrundat utåt från 0,5016 och 0,5144. Rådata finns inte i valvet.

Se även

- Konfidensintervall
- Punktskattning
- Normalfördelning
- Stickprov i par
- Två stickprov

Resurser

- Khan Academy: Confidence intervals
- Wikipedia: Students t-fördelning

Stickprov i par

19 okt. 2026, 6 min läsning

UTKAST

#matematik #statistik #metod

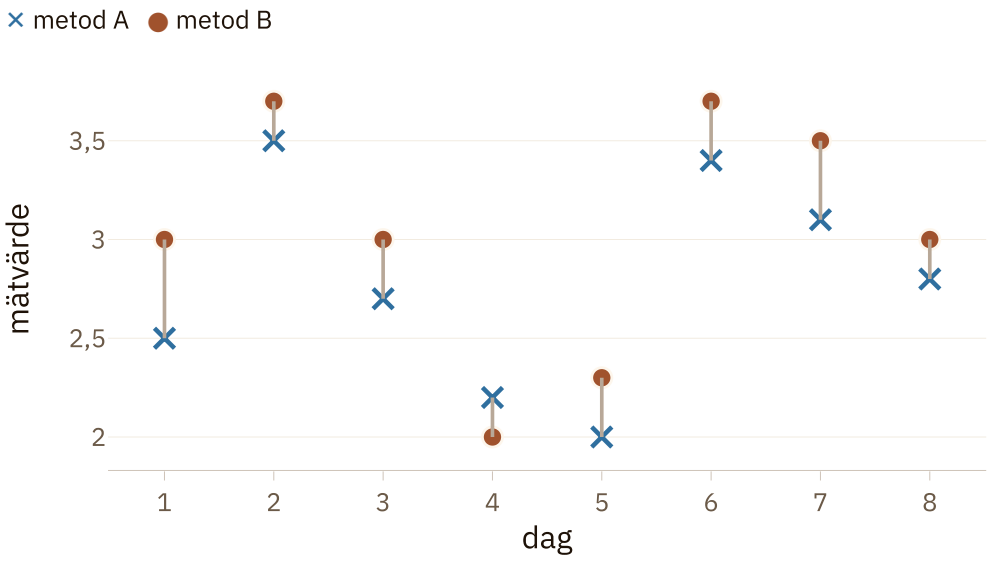
> **Kurs:** S0008M **Förkunskaper:** Konfidensintervall, t-fördelningen

Stickprov i par används när observationerna kommer parvis, till exempel två mätningar på samma individ. Vi bildar differensen inom varje par och analyserar dessa differenser som ett enda stickprov, vilket räknar bort variationen mellan individer.

1. Situationen

Vi vill jämföra två metoder, A och B, och planerar försöket så att båda används på samma objekt: samma dag, samma patient, samma vajer delad i två bitar. Då får vi par av mätvärden $(x_i,\ y_i)$, $i = 1, \dots, n$.

Objekten kan skilja sig mycket från varandra — en dag ger höga värden, en annan låga — men det intressanta är skillnaden *inom* varje par.



I figuren varierar nivån kraftigt från dag till dag, men B ligger nästan alltid lite över A. Det är den systematiska skillnaden vi vill skatta.

2. Modellen

$$x_i \text{ observation från } N(\mu_i,\ \sigma_1), \qquad y_i \text{ observation från } N(\mu_i + \Delta,\ \sigma_2), \qquad i = 1, \dots, n$$

- μ_i är nivån för objekt i. Den får variera fritt mellan objekten, och vi behöver inte skatta den.
- Δ är den **förväntade skillnaden** mellan metoderna — samma för alla objekt. Det är parametern vi vill veta något om.

Bilda differenserna $z_i = y_i - x_i$. Nivån μ_i tar ut sig:

$$\boxed{z_i \text{ är observationer från } N(\Delta,\ \sigma), \quad i = 1, \dots, n, \text{ oberoende}}$$

Problemet har blivit ett enda stickprov med okänt väntevärde Δ och okänd standardavvikelse — precis situationen i t-fördelningen.

---

3. Konfidensintervallet

Med $\bar{z}$ och s_z beräknade ur differenserna:

$$\boxed{I_\Delta = [\bar{z} - t_{\alpha/2}(n-1)\frac{s_z}{\sqrt{n}},\ \ \bar{z} + t_{\alpha/2}(n-1)\frac{s_z}{\sqrt{n}}]}$$

Tolka med nollan

- Ingår **inte** 0 i I_Δ: vi kan med den givna säkerheten påstå att metoderna skiljer sig i genomsnitt, och intervallet säger hur mycket.
- Ingår 0: vi kan inte påvisa någon skillnad. Det betyder inte att skillnaden är noll — bara att datamaterialet inte räcker för att visa den.

På räknaren: lägg mätvärdena i två kolumner i Lists & Spreadsheet, bilda differenserna i en tredje med kolumnformeln `=b[]-a[]`, och kör t Interval med **Data** på den kolumnen.

Varför paren lönar sig

Variationen mellan objekten, $\mu_1, \dots, \mu_n$, försvinner helt ur differenserna. Hade vi i stället mätt metod A på några objekt och metod B på andra hade den variationen hamnat i spridningen, och intervallet för skillnaden blivit mycket bredare. Jämför Två stickprov.

⚠ Vanliga fel

- **Analysera paren som två stickprov.** Då hamnar variationen mellan objekten i s och intervallet blir onödigt brett.
- **Blanda ordningen i differensen.** Bestäm $z_i = x_i - y_i$ en gång och tolka tecknet därefter.

---

4. Lösta lektionsuppgifter

Lektion 12, övning 5 — Två mätmetoder under åtta dagar

Två mätmetoder, A och B, används samma dag under åtta dagar:

Dag	1	2	3	4	5	6
Metod A	2,5	3,5	2,7	2,2	2,0	3,4
Metod B	3,0	3,7	3,0	2,0	2,3	3,7

a) Pricka in mätvärdena dag för dag i ett diagram. Tror du att metoderna skiljer sig åt?

b) Anta att mätvärdet för metod A dag i är en observation från $N(\mu_i, \sigma)$ och för metod B från $N(\mu_i + \Delta, \sigma)$. Tolka Δ och $\mu_1, \dots, \mu_8$.

c) Kan vi med $95\ \%$ säkerhet påstå att metoderna skiljer sig i genomsnitt? Hur stor är i så fall skillnaden?

Lösning

a) Se figuren i avsnitt 1. Nivån varierar mycket mellan dagarna, men B ligger över A alla dagar utom dag 4. Det tyder på en systematisk skillnad.

b) μ_i är den förväntade nivån dag i — det som gör att vissa dagar ger höga värden och andra låga. Det är rimligt att de skiljer sig, och figuren visar att de gör det. Δ är den förväntade skillnaden mellan metod B och metod A, densamma alla dagar.

c) Differenserna $z_i = y_i - x_i$ (B − A):

$$0{,}5 \quad 0{,}2 \quad 0{,}3 \quad -0{,}2 \quad 0{,}3 \quad 0{,}3 \quad 0{,}4 \quad 0{,}2$$

är observationer från $N(\Delta, \sigma_z)$. $\bar{z} = 0{,}25$, $s_z \approx 0{,}207$, $t_{0{,}025}(7) = 2{,}365$:

$$I_\Delta = [0{,}25 \mp 2{,}365 \cdot \frac{0{,}207}{\sqrt{8}}] \approx \boxed{[0{,}08;\ 0{,}42]}$$

0 ingår inte. Med $95\ \%$ säkerhet finns en genomsnittlig skillnad: metod B ger i genomsnitt mellan $0{,}08$ och $0{,}42$ högre mätvärden än metod A.

Rimlighetskoll. Differenserna är exakt observationerna i lektion 12, övning 2, så intervallet måste bli detsamma.

På räknaren. t Interval med **Data** på differenskolumnen, `C Level` `0.95`.

Facit: [0,07; 0,43], avrundat utåt; metod B ger i genomsnitt mellan 0,07 och 0,43 högre mätvärden än metod A.

Lektion 13, övning 1 — Torsionsprov på stålvajrar

För att studera om två maskiner, A och B, för torsionsprov på stålvajrar ger olika resultat valdes sex vajrar av olika typ ut och delades i två likadana bitar. Slumpen avgjorde vilken bit som testades i vilken maskin.

Typ av vajer	1	2	3	4	5	6
Maskin A	27	35	38	28	40	42
Maskin B	25	34	39	26	37	42

Kan vi påstå att maskinerna i genomsnitt ger olika resultat? Hur stor är i så fall skillnaden? Beräkna och tolka ett $90\ \%$ konfidensintervall, och redovisa modellantagandena.

Lösning

Modell. Mätvärdet för vajer i i maskin A är en observation från $N(\mu_i + \Delta,\ \sigma_1)$ och i maskin B från $N(\mu_i,\ \sigma_2)$, alla oberoende. μ_i är vajertypens nivå och Δ den förväntade skillnaden A – B. Stickprov i par, eftersom varje vajer testas i båda maskinerna.

Differenser z_i = A − B:

$$2 \quad 1 \quad -1 \quad 2 \quad 3 \quad 0$$

observationer från $N(\Delta, \sigma)$. $\bar{z} \approx 1{,}167$, $s_z \approx 1{,}472$, $t_{0{,}05}(5) = 2{,}015$:

$$I_\Delta = [1{,}167 \mp 2{,}015 \cdot \frac{1{,}472}{\sqrt{6}}] = [1{,}167 \mp 1{,}211] \approx \boxed{[-0{,}04;\ 2{,}38]}$$

Tolkning. 0 ingår i intervallet, så vi kan inte med $90\ \%$ säkerhet påvisa någon skillnad i förväntat mätresultat mellan maskinerna. Den nedre gränsen ligger nära 0, så fler observationer skulle behövas för ett säkrare besked.

På räknaren. t Interval på differenserna med `C Level` `0.9`.

Facit: [−0,05; 2,38], avrundat utåt; ingen skillnad kan påvisas med 90 % säkerhet, fler observationer rekommenderas.

Lektion 13, övning 3b (Vännman 8.23 b) — Blodtryck före och efter

I ett försök mäts blodtrycket hos 25 patienter före (x) och efter (y) behandling med en medicin. Resultat: $\bar{x} = 149$, $s_x = 8{,}1$, $\bar{y} = 150{,}9$, $s_y = 9{,}5$, och för differenserna $z_i = y_i - x_i$: $\bar{z} = 1{,}9$, $s_z = 1{,}6$. Beräkna ett $95\ \%$ konfidensintervall för den förväntade förändringen och tolka det.

Lösning

Modell. x_i observation från $N(\mu_i, \sigma_1)$, y_i från $N(\mu_i + \Delta, \sigma_2)$; differenserna z_i observationer från $N(\Delta, \sigma)$, $i = 1, \ldots, 25$. Stickprov i par — samma patient mäts två gånger.

$t_{0{,}025}(24) = 2{,}064$:

$$I_\Delta = [1{,}9 \mp 2{,}064 \cdot \frac{1{,}6}{\sqrt{25}}] = [1{,}9 \mp 0{,}66] \approx \boxed{[1{,}24;\ 2{,}56]}$$

Tolkning. 0 ingår inte. Med $95\ \%$ säkerhet höjer medicinen blodtrycket i genomsnitt med mellan $1{,}24$ och $2{,}56$ enheter.

Observera att $s_z = 1{,}6$ är mycket mindre än s_x och s_y: variationen mellan patienter är stor, men förändringen hos varje patient är liten och stabil. Deluppgift a), samma fråga med två stickprov, finns i Två stickprov.

Uppgiften (Vännman 8.23 b, i 2:a uppl. 8.25 b) är återgiven ur det handskrivna lösningsförslaget, och intervallet stämmer med det.

Se även

- Två stickprov
- t-fördelningen
- Konfidensintervall
- Hypotesprövning

Resurser

- Wikipedia: Paired difference test

Två stickprov

19 okt. 2026, 6 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Stickprov i par, t-fördelningen

---

Tvåstickprovsmetoden jämför väntevärdena i två grupper när observationerna inte är parvisa utan oberoende. Osäkerheten blir större än vid parvisa observationer, eftersom variationen mellan individer då inte kan elimineras.

---

1. Modellen

Två oberoende stickprov, ett från varje grupp:

$$x_1, \dots, x_{n_1} \text{ observationer från } N(\mu_1,\ \sigma), \qquad y_1, \dots, y_{n_2} \text{ observationer från } N(\mu_2,\ \sigma)$$

alla oberoende. Vi söker ett intervall för skillnaden $\mu_1 - \mu_2$.

Samma σ i båda grupperna

Metoden förutsätter att grupperna har **samma standardavvikelse**, även om den är okänd. Väntevärdena får skilja sig, spridningen inte. Det är detta som gör att vi kan väga ihop de två skattningarna av σ.

---

2. Punktskattning av skillnaden

$\bar{\xi} - \bar{\eta}$ skattar $\mu_1 - \mu_2$. Eftersom stickproven är oberoende adderas varianserna:

$$\bar{\xi} - \bar{\eta} \in N\left(\mu_1 - \mu_2,\ \ \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

---

3. Sammanvägd skattning av σ

Varje stickprov ger sin egen skattning, s_1^2 och s_2^2. Eftersom de skattar samma σ^2 väger vi ihop dem med frihetsgraderna som vikter:

$$\boxed{s^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

Det större stickprovet får större vikt. Nämnaren $n_1 + n_2 - 2$ är också antalet frihetsgrader i intervallet nedan.

---

4. Konfidensintervallet

$$\boxed{I_{\mu_1 - \mu_2} = [\bar{x} - \bar{y} \mp t_{\alpha/2}(n_1 + n_2 - 2) \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}]}$$

Är standardavvikelserna **kända** i förväg, σ_1 och σ_2, behövs varken sammanvägningen eller t -fördelningen:

$$I_{\mu_1 - \mu_2} = \left[\bar{x} - \bar{y} \mp \lambda_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

På räknaren: `menu` → Statistics → Confidence Intervals → 2-Sample t Interval med **Pooled: Yes**.

---

5. Par eller två stickprov?

	Stickprov i par	**Två stickprov**
Försöksplan	båda metoderna på **samma** objekt	olika objekt i de två grupperna
Antal observationer	lika många i båda	n_1 och n_2 får skilja sig
Analyseras	differenserna z_i som ett stickprov	två stickprov med sammanvägd s
Variation mellan objekt	tar ut sig	hamnar i s
Frihetsgrader	$n - 1$	$n_1 + n_2 - 2$

⚠ Det är försöksplanen som avgör

Frågan är inte vilken formel som ger bäst resultat, utan hur datamaterialet samlades in. Finns en naturlig parning — samma person, samma dag, samma bit material — är det stickprov i par. Är grupperna olika individer är det två stickprov.

⚠ Vanliga fel

- **Frihetsgraderna.** Två stickprov ger $n_1 + n_2 - 2$, inte $n - 1$.
- **Medelvärdet av s_1 och s_2.** Den sammanvägda skattningen viktar *varianserna* med frihetsgraderna; ta roten först efteråt.
- **Pooled: No på räknaren.** Kursens modell har samma σ i båda grupperna, alltså **Pooled: Yes**.

---

6. Lösta lektionsuppgifter

Lektion 13, övning 3a och c (Vännman 8.23) — Blodtryck i två grupper

För att undersöka om en medicin påverkar blodtrycket mättes det i två oberoende grupper: 50 personer (x) och 25 personer (y). Resultat: $\bar{x} = 148{,}2$, $s_x = 10{,}0$ och $\bar{y} = 151{,}7$, $s_y = 8{,}0$.

a) Beräkna ett $95\ \%$ konfidensintervall för $\mu_1 - \mu_2$ och tolka det.

c) Ett nytt försök, där 25 patienter mättes före och efter behandlingen, gav intervallet $[1{,}24;\ 2{,}56]$ för den förväntade ökningen (se Stickprov i par). Varför ger försöken olika svar på frågan om medicinen höjer blodtrycket, och vilken slutsats kan dras?

Lösning

Modell. x_i observationer från $N(\mu_1, \sigma)$, $i = 1, \ldots, 50$; y_j observationer från $N(\mu_2, \sigma)$, $j = 1, \ldots, 25$; alla oberoende.

a) Sammanvägd varians:

$$s^2 = \frac{49 \cdot 10{,}0^2 + 24 \cdot 8{,}0^2}{73} = \frac{6436}{73} \approx 88{,}16, \qquad s \approx 9{,}39$$

Med $t_{0{,}025}(73) \approx 1{,}993$:

$$I_{\mu_1-\mu_2} = [-3{,}5 \mp 1{,}993 \cdot 9{,}39\sqrt{\frac{1}{50}+\frac{1}{25}}] = [-3{,}5 \mp 4{,}58] \approx \boxed{[-8{,}08;\ 1{,}08]}$$

0 ingår, så försöket ger ingen tydlig indikation på att grupperna skiljer sig i genomsnitt.

c) Blodtrycket varierar mycket mellan personer ($s \approx 9$), och i tvåstickprovsförsöket hamnar den variationen i intervallets bredd. I parförsöket mäts samma person två gånger; variationen mellan personer tar ut sig och det som återstår är den lilla, stabila förändringen ($s_z = 1{,}6$). Parförsöket är därför mycket känsligare och kan påvisa en höjning som tvåstickprovsförsöket inte ser. Slutsatsen dras från parförsöket: med $95\ \%$ säkerhet höjer medicinen blodtrycket, i genomsnitt med mellan $1{,}24$ och $2{,}56$ enheter.

På räknaren. 2-Sample t Interval med **Stats** ($\bar{x}_1$ = `148.2`, `sx1` = `10`, `n1` = `50`, $\bar{x}_2$ = `151.7`, `sx2` = `8`, `n2` = `25`) och **Pooled: Yes**.

Uppgiften (Vännman 8.23 a, i 2:a uppl. 8.25 a) är återgiven ur det handskrivna lösningsförslaget, som anger [−8,09; 1,10] med $t \approx 1{,}996$ och gränserna avrundade utåt; med $t_{0{,}025}(73) = 1{,}993$ blir det [−8,08; 1,08].

Lektion 13, övning 4 (Vännman 8.28) — Böjhållfasthet vid två temperaturer

Böjhållfastheten hos ett material mäts vid två temperaturer: 5 observationer x_i vid den ena och 6 observationer y_j vid 700°. Ur mätvärdena fås

$$\bar{x} = 137, \quad \sum x_i^2 = 94\,215, \qquad \bar{y} = 208, \quad \sum y_j^2 = 260\,236$$

Beräkna och tolka i ord ett $95\ \%$ konfidensintervall för skillnaden i förväntad böjhållfasthet.

Lösning

Modell. x_i observationer från $N(\mu_1, \sigma)$, $i = 1, \ldots, 5$; y_j observationer från $N(\mu_2, \sigma)$, $j = 1, \ldots, 6$; oberoende stickprov.

Varianser med räkneformen:

$$s_x^2 = \frac{94\,215 - 5 \cdot 137^2}{4} = 92{,}5, \qquad s_y^2 = \frac{260\,236 - 6 \cdot 208^2}{5} = 130{,}4$$

$$s^2 = \frac{4 \cdot 92{,}5 + 5 \cdot 130{,}4}{9} = \frac{1022}{9} \approx 113{,}6, \qquad s \approx 10{,}66$$

Intervall. $t_{0{,}025}(9) = 2{,}262$:

$$I_{\mu_1-\mu_2} = [137 - 208 \mp 2{,}262 \cdot 10{,}66\sqrt{\frac{1}{5} + \frac{1}{6}}] = [-71 \mp 14{,}6] \approx \boxed{[-85{,}6;\ -56{,}4]}$$

Tolkning. 0 ingår inte. Med 95 % säkerhet är den genomsnittliga böjhållfastheten mellan 56,4 och 85,6 enheter högre vid 700°.

På räknaren. 2-Sample t Interval med **Data** (de två listorna) och **Pooled: Yes**.

Uppgiften (Vännman 8.28, i 2:a uppl. 8.30) är återgiven ur det handskrivna lösningsförslaget, och intervallet stämmer med det.

Lektion 14, övning 1 — Mäter två laboratorier likvärdigt?

Ett prov delades i två delar som skickades till två laboratorier. Vart och ett gjorde 7 oberoende mätningar (kodade enheter), som kan beskrivas som oberoende observationer av normalfördelade stokastiska variabler.

Lab 1	**10,084**	**9,984**	**10,076**	**9,894**	**10,105**	**9,**
Lab 2	10,036	10,060	10,207	10,207	10,077	10

Räknehjälp: $\bar{x}_1 = 10{,}0081$, $s_1 = 0{,}0917$; $\bar{x}_2 = 10{,}1380$, $s_2 = 0{,}0874$; differenser Lab 2 − Lab 1: $\bar{z} = 0{,}1299$, $s_z = 0{,}1379$.

a) Ange de fullständiga modellantagandena.

b) Finns en genomsnittlig systematisk skillnad mellan laboratorierna? Hur stor? Beräkna och tolka ett 95 % konfidensintervall.

c) Varför är det två stickprov och inte stickprov i par?

Lösning

a) x_{1i} = mätning i vid lab 1, observation från $N(\mu_1, \sigma)$; x_{2j} = mätning j vid lab 2, observation från $N(\mu_2, \sigma)$; $i, j = 1, \ldots, 7$, alla oberoende och med samma σ.

c) Mätningarna vid de två laboratorierna hör inte ihop parvis: mätning nr 1 vid lab 1 har ingenting gemensamt med mätning nr 1 vid lab 2. Alla fjorton mätningar görs på samma prov. Ordningen i tabellen är godtycklig, så differenserna z i räknehjälpen saknar mening.

b) Sammanvägd standardavvikelse:

$$s = \sqrt{\frac{6 \cdot 0{,}0917^2 + 6 \cdot 0{,}0874^2}{12}} \approx 0{,}0896$$

Med $t_{0{,}025}(12) = 2{,}179$:

$$I_{\mu_2 - \mu_1} = [0{,}1299 \mp 2{,}179 \cdot 0{,}0896\sqrt{\frac{1}{7} + \frac{1}{7}}] = [0{,}1299 \mp 0{,}1043] \approx \boxed{[0{,}0255;\ 0{,}2342]}$$

0 ingår inte. Med $95\ \%$ säkerhet finns en systematisk skillnad: lab 2 ger i genomsnitt mellan $0{,}026$ och $0{,}234$ enheter högre värden än lab 1.

På räknaren. 2-Sample t Interval med **Pooled: Yes**; räknaren ger $\mu_1 - \mu_2$, alltså samma intervall med ombytta tecken.

Facit: [0,0255; 0,2342] för $\mu_2 - \mu_1$, med samma tolkning.

Se även

- Stickprov i par
- t-fördelningen
- Konfidensintervall
- Punktskattning
- Hypotesprövning

Resurser

- Wikipedia: Student’s t-test — independent two-sample

Hypotesprövning

19 okt. 2026, 5 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Konfidensintervall

Vid hypotesprövning ställer vi upp en nollhypotes och prövar om stickprovet talar tillräckligt starkt emot den. Signifikansnivån anger hur stor risk vi accepterar att förkasta en nollhypotes som i själva verket är sann.

---

1. Noll- och mothypotes

Ett test är en beslutsregel för att välja mellan två påståenden om en okänd parameter, till exempel väntevärdet μ:

- **Nollhypotesen** H_0 är det "ingenting har hänt"-läge vi utgår från, till exempel $H_0 : \mu = \mu_0$.
- **Mothypotesen** H_1 är det vi vill kunna påvisa, till exempel $H_1 : \mu > \mu_0$.

> **Det vi vill visa läggs i mothypotesen**
>
> Testet är byggt så att vi försöker **förkasta** H_0 med en liten felrisk. Kan vi förkasta H_0 har vi påvisat H_1, och vi har gjort det på ett sätt där risken att ha fel är liten om H_0 skulle vara sann. Kan vi inte förkasta H_0 har vi inte bevisat H_0 — vi har bara inte tillräckligt starka belägg mot den.

---

2. Testvariabel och kritiskt område

Testet bygger på en **testvariabel**, ett tal vi räknar ut ur stickprovet. För hypoteser om μ i en normalfördelning är det naturligt att använda medelvärdet $\bar{x}$, eller en standardiserad variant av det.

Vi bestämmer i förväg ett **kritiskt område**: de värden på testvariabeln som talar så starkt emot H_0 att vi förkastar den. Gränsen kallas den **kritiska gränsen** k. Med $H_1 : \mu > \mu_0$ blir testet

$$\text{förkasta } H_0 \text{ om } \bar{x} > k$$

Var k hamnar bestäms av signifikansnivån, se Signifikansnivå och styrka.

---

3. Två sorters fel

Beslutet kan bli fel på två sätt:

	H_0 **sann**	H_0 **falsk**
Förkasta H_0	fel av första slaget, sannolikhet α	rätt beslut, sannolikhet = styrkan
Förkasta inte H_0	rätt beslut	fel av andra slaget

Signifikansnivån α är sannolikheten för fel av första slaget, $P(\text{förkasta } H_0 \mid H_0 \text{ sann})$. Vi väljer den själva i förväg, typiskt $5\ \%$ eller $1\ \%$. **Styrkan** är sannolikheten att förkasta H_0 när den är falsk, alltså att upptäcka en verklig avvikelse.

---

4. Arbetsgång

Så ställer vi upp ett test

1. **Modellantaganden.** Inför och definiera de stokastiska variablerna och ange deras fördelning, till exempel $\xi_i \in N(\mu, \sigma)$, oberoende, $i = 1, \ldots, n$.
2. **Hypoteser.** H_0 och H_1, med det vi vill visa i H_1.
3. **Testvariabel och strategi.** Välj testvariabel och bestäm det kritiska området utifrån α: "förkasta H_0 om ...".
4. **Slutsats.** Stoppa in observationerna och formulera slutsatsen i ord, med signifikansnivån.

Vilken testvariabel och vilket kritiskt område som gäller beror på hypoteserna: Enkla hypoteser har ett enda värde i varje hypotes, Sammansatta hypoteser har ett helt intervall i mothypotesen. I stället för en kritisk gräns kan vi också använda Direktmetoden, och testet kan kopplas direkt till ett konfidensintervall.

På räknaren: Stat Tests

`menu` → **Statistics** → **Stat Tests** har färdiga test för kursens situationer. Välj **Alternate Hyp** så att den stämmer med H_1 ($\mu \neq \mu_0$, $\mu < \mu_0$ eller $\mu > \mu_0$).

Situation	Test
μ, känt σ	z Test
μ, okänt σ	t Test
$\mu_1 - \mu_2$, två stickprov med samma σ	2-Sample t Test, **Pooled: Yes**
μ för differenser, stickprov i par	t Test på differenslistan
p i en binomialfördelning (normalapproximation)	1-Prop z Test

Räknaren svarar med testvariabeln (`z` eller `t`) och `PVal`, alltså direktmetodens α_0. Den ger ingen kritisk gräns k och ingen styrka — dem räknar vi själva med `invNorm` / `invt` och `normCdf`. Steg 1, 2 och 4 i arbetsgången ovan ska skrivas ut på tentan oavsett.

Exempel — Fyller maskinen för mycket?

Påsfyllningen i Statistisk inferens: $\xi_i \in N(\mu,\ 4)$ oberoende, $n = 9$ och $\bar{x} = 502{,}9$ g. Pröva på nivån $5\ \%$ om maskinen fyller mer än 500 g i genomsnitt.

Lösning

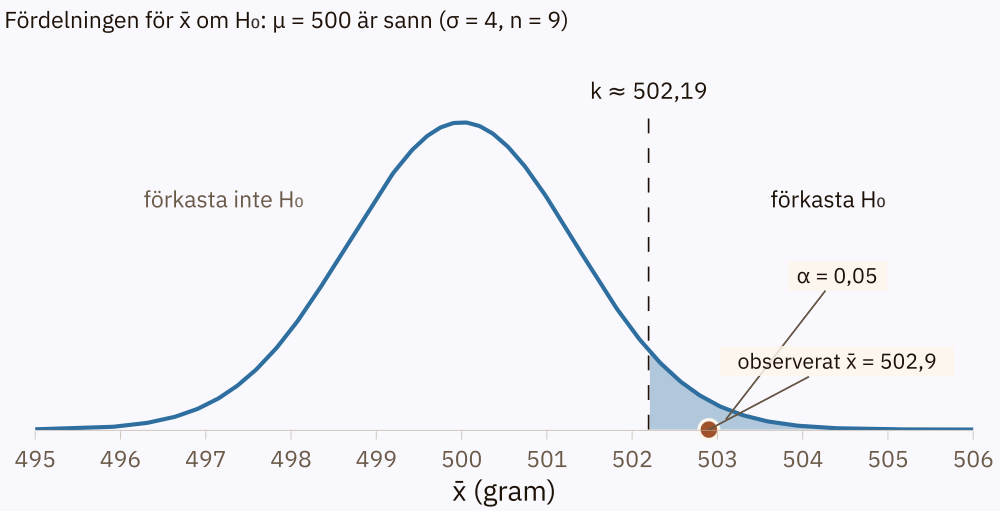
1. Modell. ξ_i = vikten hos påse i, $\xi_i \in N(\mu,\ 4)$ oberoende, $i = 1, \ldots, 9$.

2. Hypoteser. $H_0 :\ \mu = 500$ mot $H_1 :\ \mu > 500$.

3. Testvariabel och strategi. Om H_0 är sann är $\bar{\xi} \in N(500,\ 4/3)$. Förkasta H_0 om $\bar{x} > k$, där

$$k = 500 + \lambda_{0{,}05} \cdot \frac{4}{3} = 500 + 1{,}6449 \cdot 1{,}333 \approx 502{,}19$$

4. Slutsats. $\bar{x} = 502{,}9 > 502{,}19$, så H_0 **förkastas** på nivån $5\ \%$. Maskinen fyller signifikant mer än 500 g.



Med direktmetoden blir $\alpha_0 = P(\bar{\xi} \geq 502{,}9 \mid \mu = 500) = 1 - \Phi(2{,}175) \approx 0{,}015 < 0{,}05$, samma beslut.

På räknaren. Stat Tests → z Test med $\mu_0 = 500$, $\sigma = 4$, $\bar{x} = 502{,}9$, $n = 9$ och Alternate Hyp $\mu > \mu_0$: `z` $\approx 2{,}175$, `PVal` $\approx 0{,}0148$.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

⚠ Vanliga fel

- **Det man vill visa i H_0.** Det som ska påvisas hör hemma i H_1; H_0 är det vi försöker förkasta.
- **"Vi har visat H_0."** Att inte förkasta H_0 betyder bara att beläggen inte räckte.
- **Välja H_1 efter att ha sett data.** Riktningen ska bestämmas av frågan, inte av vilken sida $\bar{x}$ hamnade på.
- **Ensidigt eller tvåsidigt.** $H_1 : \mu \neq \mu_0$ kräver $\lambda_{\alpha/2}$, ensidiga H_1 kräver λ_α.

Se även

- Konfidensintervall
- Enkla hypoteser
- Signifikansnivå och styrka
- Sammansatta hypoteser
- Direktmetoden
- Konfidensintervall och hypotesprövning

Resurser

- Khan Academy: Significance tests (hypothesis testing)
- Wikipedia: Hypotesprövning

Enkla hypoteser

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Hypotesprövning

En enkel hypotes anger ett enda bestämt värde på parametern. Eftersom både noll- och mothypotes då är fullständigt specificerade kan vi beräkna både risken att felaktigt förkasta och risken att felaktigt behålla nollhypotesen.

---

1. Situationen

Låt $\xi_1, \ldots, \xi_n$ vara oberoende och $\in N(\mu, \sigma)$ med känt σ. Vi prövar

$$H_0 : \ \mu = \mu_0 \qquad \text{mot} \qquad H_1 : \ \mu = \mu_1, \qquad \mu_1 > \mu_0$$

Båda hypoteserna innehåller bara ett värde, så båda är **enkla**. Situationen är inte särskilt realistisk — sällan finns bara två möjliga värden — men den visar hela maskineriet i hypotesprövningen utan komplikationer.

---

2. Testet

Stora värden på $\bar{x}$ talar för H_1, så testet blir: förkasta H_0 om $\bar{x} > k$. Under H_0 är $\bar{\xi} \in N(\mu_0,\ \sigma/\sqrt{n})$, och signifikansnivån α ger

$$\alpha = P(\bar{\xi} > k \mid \mu = \mu_0) = P\left(\frac{\bar{\xi} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) \quad \Longrightarrow \quad \boxed{k = \mu_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}}$$

där λ_α är kvantilen i $N(0, 1)$ med $P(\eta > \lambda_\alpha) = \alpha$, på räknaren `invNorm(1-α)`.

---

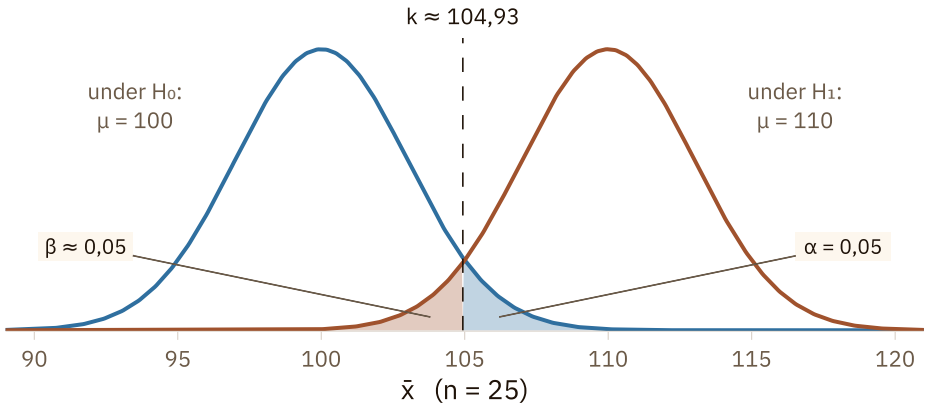
3. Båda felriskerna

Med en enkel mothypotes kan vi också räkna ut sannolikheten att **inte** förkasta H_0 fast H_1 gäller, felet av andra slaget:

$$\beta = P(\bar{\xi} \leq k \mid \mu = \mu_1) = \Phi(\frac{k - \mu_1}{\sigma/\sqrt{n}})$$

Styrkan är $1 - \beta$. På räknaren är β = `normCdf(-∞,k,μ1,σ/√(n))` — bocka i **Draw** så ritas arean.

Figuren visar fördelningen för $\bar{\xi}$ under respektive hypotes i övning 4 nedan ($\mu_0 = 100$, $\mu_1 = 110$, $\sigma/\sqrt{n} = 3$), med α till höger om k under H_0 och β till vänster om k under H_1:



Avvägningen

Flyttar vi k åt höger minskar α men β ökar, och tvärtom. Det enda sättet att minska **båda** är att öka n: då krymper $\sigma/\sqrt{n}$, kurvorna blir smalare och överlappar mindre. Se Signifikansnivå och styrka.

Vanliga fel

- **Blanda ihop λ_α och λ_β.** k ligger λ_α medelfel från μ_0 och λ_β medelfel från μ_1.
- **Tro att $\alpha + \beta = 1$.** Båda kan göras små samtidigt, men bara genom större n.

4. Lösta lektionsuppgifter

Lektion 14, övning 4 (Vännman 9.3) — Hur många observationer?

Observationerna är oberoende och $\in N(\mu,\ 15)$. Vi prövar $H_0:\ \mu = 100$ mot $H_1:\ \mu = 110$ med testet "förkasta H_0 om $\bar{x} > k$".

a) Bestäm antalet observationer n och den kritiska gränsen k så att testet får signifikansnivån $5\ \%$ och styrkan minst $95\ \%$.

b) Hur blir det nya testet?

Lösning

Fördelning. $\bar{\xi} \in N(\mu,\ 15/\sqrt{n})$. Den kritiska gränsen beror av n, så vi får två villkor på de två obekanta.

Villkor 1, signifikansnivån.

$$0{,}05 = P(\bar{\xi} > k \mid \mu = 100) \quad\Longrightarrow\quad \frac{k-100}{15/\sqrt{n}} = \lambda_{0{,}05} \approx 1{,}6449 \quad\Longrightarrow\quad k = 100 + 1{,}6$$

Villkor 2, styrkan.

$$0{,}95 = P(\bar{\xi} > k \mid \mu = 110) \quad\Longrightarrow\quad \frac{110-k}{15/\sqrt{n}} = \lambda_{0{,}05} \quad\Longrightarrow\quad k = 110 - 1{,}6449 \cdot \frac{15}{\sqrt{n}}$$

Lös ut n. Sätt uttrycken lika:

$$2 \cdot 1{,}6449 \cdot \frac{15}{\sqrt{n}} = 10 \quad\Longrightarrow\quad n = (3 \cdot 1{,}6449)^2 \approx 24{,}35$$

n måste vara ett heltal. Med $n = 24$ och k från villkor 1 blir styrkan $\approx 0{,}948 < 0{,}95$; med $n = 25$ blir den $\approx 0{,}954$. Alltså $\boxed{n = 25}$, och villkor 1 ger

$$k = 100 + 1{,}6449 \cdot \frac{15}{5} \approx \boxed{104{,}93}$$

b) Det nya testet: förkasta H_0 om $\bar{x} > 104{,}93$, baserat på 25 observationer. Det är testet i figuren i avsnitt 3.

På räknaren. $\lambda_{0{,}05}$ = `invNorm(0.95)`. Styrkorna för $n = 24$ och $n = 25$ kontrolleras med `normCdf(k,∞,110,15/√(n))`, där `k` är gränsen för respektive n.

Uppgiften (Vännman 9.3) är återgiven ur det handskrivna lösningsförslaget ($n = 25$, $k \approx 104{,}93$), och svaret stämmer med det.

Se även

- Hypotesprövning
- Signifikansnivå och styrka
- Sammansatta hypoteser
- Normalfördelning

Resurser

- Wikipedia: Type I and type II errors

Signifikansnivå och styrka

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #begrepp

> **Kurs:** S0008M **Förkunskaper:** Hypotesprövning, Enkla hypoteser

Signifikansnivån är sannolikheten att förkasta en sann nollhypotes, och styrkan är sannolikheten att förkasta en falsk. Vi väljer signifikansnivån själva; styrkan styr vi med antalet observationer.

1. Definitionerna

$$\boxed{\alpha = P(\text{förkasta } H_0 \mid H_0 \text{ sann})}$$

$$\boxed{\text{styrkan} = P(\text{förkasta } H_0 \mid H_1 \text{ sann})}$$

Signifikansnivån är risken för **fel av första slaget** i hypotesprövningen. Styrkan är $1-\beta$, där β är risken för **fel av andra slaget**. Ett bra test har litet α och stor styrka.

Är mothypotesen sammansatt, till exempel $\mu > \mu_0$, har styrkan ett värde för varje μ i mothypotesen. Då blir den en funktion, styrkefunktionen $S(\mu)$.

2. Vad påverkar styrkan

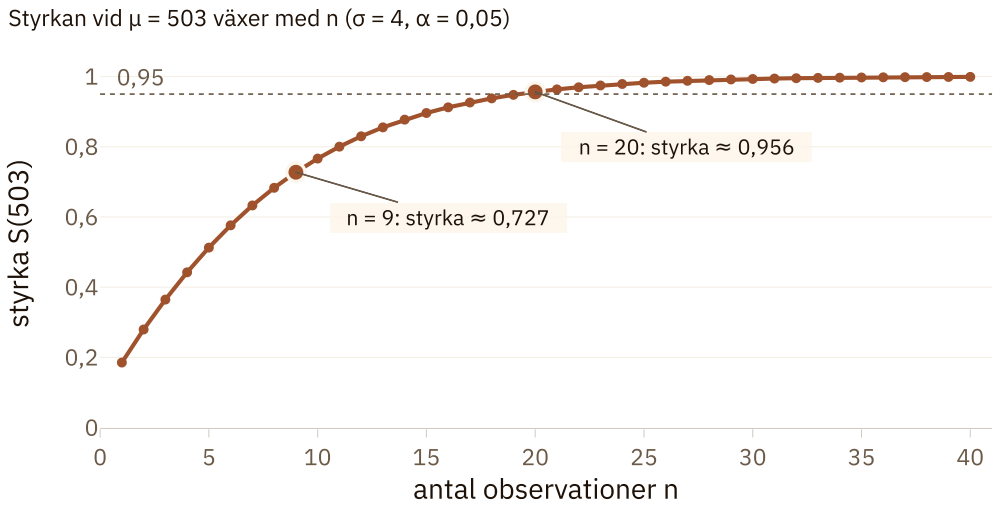
För testet "förkasta $H_0: \mu = \mu_0$ om $\bar{x} > k$" med $\bar{\xi} \in N(\mu,\ \sigma/\sqrt{n})$ är styrkan i ett värde $\mu_1 > \mu_0$

$$P(\bar{\xi} > k \mid \mu = \mu_1) = 1 - \Phi\left(\frac{k-\mu_1}{\sigma/\sqrt{n}}\right)$$

Ur formeln syns tre saker:

Ändring	Effekt på styrkan
Mindre α (större k)	minskar

Ändring	Effekt på styrkan
Större avstånd $\mu_1 - \mu_0$	ökar — stora avvikelser är lättare att upptäcka
Större n	ökar — $\sigma/\sqrt{n}$ krymper



Figuren gäller påsfyllningen i Hypotesprövning ($\mu_0 = 500$, $\sigma = 4$, $\alpha = 0{,}05$) och visar styrkan i $\mu_1 = 503$. Med $n = 9$ upptäcks en avvikelse på 3 g bara i knappt tre fall av fyra; med $n = 20$ i minst $95\,\%$ av fallen.

---

3. Att bestämma n

Kräver vi både en viss signifikansnivå α och minst styrkan $1 - \beta$ i ett visst μ_1 får vi två villkor på k:

$$k = \mu_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}, \qquad k = \mu_1 - \lambda_\beta \frac{\sigma}{\sqrt{n}}$$

Sätter vi dem lika och löser ut n:

$$\boxed{n = \left(\frac{(\lambda_\alpha + \lambda_\beta)\,\sigma}{\mu_1 - \mu_0} \right)^2}$$

> ⚠ **Avrunda uppåt och bestäm k efteråt**
>
> n blir sällan ett heltal. Avrunda till **närmast högre** heltal, så att styrkan blir minst den krävda, och bestäm sedan k ur signifikansvillkoret med det nya n, så att signifikansnivån blir exakt rätt.

---

4. Lösta lektionsuppgifter

Lektion 15, övning 4 (Vännman 9.6) — Livslängd hos däck

Livslängden hos däck antas vara $\xi_i \in N(\mu,\ 900)$, oberoende, $i = 1, \dots, n$. Vi vill pröva $H_0: \mu = 5400$ mot $H_1: \mu > 5400$ med testet "förkasta H_0 om $\bar{x} > k$". Bestäm n och k så att signifikansnivån blir $0{,}01$ och styrkan minst $0{,}95$ om $\mu = 6300$.

Lösning

Fördelning. $\bar{\xi} \in N(\mu,\ 900/\sqrt{n})$.

Signifikansnivån. $\lambda_{0{,}01} \approx 2{,}3263$:

$$0{,}01 = P(\bar{\xi} > k \mid \mu = 5400) \quad \Longrightarrow \quad k = 5400 + 2{,}3263 \cdot \frac{900}{\sqrt{n}} \tag{1}$$

Styrkan i $\mu = 6300$. $\lambda_{0{,}05} \approx 1{,}6449$:

$$0{,}95 = P(\bar{\xi} > k \mid \mu = 6300) \quad \Longrightarrow \quad k = 6300 - 1{,}6449 \cdot \frac{900}{\sqrt{n}} \tag{2}$$

Lös ut n. (1) = (2) ger

$$\frac{(2{,}3263 + 1{,}6449) \cdot 900}{\sqrt{n}} = 900 \quad \Longrightarrow \quad n = 3{,}9712^2 \approx 15{,}77$$

Med $n = 15$ blir styrkan $\approx 0{,}939$, för lite. Med $\boxed{n = 16}$ blir den $\approx 0{,}953$. Villkor (1) ger sedan

$$k = 5400 + 2{,}3263 \cdot \frac{900}{\sqrt{16}} \approx \boxed{5923{,}4}$$

På räknaren. $\lambda_{0{,}01}$ = `invNorm(0.99)`, $\lambda_{0{,}05}$ = `invNorm(0.95)`. Styrkan för $n = 16$: `normCdf(5923.4,∞,6300,900/4)` ≈ `0.953`.

Uppgiften (Vännman 9.6) är återgiven ur det handskrivna lösningsförslaget ($n = 16$, $k \approx 5923{,}42$), och svaret stämmer med det.

Exempel — Styrka och antal påsar

Påsfyllningen: $H_0 : \mu = 500$ mot $H_1 : \mu > 500$, $\sigma = 4$ känd, $\alpha = 0{,}05$.

a) Beräkna styrkan i $\mu = 503$ med $n = 9$.

b) Hur många påsar behövs för att styrkan i $\mu = 503$ ska bli minst $0{,}95$?

Lösning

a) Med $n = 9$ är $k \approx 502{,}19$ (se Hypotesprövning), och

$$P(\bar{\xi} > 502{,}193 \mid \mu = 503) = 1 - \Phi\left(\frac{502{,}193 - 503}{4/3}\right) = \Phi(0{,}605) \approx \boxed{0{,}727}$$

b) $\lambda_{0{,}05} = 1{,}6449$ för både α och $\beta = 0{,}05$:

$$n = \left(\frac{(1{,}6449 + 1{,}6449) \cdot 4}{503 - 500}\right)^2 \approx 19{,}24 \quad \Rightarrow \quad \boxed{n = 20}$$

På räknaren. a) `1-normCdf(-∞,502.19,503,4/3)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Vanliga fel

- **β är inte $1 - \alpha$.** α och β räknas under olika hypoteser och summerar inte till 1.
- **Styrka i μ_0.** Där är "styrkan" α, inte något stort tal.
- **Avrunda n nedåt.** $19{,}24$ ger $n = 20$; med 19 når styrkan inte kravet.

Se även

- Hypotesprövning
- Enkla hypoteser
- Styrkefunktion

- Sammansatta hypoteser

Resurser

- Wikipedia: Power of a test
- Wikipedia: Statistical significance

Sammansatta hypoteser

19 okt. 2026, 6 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Enkla hypoteser

En sammansatt hypotes omfattar ett helt intervall av parametervärden, till exempel $\mu > \mu_0$. Testets förmåga att upptäcka en avvikelse beror då på var i intervallet det sanna värdet ligger, och beskrivs av en styrkefunktion.

1. Tre mothypoteser

Nollhypotesen är fortfarande enkel, $H_0 : \mu = \mu_0$, men mothypotesen innehåller ett helt intervall. Vilket kritiskt område vi väljer beror på vad vi vill kunna upptäcka:

Mothypotes	Vi vill upptäcka att μ …	Förkasta H_0 om (σ känd)
$H_1 : \mu > \mu_0$	har ökat	$\bar{x} > \mu_0 + \lambda_\alpha \sigma/\sqrt{n}$
$H_1 : \mu < \mu_0$	har minskat	$\bar{x} < \mu_0 - \lambda_\alpha \sigma/\sqrt{n}$
$H_1 : \mu \neq \mu_0$	har ändrats åt något håll	$\lvert\bar{x} - \mu_0\rvert > \lambda_{\alpha/2} \sigma/\sqrt{n}$

De två första testen är **ensidiga**, det tredje är **tvåsidigt**: där delas α lika mellan de två svansarna. Gränserna härleds som i Enkla hypoteser, eftersom signifikansnivån bara beror på fördelningen under H_0.

Styrkan har nu ett värde för varje μ i mothypotesen och blir en styrkefunktion $S(\mu)$.

2. Standardiserad testvariabel

Att förkasta när $\bar{x} > \mu_0 + \lambda_\alpha \sigma/\sqrt{n}$ är samma sak som att förkasta när

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > \lambda_\alpha$$

Vi kan alltså lika gärna använda z som testvariabel. Den är en observation från $N(0, 1)$ när H_0 är sann, och den kritiska gränsen blir λ_α oavsett μ_0, σ och n.

---

3. Okänd σ – t-test

Oftast känner vi inte σ i förväg utan måste skatta den med stickprovets standardavvikelse s. Vi ersätter då σ med s i z:

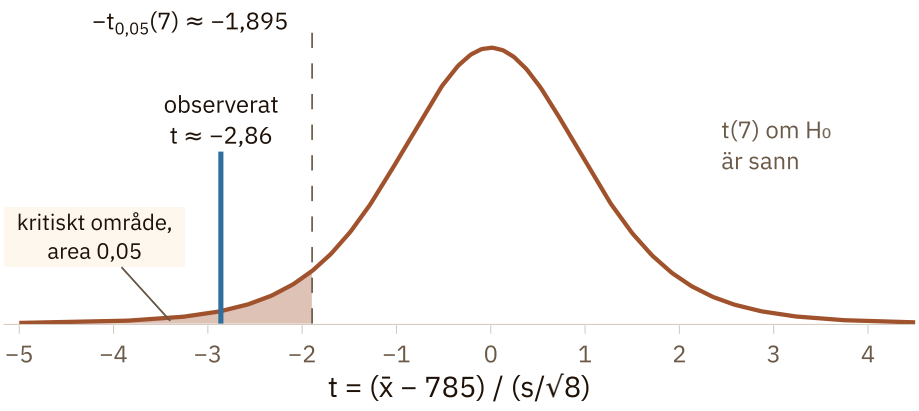
$$\boxed{t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$

Den här testvariabeln är en observation från t-fördelningen med $n - 1$ frihetsgrader när H_0 är sann. Den kritiska gränsen tas därför ur t-tabellen i stället för ur normalfördelningstabellen:

Mothypotes	Förkasta H_0 om (σ okänd)
$H_1 : \mu > \mu_0$	$t > t_\alpha(n-1)$
$H_1 : \mu < \mu_0$	$t < -t_\alpha(n-1)$
$H_1 : \mu \neq \mu_0$	$\lvert t \rvert > t_{\alpha/2}(n-1)$

På räknaren: `menu` → **Statistics** → **Stat Tests** → **z Test** när σ är känd och **t Test** när den är okänd. Välj **Stats** om du har $\bar{x}$ och s, **Data** om du har mätvärdena i en lista, och **Alternate Hyp** efter H_1. De kritiska gränserna är λ_α = `invNorm(1-α)` och $t_\alpha(f)$ = `invt(1-α,f)`.

Figuren visar det kritiska området för det ensidiga t-testet i övning 6 nedan, där $t(7)$ är fördelningen för testvariabeln om H_0 är sann:



⚠ Vanliga fel

- **λ när σ är skattad.** Med s i stället för σ blir testvariabeln t-fördelad med $n - 1$ frihetsgrader.
- **Absolutbelopp i ensidigt test.** $|t| > t_{\alpha/2}$ gäller bara $H_1 : \mu \neq \mu_0$; ensidiga test jämför t med $\pm t_\alpha$ utan belopp.

---

4. Lösta lektionsuppgifter

Lektion 15, övning 1 — Coca-Colaburkar

Låt $x_1, \ldots, x_{10}$ vara observerade värden på $\xi_1, \ldots, \xi_{10}$, där $\xi_i \in N(\mu,\ 1{,}72)$. Enheten är ml. Väntevärdet bör vara 330. Bestäm ett test som ska upptäcka om $\mu < 330$ och som har $5\ \%$ felrisk om μ är 330.

a) Sätt upp noll- och mothypotes.

b) Bestäm testet (strategin).

c) Vad blir slutsatsen om $\bar{x} = 329{,}4$?

d) Vad blir slutsatsen om $\bar{x} = 328{,}5$?

Lösning

a) Vi vill kunna påvisa att burkarna innehåller för lite, så det läggs i mothypotesen:

$$H_0 : \mu = 330, \qquad H_1 : \mu < 330$$

b) Små värden på $\bar{x}$ talar för H_1: förkasta H_0 om $\bar{x} < k$. Under H_0 är $\bar{\xi} \in N(330,\ 1{,}72/\sqrt{10}) = N(330,\ 0{,}544)$, och

$$0{,}05 = P\left(\bar{\xi} < k \mid \mu = 330\right) \quad \Longrightarrow \quad k = 330 - \lambda_{0{,}05} \cdot \frac{1{,}72}{\sqrt{10}} = 330 - 1{,}6449 \cdot 0{,}544 \approx 3$$

Testet: $\boxed{\text{förkasta } H_0 \text{ om } \bar{x} < 329{,}1}$.

c) $\bar{x} = 329{,}4$ ligger inte under $329{,}1$. H_0 **kan inte förkastas** på $5\ \%$ signifikansnivå — vi har inte påvisat att burkarna innehåller för lite.

d) $\bar{x} = 328{,}5 < 329{,}1$. H_0 **förkastas** på $5\ \%$ signifikansnivå: med $5\ \%$ felrisk kan vi påstå att $\mu < 330$.

På räknaren. k direkt: `invNorm(0.05,330,1.72/√(10))` ≈ `329.105`. Eller z Test med **Stats**: `μ0: 330`, `σ: 1.72`, `x̄: 328.5`, `n: 10`, **Alternate Hyp** `μ < μ0` ger $z \approx -2{,}76$ och `PVal` ≈ $0{,}003 < 0{,}05$ — samma slutsats som d), se Direktmetoden.

Facit: a) $H_0:\ \mu = 330$, $H_1:\ \mu < 330$; b) förkasta H_0 om $\bar{x} < 329{,}1$; c) kan inte förkastas; d) förkastas.

Lektion 15, övning 6 (Vännman 9.16) — t-test med åtta mätvärden

Åtta oberoende mätvärden, $\xi_i \in N(\mu, \sigma)$ med okänt σ, gav $\bar{x} = 782{,}875$ och $s = 2{,}1002$. Pröva $H_0:\ \mu = 785$ mot $H_1:\ \mu < 785$ på signifikansnivån $5\ \%$.

Lösning

Testvariabel. σ är okänd, så vi använder

$$t = \frac{\bar{x} - 785}{s/\sqrt{8}}$$

som är en observation från $t(7)$ om H_0 är sann.

Strategi. Små värden talar för H_1: förkasta H_0 om $t < -t_{0{,}05}(7) \approx -1{,}895$.

Slutsats.

$$t = \frac{782{,}875 - 785}{2{,}1002/\sqrt{8}} \approx -2{,}86 < -1{,}895$$

$\boxed{H_0 \text{ förkastas}}$ på $5\ \%$ signifikansnivå: vi har påvisat att $\mu < 785$.

På räknaren. t Test med **Stats**: `μ0: 785`, `x̄: 782.875`, `sx: 2.1002`, `n: 8`, **Alternate Hyp** `μ < μ0`. Räknaren ger $t \approx -2{,}86$ och `PVal` $\approx 0{,}012$. Gränsen: `invt(0.05,7)` $\approx$ `-1.895`.

Uppgiften (Vännman 9.16) är återgiven ur det handskrivna lösningsförslaget ($t \approx -2{,}86 < -1{,}895$), och slutsatsen stämmer med det.

Lektion 16, övning 1 — Enheter som ska väga 500 gram

En maskin tillverkar enheter som bör väga 500 gram. För att se att den genomsnittliga vikten håller sig på 500 gram tas regelbundet ett stickprov av storlek 8. Bilda ett test baserat på 8 mätvärden $x_1, \ldots, x_8$ som ska upptäcka en genomsnittlig systematisk avvikelse från 500, alltså om $\mu \neq 500$. Använd $5\ \%$ signifikansnivå och anta att mätvärdena är normalfördelade.

a) Ange fullständiga modellantaganden och hypoteser.

b) Bestäm testet, det vill säga beslutsstrategin.

c) Ett stickprov gav $\bar{x} = 497{,}8$ gram och $s = 2{,}1$. Vad blir slutsatsen?

Lösning

a) ξ_i = vikten hos enhet nummer i, $i = 1, \ldots, 8$. De är oberoende och $\xi_i \in N(\mu, \sigma)$ med både μ och σ okända.

$$H_0 : \mu = 500, \qquad H_1 : \mu \neq 500$$

b) σ är okänd och mothypotesen tvåsidig. Med $t_{0{,}025}(7) \approx 2{,}365$:

$$\text{förkasta } H_0 \text{ om } \left| \frac{\bar{x} - 500}{s/\sqrt{8}} \right| > 2{,}365$$

c)

$$\left| \frac{497{,}8 - 500}{2{,}1/\sqrt{8}} \right| \approx |-2{,}96| = 2{,}96 > 2{,}365$$

$\boxed{H_0 \text{ förkastas}}$ på 5 % signifikansnivå: det finns en genomsnittlig systematisk avvikelse från 500 gram.

Samma slutsats med ett konfidensintervall finns i Konfidensintervall och hypotesprövning.

På räknaren. t Test med **Stats**: `μ0: 500`, `x̄: 497.8`, `sx: 2.1`, `n: 8`, **Alternate Hyp** `μ ≠ μ0`. Räknaren ger $t \approx -2{,}96$ och `PVal` $\approx 0{,}021 < 0{,}05$. Gränsen: `invt(0.975,7)` $\approx$ `2.365`.

Facit: a) $\xi_i \in N(\mu, \sigma)$, $H_0: \mu = 500$, $H_1: \mu \neq 500$*; b) förkasta* H_0 *om* $|\bar{x} - 500|/(s/\sqrt{8}) > 2{,}365$*; c)* H_0 *förkastas.*

Se även

- Hypotesprövning
- Enkla hypoteser
- Styrkefunktion
- Direktmetoden
- t-fördelningen
- Konfidensintervall och hypotesprövning

Resurser

- Khan Academy: Tests about a population mean
- Wikipedia: Student's t-test

Styrkefunktion

19 okt. 2026, 3 min läsning

UTKAST

#matematik #statistik #begrepp

> **Kurs:** S0008M **Förkunskaper:** Sammansatta hypoteser, Signifikansnivå och styrka

---

Styrkefunktionen $S(\mu)$ anger sannolikheten att ett test förkastar nollhypotesen, som funktion av det sanna värdet μ. I nollhypotesens värde är den lika med signifikansnivån, och den växer mot 1 ju längre in i mothypotesen μ ligger.

---

1. Definitionen

För ett test av $H_0 : \mu = \mu_0$ är **styrkefunktionen**

$$\boxed{S(\mu) = P\,(\text{förkasta } H_0 \mid \mu)}$$

Den är definierad för alla μ. I $\mu = \mu_0$ är $S(\mu_0) = \alpha$, signifikansnivån. För μ i mothypotesen är $S(\mu)$ testets styrka i just det värdet.

---

2. Styrkefunktionen för ett ensidigt test

För testet "förkasta H_0 om $\bar{x} < k$" med $\bar{\xi} \in N(\mu,\ \sigma/\sqrt{n})$ är

$$S(\mu) = P\,(\bar{\xi} < k \mid \mu) = \Phi\left(\frac{k-\mu}{\sigma/\sqrt{n}}\right)$$

Den är avtagande i μ: ju lägre det sanna väntevärdet ligger, desto troligare är det att $\bar{x}$ hamnar under k. För testet "förkasta om $\bar{x} > k$" blir $S(\mu) = 1 - \Phi((k-\mu)/(\sigma/\sqrt{n}))$, som är växande.

> **Formen på kurvan**
>
> Styrkefunktionen för ett ensidigt test är en S-formad kurva som går från nära 0 på nollhypotesens sida, genom α i μ_0, till nära 1 långt in i mothypotesen. Med större n blir övergången brantare: testet skiljer bättre mellan värden nära μ_0.

⚠ Vanliga fel

- **Räkna under H_0.** Styrkan i μ räknas med det μ, inte med μ_0; bara k kommer från H_0.
- **Fel riktning.** För "förkasta om $\bar{x} < k$" är $S(\mu) = \Phi(\cdot)$, för "förkasta om $\bar{x} > k$" är det $1 - \Phi(\cdot)$.

---

3. Lösta lektionsuppgifter

Lektion 15, övning 2 och 3 — Styrkan i Coca-Colatestet

I testet för Coca-Colaburkarna i Sammansatta hypoteser är $\xi_i \in N(\mu,\ 1{,}72)$, $n = 10$, och vi förkastar $H_0 :\ \mu = 330$ till förmån för $H_1 :\ \mu < 330$ om $\bar{x} < 329{,}1$.

Övning 2. Anta att påfyllnadsmekanismen har börjat fylla för lite, så att $\mu = 328$. Vad är sannolikheten att testet upptäcker att H_0 är falsk? Beräkna alltså styrkan i $\mu = 328$.

Övning 3. Skissa styrkefunktionen för testet.

Lösning

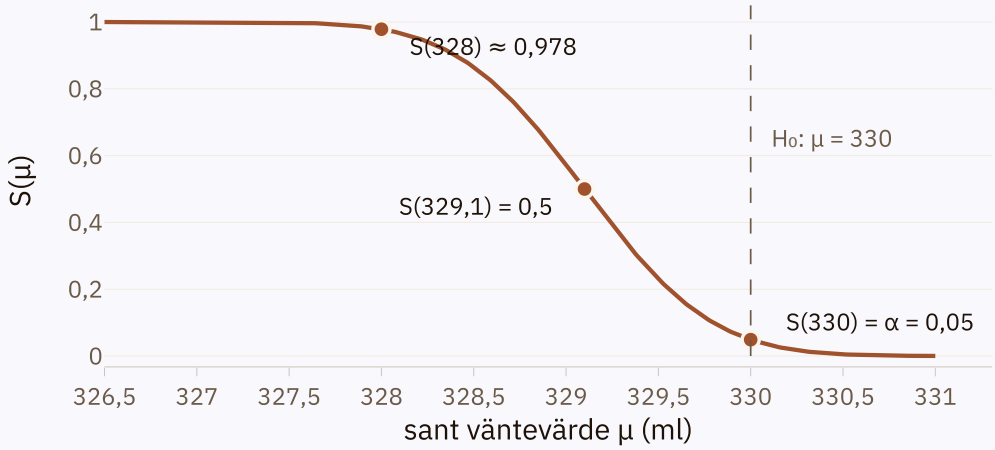
Övning 2. Med $\mu = 328$ är $\bar{\xi} \in N(328,\ 1{,}72/\sqrt{10}) = N(328,\ 0{,}544)$:

$$S(328) = P\left(\bar{\xi} < 329{,}1 \mid \mu = 328\right) = \Phi\left(\frac{329{,}1 - 328}{0{,}544}\right) = \Phi(2{,}02) \approx \boxed{0{,}978}$$

Testet upptäcker alltså en minskning till 328 ml med sannolikheten $97{,}8\ \%$.

På räknaren. `normCdf(-∞,329.1,328,1.72/√(10))` ≈ `0.978`.

Övning 3. Vi vet två punkter: $S(330) = \alpha = 0{,}05$ och $S(328) \approx 0{,}978$. Kurvan är avtagande i μ, går mot 1 när μ minskar och mot 0 när μ ökar. Den passerar $0{,}5$ precis i $\mu = k = 329{,}1$, där $\bar{x}$ lika gärna hamnar över som under gränsen:



På räknaren (hela kurvan). Definiera `s(m):=normCdf(-∞,329.1,m,1.72/√(10))` i en Calculator-sida och rita `s(x)` i **Graphs** med fönstret $326 \le x \le 331$.

Facit: $S(328) = 0{,}978$, *det vill säga styrkan är* $97{,}8\ \%$.

Se även

- Signifikansnivå och styrka
- Sammansatta hypoteser
- Enkla hypoteser
- Normalfördelning

Resurser

- Wikipedia: Power of a test

Direktmetoden

19 okt. 2026, 5 min läsning

UTKAST

#matematik #statistik #metod

> **Kurs:** S0008M **Förkunskaper:** Sammansatta hypoteser

Direktmetoden innebär att vi beräknar p-värdet — sannolikheten att få ett minst lika extremt utfall som det observerade om nollhypotesen är sann — och jämför det direkt med signifikansnivån. I stället för en kritisk gräns får vi då ett mått på hur starkt data talar emot nollhypotesen.

---

1. α_0-värdet

I stället för att först räkna ut den kritiska gränsen k och jämföra testvariabeln med den, räknar vi direkt ut

$$\boxed{\alpha_0 = P\,(\text{testvariabeln blir lika extrem eller extremare än det observerade} \mid H_0 \text{ sann})}$$

"Extrem" betyder i mothypotesens riktning. Med $H_1 : \mu < \mu_0$ och observerat $\bar{x}$ är $\alpha_0 = P\,(\bar{\xi} \leq \bar{x} \mid \mu = \mu_0)$.

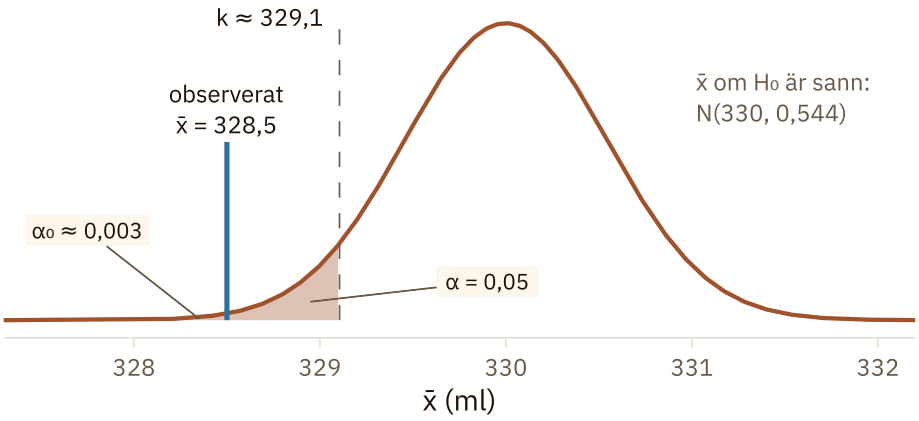
Vännman kallar talet α_0. Miniräknare och datorprogram kallar det **p-värdet**; på TI-Nspire heter raden `PVal` i resultatet från z Test, t Test och de andra testen under `menu` → **Statistics** → **Stat Tests**.

---

2. Beslutsregeln

$$\text{förkasta } H_0 \text{ om } \alpha_0 < \alpha$$

Det ger samma beslut som testet med kritisk gräns: $\alpha_0 < \alpha$ betyder precis att det observerade värdet ligger i det kritiska området.



Figuren visar Coca-Colatestet i uppgift A nedan. Arean till vänster om k är $\alpha = 0{,}05$. Arean till vänster om det observerade $\bar{x}$ är $\alpha_0 \approx 0{,}003$ – så liten att den knappt syns. Att $\bar{x}$ ligger till vänster om k är samma sak som att $\alpha_0 < \alpha$.

⚠ Tvåsidiga test

Med $H_1 : \mu \neq \mu_0$ fördelas α på båda svansarna. Räknar vi α_0 som sannolikheten i **en** svans ska den jämföras med $\alpha/2$. Miniräknarens p-värde för ett tvåsidigt test är redan dubblerat och jämförs med α.

✎ Vad direktmetoden inte ger

Direktmetoden räcker när vi bara vill veta om H_0 kan förkastas. Vill vi räkna ut testets styrka behöver vi den kritiska gränsen k.

α_0 säger också hur marginellt beslutet är: $\alpha_0 = 0{,}049$ och $\alpha_0 = 0{,}0001$ ger båda förkastning på $5\ \%$-nivån, men det senare är ett mycket starkare belägg mot H_0.

⚠ Vanliga fel

- **Svansen åt fel håll.** α_0 räknas i mothypotesens riktning: för $H_1 : \mu < \mu_0$ är det $P(\bar{\xi} \leq \bar{x} \mid H_0)$.
- **Tolka α_0 som $P(H_0$ sann$)$.** Det är sannolikheten för ett minst lika extremt utfall *om* H_0 är sann.

3. Lösta lektionsuppgifter

Inför lektion 16, uppgift A — Coca-Colatestet med direktmetoden

Använd direktmetoden på Coca-Colatestet i Sammansatta hypoteser: $\xi_i \in N(\mu,\ 1{,}72)$, $n = 10$, $H_0 :\ \mu = 330$ mot $H_1 :\ \mu < 330$ på 5 % signifikansnivå, och det observerade medelvärdet är $\bar{x} = 328{,}5$. Beräkna sannolikheten att få ett lika extremt eller extremare värde, $P(\bar{\xi} \le 328{,}5)$, givet att H_0 är sann, och jämför med signifikansnivån.

Lösning

Under H_0 är $\bar{\xi} \in N(330,\ 0{,}544)$, och "extremare" betyder mindre, eftersom H_1 säger $\mu < 330$:

$$\alpha_0 = P(\bar{\xi} \le 328{,}5 \mid \mu = 330) = \Phi\left(\frac{328{,}5 - 330}{0{,}544}\right) = \Phi(-2{,}76) \approx \boxed{0{,}003}$$

$\alpha_0 \approx 0{,}003 < 0{,}05$, så H_0 förkastas — samma slutsats som med den kritiska gränsen $329{,}1$. Figuren i avsnitt 2 visar just den här situationen.

På räknaren. `normCdf(-∞,328.5,330,1.72/√(10))` ≈ `0.0029`, eller `PVal` i z Test (se Sammansatta hypoteser).

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens. Slutsatsen stämmer också med facit till lektion 15, övning 1 d).

Lektion 16, övning 3 (Vännman 9.22) — Nio lyckade av elva

Ett försök upprepas 11 gånger oberoende av varandra, med sannolikheten p att lyckas varje gång. Utfallet blev 9 lyckade försök. Använd direktmetoden för att pröva $H_0 :\ p = 0{,}5$ på 5 % signifikansnivå mot

a) $H_1 :\ p > 0{,}5$,

b) $H_1 :\ p \neq 0{,}5$.

Lösning

Modell. ξ = antal lyckade försök, $\xi \in \text{Bin}(11,\ p)$. Under H_0 är $\xi \in \text{Bin}(11,\ 0{,}5)$.

α_0**.** Nio lyckade ligger klart i den högra svansen. Sannolikheten för lika extremt eller extremare utfall:

$$\alpha_0 = P(\xi \geq 9 \mid p = 0{,}5) = 1 - P(\xi \leq 8 \mid p = 0{,}5) = \frac{\binom{11}{9} + \binom{11}{10} + \binom{11}{11}}{2^{11}} = \frac{67}{2048} \approx 0{,}0327$$

a) Ensidig mothypotes: jämför med $\alpha = 0{,}05$. Eftersom $0{,}0327 < 0{,}05$ $\boxed{\text{förkastas } H_0}$.

b) Tvåsidig mothypotes: signifikansnivån fördelas på båda svansarna, så enkelsvanssannolikheten jämförs med $\alpha/2 = 0{,}025$. Eftersom $0{,}0327 > 0{,}025$ $\boxed{\text{kan } H_0 \text{ inte förkastas}}$.

Samma data ger alltså olika slutsats beroende på mothypotesen. Därför måste hypoteserna bestämmas **innan** vi tittar på data.

På räknaren. `binomCdf(11,0.5,9,11)` ≈ `0.0327`.

Uppgiften (Vännman 9.22) är återgiven ur det handskrivna lösningsförslaget ($\alpha_0 \approx 0{,}0327$), och slutsatserna stämmer med det.

Lektion 16, övning 4 (Vännman 9.23) — Defekta brickor

Sannolikheten att en bricka är defekt ska vara $0{,}01$. I en sändning om 300 brickor hittades 8 defekta. Pröva med direktmetoden $H_0:\ p = 0{,}01$ mot $H_1:\ p > 0{,}01$ på 2 % signifikansnivå.

Lösning

Modell. η = antal defekta brickor bland 300, $\eta \in \text{Bin}(300,\ p)$. Under H_0 är $p = 0{,}01 \leq 0{,}1$ och $n \geq 10$, så approximativt $\eta \in \text{Po}(300 \cdot 0{,}01) = \text{Po}(3)$.

α_0**.** Stora värden talar för H_1:

$$\alpha_0 = P(\eta \geq 8 \mid p = 0{,}01) = 1 - P(\eta \leq 7) \approx 1 - 0{,}9881 = \boxed{0{,}0119}$$

(Exakt binomialsannolikhet: $0{,}0115$.)

Slutsats. $\alpha_0 \approx 0{,}012 < 0{,}02$, så H_0 förkastas på $2\ \%$ signifikansnivå: andelen defekta är större än $0{,}01$.

På räknaren. `1-poissCdf(3,0,7)` ≈ `0.0119`. Den exakta binomialsannolikheten: `binomCdf(300,0.01,8,300)` ≈ `0.0115`.

Uppgiften (Vännman 9.23) är återgiven ur det handskrivna lösningsförslaget ($\alpha_0 \approx 0{,}0119 < 0{,}02$), och svaret stämmer med det.

Se även

- Hypotesprövning
- Sammansatta hypoteser
- Styrkefunktion
- Binomialfördelning
- Poissonfördelning

Resurser

- Khan Academy: P-values and significance tests
- Wikipedia: p-värde

Konfidensintervall och hypotesprövning

19 okt. 2026, 3 min läsning

UTKAST

#matematik #statistik #samband

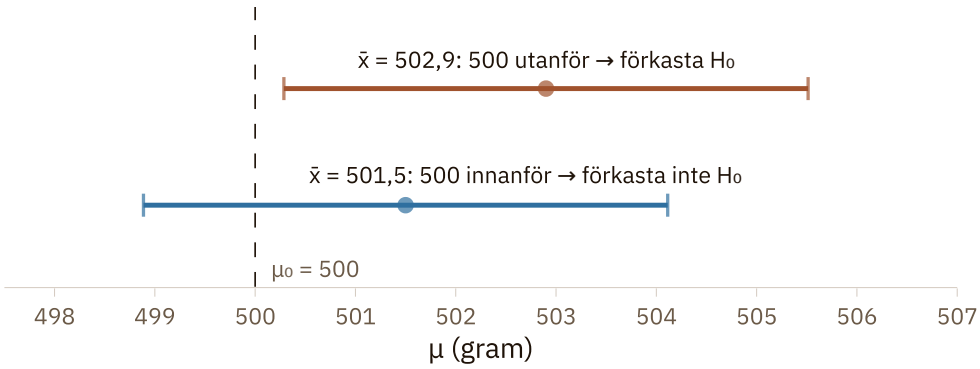
> **Kurs:** S0008M **Förkunskaper:** Konfidensintervall, Sammansatta hypoteser

Ett tvåsidigt test av $H_0 : \mu = \mu_0$ på signifikansnivån α förkastar nollhypotesen precis när μ_0 inte ligger i konfidensintervallet för μ med konfidensgraden $1 - \alpha$. Intervallet ger därmed testbeslutet för alla μ_0 på en gång.

1. Sambandet

Låt I_μ vara ett konfidensintervall för μ med konfidensgrad $1 - \alpha$. Då gäller:

$$H_0 : \mu = \mu_0 \text{ förkastas mot } H_1 : \mu \neq \mu_0 \text{ på nivån } \alpha \iff \mu_0 \notin I_\mu$$



Figuren visar två 95 %-intervall för påsfyllningen ($\sigma = 4$, $n = 9$), $\bar{x} \pm 1{,}96 \cdot 4/3$. Med $\bar{x} = 502{,}9$ blir intervallet $[500{,}29,\ 505{,}51]$; 500 ligger utanför, så det tvåsidiga testet förkastar $H_0 : \mu = 500$. Med $\bar{x} = 501{,}5$ blir det $[498{,}89,\ 504{,}11]$, som innehåller 500.

2. Varför

Ta fallet med okänd σ. Det tvåsidiga t-testet i Sammansatta hypoteser förkastar **inte** H_0 när

$$\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \le t_{\alpha/2}(n-1) \quad \Longleftrightarrow \quad \bar{x}-t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} \le \mu_0 \le \bar{x}+t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

Högerledet säger exakt att μ_0 ligger i konfidensintervallet. De två metoderna bygger på samma olikhet, bara löst för olika storheter.

Vad intervallet ger utöver testet

Testet svarar bara ja eller nej för ett enda μ_0. Intervallet visar dessutom **hur stor** avvikelsen kan vara. Därför är det ofta mer informativt att redovisa intervallet.

Ensidiga test

Ett ensidigt test på nivån α svarar på samma sätt mot ett **ensidigt** konfidensintervall med konfidensgrad $1-\alpha$, till exempel $\mu \le \bar{x}+t_{\alpha}(n-1)\,s/\sqrt{n}$ för $H_1:\ \mu<\mu_0$.

Vanliga fel

- **Tvåsidigt intervall mot ensidigt test.** Sambandet gäller tvåsidigt test med tvåsidigt intervall, eller ensidigt med ensidigt, på samma α.
- **Fel konfidensgrad.** Testet på nivån α hör ihop med konfidensgraden $1-\alpha$, inte $1-\alpha/2$.

---

3. Lösta lektionsuppgifter

Lektion 16, övning 2 — Vikten 500 gram via konfidensintervall

Fortsättning på testet av enheterna som ska väga 500 gram i Sammansatta hypoteser: 8 normalfördelade mätvärden gav $\bar{x}=497{,}8$ gram och $s=2{,}1$. Bestäm vad vi kan säga om hur stort μ är med $95\ \%$ säkerhet genom att beräkna och tolka ett $95\ \%$ konfidensintervall. Hur kan vi dra slutsatsen om $H_0:\ \mu=500$ ur intervallet i stället för att genomföra testet?

Lösning

Intervallet. σ är okänd, så med $t_{0{,}025}(7) \approx 2{,}365$:

$$I_\mu = \bar{x} \pm t_{0{,}025}(7)\frac{s}{\sqrt{8}} = 497{,}8 \pm 2{,}365 \cdot \frac{2{,}1}{\sqrt{8}} = 497{,}8 \pm 1{,}76 = \boxed{[496{,}0,\ 499{,}6]}$$

Tolkning. Med $95\ \%$ säkerhet ligger väntevärdet μ mellan $496{,}0$ och $499{,}6$ gram.

Testet. 500 ligger inte i intervallet, så $H_0 : \mu = 500$ förkastas på $5\ \%$ signifikansnivå mot $H_1 : \mu \neq 500$ — samma slutsats som t-testet gav. Intervallet säger dessutom att enheterna i genomsnitt väger mellan $0{,}4$ och $4{,}0$ gram för lite.

På räknaren. `menu` → **Statistics** → **Confidence Intervals** → **t Interval** med **Stats**: `x̄: 497.8`, `sx: 2.1`, `n: 8`, `C Level: 0.95`.

Facit: intervallet $[496{,}0,\ 499{,}6]$*; eftersom* 500 *inte ingår kan* H_0 *förkastas på* $5\ \%$ *signifikansnivå.*

Se även

- Konfidensintervall
- Hypotesprövning
- Sammansatta hypoteser
- t-fördelningen

Resurser

- Wikipedia: Confidence interval — relationship with hypothesis testing

Flerdimensionella fördelningar

19 okt. 2026, 9 min läsning

UTKAST

#matematik #sannolikhetslära #grundbegrepp

> **Kurs:** S0008M **Förkunskaper:** Fördelningsfunktion, Funktioner av flera variabler

En flerdimensionell fördelning beskriver två eller flera stokastiska variabler samtidigt genom en simultan fördelning. Ur den får vi marginalfördelningarna för varje variabel för sig, men marginalerna räcker inte för att återskapa simultanfördelningen om variablerna är beroende.

1. Tvådimensionella stokastiska variabler

En **tvådimensionell stokastisk variabel** är ett par (ξ, η) av stokastiska variabler som betraktas tillsammans — en punkt i planet vars koordinater är slumpmässiga. Ibland skriver vi (ξ_1, ξ_2) i stället för (ξ, η).

Vi behöver den när variablerna **inte** kan antas vara oberoende: utsänd och mottagen signal, längd och vikt hos samma person, två studenters betyg på samma tenta. Är variablerna oberoende räcker det att känna fördelningarna var för sig; är de beroende måste vi beskriva dem tillsammans.

Som i det endimensionella fallet skiljer vi mellan diskreta och kontinuerliga fördelningar.

2. Diskret fall: simultan sannolikhetsfunktion

(ξ, η) är diskret om både ξ och η är diskreta. Fördelningen anges med den **simultana sannolikhetsfunktionen**

$$\boxed{p(x, y) = P(\xi = x,\ \eta = y)}$$

för de värden (x, y) som paret kan anta, och $p(x, y) = 0$ annars. Den skrivs ofta som en tabell med x i raderna och y i kolumnerna.

Sannolikheten att paret hamnar i ett område C i planet är summan av massan där:

$$P((\xi,\eta)\in C)=\sum_{(x,y)\in C}p(x,y)$$

---

3. Kontinuerligt fall: simultan frekvensfunktion

Enligt kompendiets definition 1.3 har (ξ,η) en **kontinuerlig tvådimensionell fördelning** om det finns en funktion $f(x,y)$ sådan att för varje område D i planet

$$\boxed{P((\xi,\eta)\in D)=\iint_D f(x,y)\,dx\,dy}$$

$f(x,y)$ kallas den **simultana frekvensfunktionen** och uppfyller

$$f(x,y)\geq 0,\qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$$

Sannolikhet är alltså volym under ytan $z=f(x,y)$, precis som den i en dimension är area under frekvensfunktionen. Beräkningarna är dubbelintegraler.

Simultan fördelningsfunktion

Oavsett typ bestäms fördelningen av $F(x,y)=P(\xi\leq x,\ \eta\leq y)$. I det kontinuerliga fallet är $f(x,y)=\dfrac{\partial^2 F}{\partial x\,\partial y}$. Vi använder den sällan direkt, men idén att först ta fram en fördelningsfunktion och sedan derivera återkommer i Icke-linjära funktioner av stokastiska variabler.

---

4. Väntevärden av funktioner

Är $g(x,y)$ en funktion av två variabler gäller

$$E(g(\xi,\eta))=\sum_x\sum_y g(x,y)\,p(x,y)\qquad\text{respektive}\qquad E(g(\xi,\eta))=\iint_{\mathrm{R}^2} g(x,y)\,f(x,y)\,dx\,dy$$

Med $g(x,y)=x+y$ följer till exempel $E(\xi+\eta)=E(\xi)+E(\eta)$ direkt, och med $g(x,y)=xy$ får vi $E(\xi\eta)$ som behövs för kovariansen.

---

5. Likformig tvådimensionell fördelning

(ξ, η) är **likformigt fördelad** på ett område K om frekvensfunktionen är konstant på K och noll utanför (kompendiet, definition 1.4):

$$f(x,y) = \begin{cases} \dfrac{1}{|K|}, & (x,y) \in K \\ 0, & \text{annars} \end{cases}$$

där $|K|$ är arean av K. För ett delområde $A \subset K$ blir sannolikheten ett rent areaförhållande:

$$\boxed{P((\xi,\eta) \in A) = \frac{|A|}{|K|}}$$

Rita området

Med likformig fördelning blir nästan varje uppgift geometri: rita K, rita området som händelsen motsvarar, och räkna areor. Ingen integral behövs.

---

6. Oberoende

ξ och η är oberoende om och endast om den simultana fördelningen är produkten av marginalfördelningarna:

$$p(x,y) = p_\xi(x)\,p_\eta(y) \qquad \text{respektive} \qquad f(x,y) = f_\xi(x)\,f_\eta(y) \qquad \text{för alla } x, y$$

Räcker det med ett enda par (x, y) där likheten inte gäller är variablerna beroende. Omvänt: är ξ och η oberoende och kända var för sig får vi den simultana fördelningen som produkten.

Vanliga fel

- $f(x,y)$ **som sannolikhet.** Frekvensfunktionen kan vara större än 1; sannolikhet är *volym* under ytan, inte funktionsvärdet.
- **Integrerat över hela kvadraten.** f är 0 utanför området, så integrationsgränserna följer området där $f > 0$.
- **Oberoende på ett icke-rektangulärt område.** Är området där $f > 0$ inte en rektangel, t.ex. triangeln $0 < y < x < 1$, är variablerna beroende — även om f är konstant och formeln ser ut att faktorisera.
- **Glömt $|K|$.** Likformig fördelning har höjden $1/|K|$, inte 1.

---

7. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar och rekommenderas inför tentan. Kompendiet har inget facit, så varje sådan övning är löst två gånger oberoende av varandra.

Lektion 17, övning 3 — Två oberoende rektangelfördelade variabler

Antag att ξ och η är oberoende och rektangelfördelade på intervallet $(0,1)$.

a) Bestäm den simultana frekvensfunktionen för (ξ,η). Vad kallas denna tvådimensionella fördelning?

b) Händelsen $A=\{\xi+\eta\le 1{,}5\}$ kan skrivas $\{(\xi,\eta)\in C\}$ för ett lämpligt område C i planet. Vilket? Rita en bild.

c) Beräkna $P(\xi+\eta\le 1{,}5)$.

d) Beräkna $P(\xi+\eta\le 0{,}75)$.

e) Två personer ska träffas på en viss plats mellan kl 12.00 och 13.00 och väntar högst 20 minuter på varandra. Om de anländer oberoende av varandra vid slumpmässigt valda tider under timmen, hur stor är sannolikheten att de träffas?

Lösning

a) Oberoende ger produkten av marginalerna, $f_\xi(x)=1$ på $(0,1)$ och $f_\eta(y)=1$ på $(0,1)$:

$$f(x,y)=f_\xi(x)\,f_\eta(y)=\begin{cases}1, & 0\le x\le 1,\ 0\le y\le 1\\ 0, & \text{annars}\end{cases}$$

Det är den **likformiga fördelningen på enhetskvadraten** K, med $|K|=1$.

b) $C=\{(x,y):x+y\le 1{,}5\}$, halvplanet under linjen $y=1{,}5-x$. Inom kvadraten är det allt utom ett hörn uppe till höger.



c) Komplementet inom kvadraten är triangeln med hörn i $(0{,}5;\ 1)$, $(1;\ 1)$ och $(1;\ 0{,}5)$, med kateterna $0{,}5$:

$$P(\xi + \eta \le 1{,}5) = 1 - \frac{0{,}5 \cdot 0{,}5}{2} = 1 - \frac{1}{8} = \boxed{\frac{7}{8}}$$

d) Nu är området själv en triangel med kateterna $0{,}75$ (mittenpanelen):

$$P(\xi + \eta \le 0{,}75) = \frac{0{,}75^2}{2} = \boxed{\frac{9}{32}}$$

e) Mät tiden i timmar efter kl 12: ankomsttiderna ξ och η är oberoende och $R(0,1)$, alltså likformigt fördelade på kvadraten. De träffas om $|\xi - \eta| \le 1/3$, bandet längs diagonalen i högra panelen. Utanför bandet ligger två trianglar med kateterna $2/3$:

$$P(|\xi - \eta| \le \frac{1}{3}) = 1 - 2 \cdot \frac{(2/3)^2}{2} = 1 - \frac{4}{9} = \boxed{\frac{5}{9}}$$

Facit: a) $f(x,y) = 1$ *om* $0 \le x, y \le 1$, 0 *annars; b)* $C = \{(x,y) : x + y \le 1{,}5\}$*; c)* $7/8$*; d)* $9/32$ *; e)* $5/9$*.*

Kompendiet, övning 1.1 — Seminarier och tentamensbetyg

För en universitetskurs fördes statistik över tentamensresultat och närvaro vid två frivilliga seminarier. Låt ξ vara betyget på tentamen (2 betyder underkänt) och η antalet seminarier som en slumpmässigt vald student har gått på. Den simultana sannolikhetsfunktionen $p(x,y)$ är

$x\backslash y$	0	1	2
2	0,18	0,07	0,05
3	0,11	0,16	0,08
4	0,06	0,08	0,11
5	0,02	0,03	0,05

Beräkna sannolikheten att en slumpmässigt vald student går på minst ett seminarium och får betyg 3 eller bättre.

Lösning

Området. Händelsen är $\{\xi \geq 3,\ \eta \geq 1\}$: raderna $x = 3, 4, 5$ och kolumnerna $y = 1, 2$, alltså sex celler i tabellens nedre högra hörn.

Summera massan där.

$$P(\xi \geq 3,\ \eta \geq 1) = (0{,}16 + 0{,}08) + (0{,}08 + 0{,}11) + (0{,}03 + 0{,}05) = \boxed{0{,}51}$$

Rimlighetskoll. $P(\xi \geq 3) = 0{,}70$ och $P(\eta \geq 1) = 0{,}63$. Vore variablerna oberoende skulle svaret bli $0{,}70 \cdot 0{,}63 \approx 0{,}44$. Vi får mer, eftersom seminarienärvaro och bra betyg hänger ihop.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

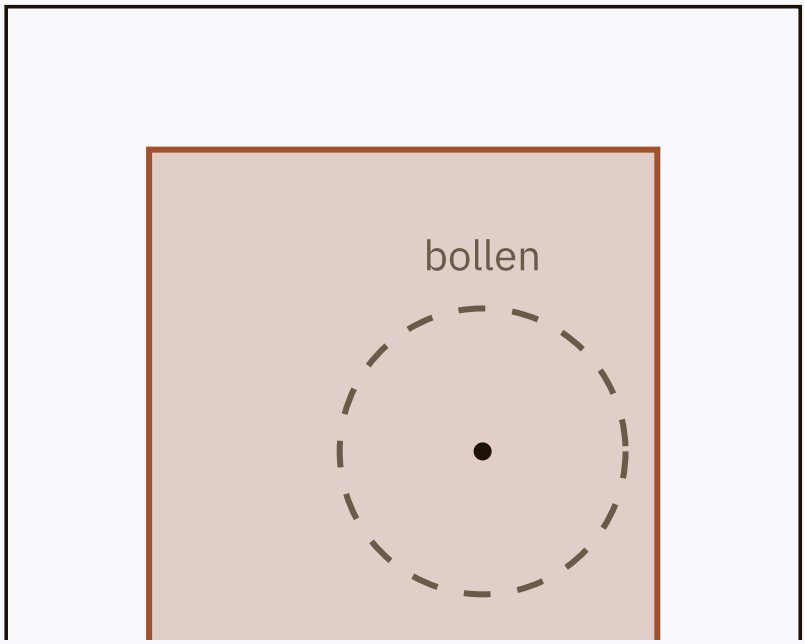
Kompendiet, övning 1.3 — Tennisboll genom ett nät

En tennisboll med diametern a cm kastas mot ett volleybollnät. Nätmaskorna är kvadratiska med sidan $b > a$. Vad är sannolikheten att bollen går igenom nätet utan att röra någon tråd? (Vi bortser från trådarnas tjocklek.)

Lösning

Modell. Bollens mittpunkt (ξ, η) hamnar i någon maska. Inom den maskan antar vi att mittpunkten är likformigt fördelad på kvadraten K med sidan b, så $|K| = b^2$.

Händelsen. Bollen rör ingen tråd om mittpunkten ligger minst en radie, $a/2$, från alla fyra sidor. Det är en inre kvadrat med sidan $b - a$:



Areaförhållandet.

$$P(\text{igenom}) = \frac{(b-a)^2}{b^2} = \boxed{\left(1 - \frac{a}{b}\right)^2}$$

Rimlighetskoll. $a \to 0$ ger sannolikheten 1 och $a \to b$ ger 0, som sig bör.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.6 — Vilken butik får första kunden?

En butik i Stockholm öppnar kl 9.00, och tiden tills den första kunden kommer in är exponentialfördelad med väntevärdet 4 minuter. En butik i Umeå öppnar också kl 9.00, med exponentialfördelad tid till första kunden med väntevärdet 7 minuter. Vad är sannolikheten att butiken i Stockholm får sin första kund före butiken i Umeå?

Lösning

Modell. ξ = tiden till första kunden i Stockholm, $\xi \in \mathrm{Exp}(1/4)$; η = tiden i Umeå, $\eta \in \mathrm{Exp}(1/7)$. Butikerna påverkar inte varandra, så ξ och η är oberoende och $f(x,y) = \frac{1}{4}e^{-x/4} \cdot \frac{1}{7}e^{-y/7}$ för $x, y > 0$.

Händelsen är området $\{(x,y) : x < y\}$ ovanför diagonalen. Integrera först över y från x till ∞, vilket ger $P(\eta > x) = e^{-x/7}$:

$$P(\xi < \eta) = \int_0^\infty \frac{1}{4}e^{-x/4}\, e^{-x/7}\, dx = \frac{1}{4} \cdot \frac{1}{\frac{1}{4}+\frac{1}{7}} = \boxed{\frac{7}{11}} \approx 0{,}636$$

Tolkning. För två oberoende exponentialfördelningar gäller allmänt $P(\xi < \eta) = \lambda_\xi/(\lambda_\xi + \lambda_\eta)$. Stockholm har den högre intensiteten ($1/4$ kund per minut mot $1/7$) och vinner därför oftare.

På räknaren. `∫((1/4)*e^(-x/4)*e^(-x/7),x,0,∞)` ger `7/11` i en Calculator-sida.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.8 — Var ska sjukhuset ligga?

Ett sjukhus ska placeras i ett kvadratiskt samhälle $K = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le 1\}$, där $(0,0)$ är sydvästra hörnet. Man kan bara färdas parallellt med koordinataxlarna, så avståndet från $(0,0)$ till (x,y) är $x + y$. Antalet olyckor ökar linjärt i både x- och y-led, och platsen för en olycka har frekvensfunktionen

$$f(x,y) = \begin{cases} c\,xy, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{annars} \end{cases}$$

a) Bestäm konstanten c.

b) Sjukhuset placeras i $(0, 0)$. Vad är sannolikheten att ambulansen behöver färdas högst $0{,}75$?

c) Vad är den förväntade utryckningssträckan med sjukhuset i $(0, 0)$?

d) ∗ Vilken placering minimerar den förväntade utryckningssträckan?

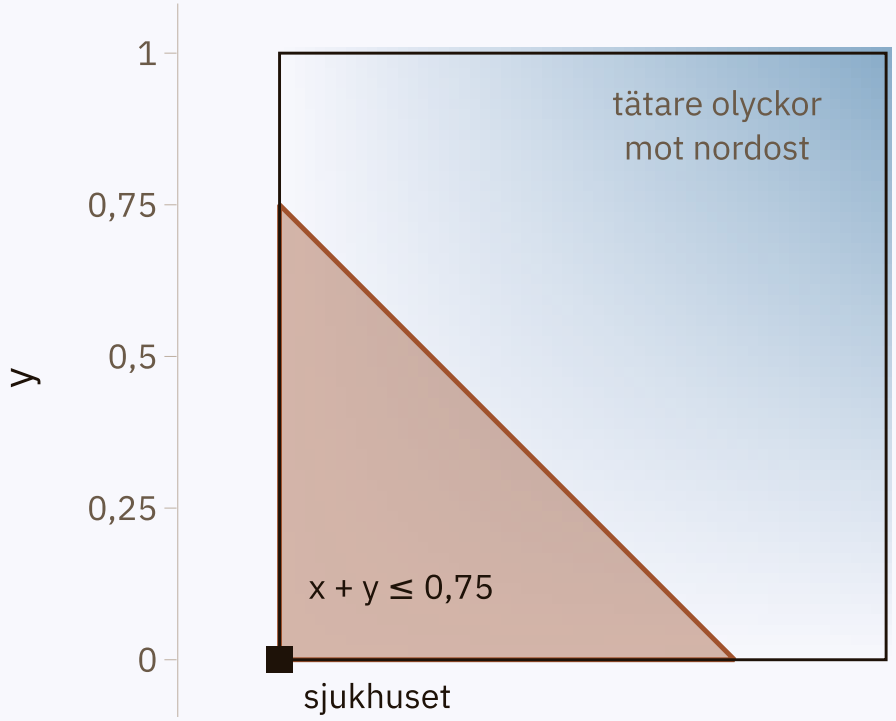
Lösning

a) Totala sannolikheten ska vara 1:

$$\int_0^1 \int_0^1 c\,xy\,dy\,dx = c \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{c}{4} = 1 \quad \Longrightarrow \quad \boxed{c = 4}$$

Frekvensfunktionen faktoriserar, $4xy = 2x \cdot 2y$, så olycksplatsens koordinater ξ och η är **oberoende**, var och en med frekvensfunktionen $2x$ på $[0, 1]$.

b) Händelsen är triangeln $\{x + y \leq 0{,}75\}$ i sydvästra hörnet, där olyckorna är som glesast:



$$P(\xi+\eta\le 0{,}75)=\int_0^{0{,}75}\int_0^{0{,}75-x} 4xy\,dy\,dx=\int_0^{0{,}75} 2x(0{,}75-x)^2\,dx=\frac{0{,}75^4}{6}=\boxed{\frac{27}{512}}\approx 0$$

Triangeln täcker $28\ \%$ av kvadratens area men bara $5\ \%$ av sannolikheten.

c) Väntevärdets linjäritet och $E(\xi)=\int_0^1 x\cdot 2x\,dx=\frac{2}{3}$:

$$E(\xi+\eta)=E(\xi)+E(\eta)=\boxed{\frac{4}{3}}$$

d) Med sjukhuset i (u,v) är sträckan $|\xi-u|+|\eta-v|$, och de två termerna kan minimeras var för sig. $E|\xi-u|$ minimeras när u är medianen för ξ — då ligger lika mycket sannolikhet på varje sida, och en liten flytt åt något håll förlänger lika mycket som den förkortar. $F_\xi(u)=u^2=1/2$ ger

$$\boxed{(u, v) = \left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right)} \approx (0{,}71;\ 0{,}71)$$

På räknaren. b) `∫(∫(4*x*y,y,0,0.75-x),x,0,0.75)` ger `27/512`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Marginalfördelning
- Betingad fördelning och betingat väntevärde
- Kovarians
- Korrelation
- Tvådimensionell normalfördelning
- Dubbelintegraler

Resurser

- Khan Academy: Joint distributions
- Wikipedia: Joint probability distribution

Marginalfördelning

19 okt. 2026, 4 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Flerdimensionella fördelningar

---

Marginalfördelningen är fördelningen för en av variablerna i en flerdimensionell fördelning, betraktad för sig. Den fås genom att summera eller integrera bort den andra variabeln.

---

1. Diskret fall

Om $p(x, y)$ är den simultana sannolikhetsfunktionen för (ξ, η) ges marginalfördelningarna av

$$\boxed{p_\xi(x) = \sum_y p(x, y), \qquad p_\eta(y) = \sum_x p(x, y)}$$

där summorna tas över alla värden som η respektive ξ kan anta.

Att $\xi = x$ inträffar är detsamma som att (ξ, η) hamnar på raden x i tabellen, oavsett vilket y. Raderna är disjunkta händelser, så sannolikheterna adderas — det är lagen om total sannolikhet i tabellform.

Summera raderna och kolumnerna

Skriv den simultana tabellen och lägg till en summakolumn och en summarad. Radsummorna är $p_\xi(x)$, kolumnsummorna $p_\eta(y)$ — därav namnet: de står i *marginalen*. Båda ska summera till 1.

	y = 3	y = 4	y = 5	$p_\xi(x)$
x = 3	0,31	0,11	0,05	**0,47**
x = 4	0,02	0,22	0,12	**0,36**
x = 5	0,01	0,03	0,13	**0,17**
$p_\eta(y)$	**0,34**	**0,36**	**0,30**	**1**

Den simultana tabellen från lektion 17, övning 1a (nedan) med radsummorna $p_\xi(x)$ och kolumnsummorna $p_\eta(y)$ i marginalen. Mörkare ruta betyder större sannolikhet.

---

2. Kontinuerligt fall

Med den simultana frekvensfunktionen $f(x, y)$ ersätts summorna med integraler:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad f_\eta(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

För fixt x integrerar vi längs den lodräta linjen genom x. Integrationsgränserna bestäms av var $f(x, y)$ är skild från noll för det x-värdet, och det är där de flesta fel görs.

⚠ Gränserna beror på x

Är (ξ, η) likformigt fördelad på en cirkelskiva med radie 1 går y från $-\sqrt{1 - x^2}$ till $\sqrt{1 - x^2}$ för ett givet x. Marginalen blir då $f_\xi(x) = \frac{2}{\pi}\sqrt{1 - x^2}$, en halvcirkel och inte en konstant — trots att den simultana fördelningen är likformig.

---

3. Marginalerna räcker inte

Marginalfördelningarna beskriver varje variabel för sig men säger ingenting om hur de beror av varandra. Olika simultana fördelningar kan ha exakt samma marginaler. Bara om ξ och η är oberoende går det att återskapa den simultana fördelningen, som produkten $p(x, y) = p_\xi(x)\, p_\eta(y)$.

Hur mycket den ena variabeln säger om den andra fångas i stället av den betingade fördelningen och av korrelationen.

⚠ **Vanliga fel**

- **Integrerat över fel variabel.** $f_\xi(x)$ fås genom att integrera bort y; svaret får bara innehålla x.
- **Stödmängden saknas.** Ange för vilka x marginalen gäller, och att den är 0 utanför.
- **Produkten av marginalerna som simultan fördelning.** $p_\xi(x)\, p_\eta(y)$ är den simultana fördelningen bara vid oberoende.
- **Ingen kontroll.** Marginalen ska summera eller integrera till 1 — ett snabbt sätt att hitta fel gränser.

---

4. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 17, övning 1a — Joels och Majas betyg

Joel och Maja har pluggat tillsammans inför en tenta för överbetyg. Låt ξ och η vara betyget för Joel respektive Maja. Den simultana sannolikhetsfunktionen för (ξ, η) är

$x \backslash y$	3	4	5
3	0,31	0,11	0,05
4	0,02	0,22	0,12
5	0,01	0,03	0,13

Bestäm marginalfördelningarna för ξ och η. Bestäm även väntevärde och standardavvikelse för Majas betyg.

Lösning

Marginalerna. Radsummor ger Joels betyg, kolumnsummor Majas:

$x\backslash y$	**3**	**4**	**5**	$p_\xi(x)$
3	0,31	0,11	0,05	**0,47**
4	0,02	0,22	0,12	**0,36**
5	0,01	0,03	0,13	**0,17**
$p_\eta(y)$	**0,34**	**0,36**	**0,30**	1

Majas väntevärde.

$$E(\eta) = 3 \cdot 0{,}34 + 4 \cdot 0{,}36 + 5 \cdot 0{,}30 = \boxed{3{,}96}$$

Majas standardavvikelse. $E(\eta^2) = 9 \cdot 0{,}34 + 16 \cdot 0{,}36 + 25 \cdot 0{,}30 = 16{,}32$, så

$$V(\eta) = 16{,}32 - 3{,}96^2 = 0{,}6384, \qquad \sigma_\eta = \sqrt{0{,}6384} \approx \boxed{0{,}799}$$

På räknaren. Lägg betygen $3, 4, 5$ i en lista och sannolikheterna `{0.34,0.36,0.30}` i en annan. Kör `menu` → **Statistics** → **Stat Calculations** → **One-Variable Statistics** med sannolikhetslistan som **Frequency List**. Då är `x̄` $= E(\eta)$ och `σx` $= \sigma_\eta$ — ta `σx`, inte `sx`, eftersom vi har hela fördelningen och inte ett stickprov.

Facit: $P(\xi = x) = 0{,}47;\ 0{,}36;\ 0{,}17$ *och* $P(\eta = y) = 0{,}34;\ 0{,}36;\ 0{,}30$ *för* $x, y = 3, 4, 5$*;* $E(\eta) = 3{,}96$*,* $\sigma = 0{,}799$.

Kompendiet, övning 1.2 — Likformig fördelning på kvadraten

Låt (ξ, η) vara likformigt fördelad på $K = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le 1\}$. Beräkna marginalfördelningarna för ξ och η direkt från definitionen.

Lösning

Simultan frekvensfunktion. $|K| = 1$, så $f(x, y) = 1$ på K och 0 utanför.

Marginalen för ξ. För ett fixt x mellan 0 och 1 är $f(x, y) = 1$ precis när $0 \le y \le 1$:

$$f_\xi(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{0}^{1} 1\, dy = 1, \qquad 0 \le x \le 1$$

och $f_\xi(x) = 0$ för övriga x, eftersom hela den lodräta linjen då ligger utanför K. Alltså $\boxed{\xi \in R(0, 1)}$.

Marginalen för η. Samma räkning med rollerna ombytta ger $\boxed{\eta \in R(0, 1)}$.

Kommentar. Här är $f(x, y) = f_\xi(x)\, f_\eta(y)$ överallt, så ξ och η är oberoende. Det är omvändningen till lektion 17, övning 3 a). Jämför med cirkelskivan i varningen ovan: där är marginalerna inte likformiga, och variablerna är beroende.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Flerdimensionella fördelningar
- Betingad fördelning och betingat väntevärde
- Lagen om total sannolikhet
- Sannolikhetsfunktion
- Frekvensfunktion

Resurser

- Wikipedia: Marginal distribution

Betingad fördelning och betingat väntevärde

19 okt. 2026, 11 min läsning

UTKAST

#matematik #sannolikhetslära #definition

> **Kurs:** S0008M **Förkunskaper:** Marginalfördelning, Betingad sannolikhet

---

Den betingade fördelningen för ξ givet $\eta = y$ är fördelningen för ξ när vi vet värdet på η. Den fås genom att normera ett tvärsnitt av den simultana fördelningen, och dess väntevärde — det betingade väntevärdet — är den bästa prognosen för ξ givet informationen.

---

1. Ett tvärsnitt som normeras

Är två variabler beroende ger värdet på den ena ny information om vilka värden på den andra som är troliga — precis som för betingade sannolikheter mellan händelser. I en simultan tabell betyder "givet $\eta = y$" att vi bara tittar på kolumnen y. Kolumnens värden summerar till $p_\eta(y)$, inte till 1, så vi delar med kolumnsumman.

Definition 1.9 (kompendiet). Den **betingade fördelningen** för ξ givet $\eta = y$ ges i det diskreta fallet av sannolikhetsfunktionen

$$\boxed{p_{\xi\,|\,\eta=y}(x) = \frac{p(x,y)}{p_\eta(y)}}$$

och i det kontinuerliga fallet av frekvensfunktionen

$$\boxed{f_{\xi\,|\,\eta=y}(x) = \frac{f(x,y)}{f_\eta(y)}}, \qquad f_\eta(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Nämnaren är marginalfördelningen för den variabel vi betingar på.

Oberoende variabler

Är ξ och η oberoende är $p(x,y) = p_\xi(x)\,p_\eta(y)$, och den betingade fördelningen blir samma som marginalfördelningen för ξ. Informationen om η ändrar ingenting.

2. Betingat väntevärde

Definition 1.12 (kompendiet). Väntevärdet i den betingade fördelningen för ξ givet $\eta = y$ kallas det **betingade väntevärdet** och betecknas $E(\xi \mid \eta = y)$:

$$E(\xi \mid \eta = y) = \sum_x x\, p_{\xi\mid\eta=y}(x) \qquad \text{respektive} \qquad E(\xi \mid \eta = y) = \int_{-\infty}^{\infty} x\, f_{\xi\mid\eta=y}(x)\, dx$$

På samma sätt får vi den betingade variansen och standardavvikelsen, som mäter hur osäkra vi fortfarande är om ξ när η är känd.

Bästa prognosen

Väntevärdet μ är den prognos c för ξ som minimerar det förväntade kvadratiska felet $E((\xi - c)^2)$ (beviset är kompendiets övning 1.12 nedan). På samma sätt är $E(\xi \mid \eta = y)$ den bästa prognosen för ξ **givet** att $\eta = y$, i minsta-kvadrat-mening.

3. Lagen om totalt väntevärde

Byter vi y mot η i $E(\xi \mid \eta = y)$ får vi en stokastisk variabel, $E(\xi \mid \eta) = g(\eta)$ där $g(y) = E(\xi \mid \eta = y)$.

Lagen om totalt väntevärde (kompendiet, sats 1.13).

$$\boxed{E(\xi) = E(E(\xi \mid \eta))}$$

Det är lagen om total sannolikhet för väntevärden: räkna väntevärdet fall för fall och väg ihop fallen. Satsen gäller oavsett om variablerna är diskreta, kontinuerliga eller en av varje.

När satsen används

När ξ är svår att beskriva direkt men blir enkel om vi vet värdet på η. Kommer η kunder till en affär och var och en köper mjölk med sannolikheten $0{,}2$ är $E(\xi \mid \eta) = 0{,}2\,\eta$, och satsen ger $E(\xi) = 0{,}2\,E(\eta)$ utan att vi någonsin tar fram fördelningen för ξ.

Tillämpad på en indikatorvariabel för en händelse A blir satsen $P(A) = \int P(A \mid \eta = y)\, f_\eta(y)\, dy$, användbar när $P(A \mid \eta = y)$ är lätt att räkna.

4. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 17, övning 1b–f — Majas betyg givet Joels

Samma tabell som i Marginalfördelning: ξ = Joels betyg, η = Majas betyg, med

$x \backslash y$	**3**	**4**	**5**	$p_\xi(x)$
3	0,31	0,11	0,05	0,47
4	0,02	0,22	0,12	0,36
5	0,01	0,03	0,13	0,17
$p_\eta(y)$	0,34	0,36	0,30	

Utan villkor är $E(\eta) = 3{,}96$ och $\sigma_\eta = 0{,}799$.

b) Joel skriver tentan först och Maja får veta att han fick betyg 3. Beräkna den betingade fördelningen för Majas betyg.

c) Skissa Majas fördelning före och efter beskedet. Vad har hänt med väntevärdet och spridningen?

d) Beräkna väntevärde och standardavvikelse i den betingade fördelningen som kontroll.

e) Vad hade hänt med väntevärdet och standardavvikelsen om Maja i stället fått veta att Joel fick en femma?

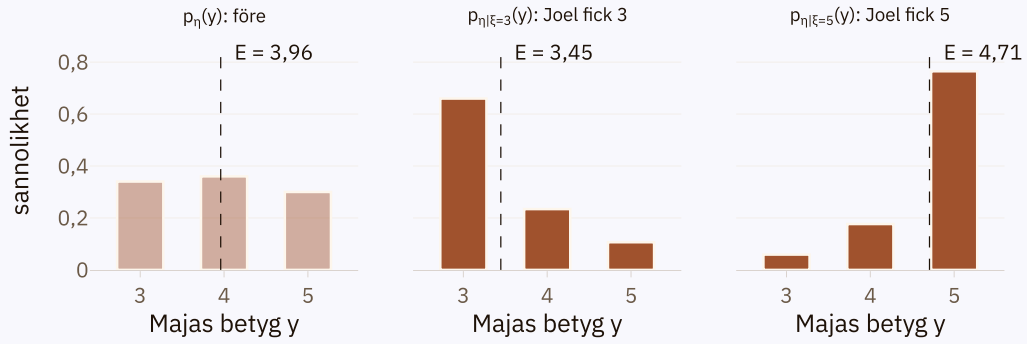
f) Kontrollera förmodan i e) genom att räkna.

Lösning

b) Ta raden $x = 3$ och dela med radsumman $p_\xi(3) = 0{,}47$:

$$p_{\eta\,|\,\xi=3}(y) = \frac{0{,}31}{0{,}47},\ \frac{0{,}11}{0{,}47},\ \frac{0{,}05}{0{,}47} \approx \boxed{0{,}6596;\ 0{,}2340;\ 0{,}1064} \qquad y = 3, 4, 5$$

c) Massan har flyttats mot trean och samlats där:



Väntevärdet har **minskat**, och eftersom massan nu är koncentrerad på ett värde har spridningen också **minskat**.

d)

$$E(\eta \mid \xi = 3) = 3 \cdot 0{,}6596 + 4 \cdot 0{,}2340 + 5 \cdot 0{,}1064 \approx \boxed{3{,}4468}$$

$E(\eta^2 \mid \xi = 3) \approx 12{,}340$, så $V(\eta \mid \xi = 3) \approx 12{,}340 - 3{,}4468^2 \approx 0{,}460$ och standardavvikelsen blir $\approx \boxed{0{,}6782}$.

e) Betygen samvarierar positivt, så en femma hos Joel bör flytta Majas fördelning uppåt: väntevärdet **ökar**. Massan bör samlas vid femman, så spridningen **minskar**.

f) Raden $x = 5$ delad med $p_\xi(5) = 0{,}17$:

$$p_{\eta\,|\,\xi=5}(y) \approx 0{,}0588;\ 0{,}1765;\ 0{,}7647$$

$$E(\eta \mid \xi = 5) \approx \boxed{4{,}7059}, \qquad \sqrt{V(\eta \mid \xi = 5)} \approx \boxed{0{,}5703}$$

Förmodan stämmer.

På räknaren. One-Variable Statistics med betygen $3, 4, 5$ som **X1 List** och radens sannolikheter, till exempel `{0.31,0.11,0.05}/0.47`, som **Frequency List**. `x̄` ger det betingade väntevärdet och `σx` den betingade standardavvikelsen.

Facit: b) 0,6596; 0,2340; 0,1064; c) minskat resp. minskat; d) 3,4468 resp. 0,6782; e) ökat resp. minskat; f) 4,7059 resp. 0,5703.

Lektion 17, övning 2 — Total livslängd för batterier

I ett lager finns 25 batterier av typ a och 10 av typ b. Livslängderna är exponentialfördelade, och sannolikheten att ett batteri räcker mer än 100 drifttimmar är $0{,}95$ för typ a och $0{,}97$ för typ b. Vi väljer ut 10 batterier på måfå. Vad är väntevärdet av den totala livslängden för de 10 batterierna?

Tips: inför ξ = total livslängd för de 10 och η = antal batterier av typ a, och använd lagen om totalt väntevärde.

Lösning

Väntevärdena per typ. För en exponentialfördelad livslängd med parameter λ är $P(\text{livslängd} > t) = e^{-\lambda t}$ och väntevärdet $1/\lambda$. Ur $e^{-100\lambda} = p$ får vi $1/\lambda = 100/(-\ln p)$:

$$\mu_a = \frac{100}{-\ln 0{,}95} \approx 1949{,}6 \text{ h}, \qquad \mu_b = \frac{100}{-\ln 0{,}97} \approx 3283{,}1 \text{ h}$$

Betingat på η. Givet att $\eta = y$ av de 10 är av typ a är resten, $10 - y$, av typ b:

$$E(\xi \mid \eta = y) = y\,\mu_a + (10 - y)\,\mu_b$$

Lagen om totalt väntevärde. η är antalet typ a vid dragning utan återläggning, $\eta \in \text{Hyp}(35,\ 10,\ 25/35)$ med $E(\eta) = 10 \cdot \frac{25}{35} \approx 7{,}143$. Eftersom $E(\xi \mid \eta)$ är linjär i η:

$$E(\xi) = E(\eta)\,\mu_a + (10 - E(\eta))\,\mu_b \approx 7{,}143 \cdot 1949{,}6 + 2{,}857 \cdot 3283{,}1 \approx \boxed{23\,306 \text{ timmar}}$$

Facit: 23 305,75 timmar.

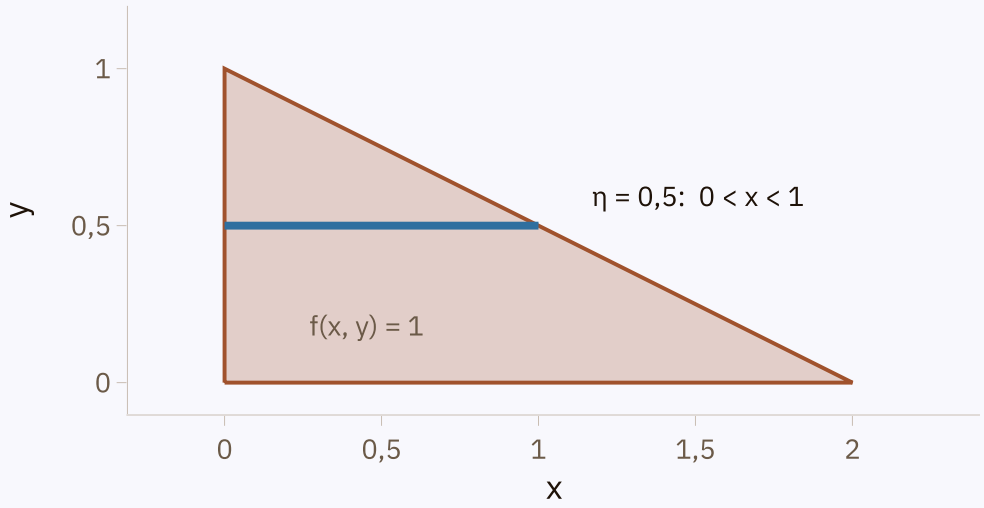
Lektion 17, övning 4 — Likformig fördelning på en triangel

Antag att (ξ, η) är likformigt fördelad på triangeln med hörn i $(0, 0)$, $(0, 1)$ och $(2, 0)$.

a) Diskutera utan beräkningar hur $f_\xi(x)$ borde se ut, och hur den betingade fördelningen för ξ givet $\eta = 0{,}5$ borde se ut.

b) Beräkna $f_\xi(x)$ och $f_{\xi\,|\,\eta=0{,}5}(x)$.

Lösning



a) Triangeln är hög vid $x = 0$ och smalnar av mot $x = 2$. Marginalen $f_\xi(x)$ borde därför också se ut som en triangel, störst vid 0 och noll vid 2. Givet $\eta = 0{,}5$ ligger punkten på den vågräta sträckan inuti triangeln, och eftersom den simultana fördelningen är likformig bör den betingade fördelningen vara likformig på den sträckan.

b) Triangelns area är $\frac{1}{2} \cdot 2 \cdot 1 = 1$, så $f(x,y) = 1$ på triangeln. Den sneda sidan är $y = 1 - x/2$.

Marginalen. För $0 < x < 2$ går y från 0 till $1 - x/2$:

$$f_\xi(x) = \int_0^{1-x/2} 1\,dy = \boxed{1 - \frac{x}{2}}, \qquad 0 < x < 2$$

en triangel, som förutsagt.

Den betingade. För $0 < y < 1$ går x från 0 till $2(1-y)$, så $f_\eta(y) = 2(1-y)$ och $f_\eta(0{,}5) = 1$. Alltså

$$f_{\xi\,|\,\eta=0{,}5}(x) = \frac{f(x;\ 0{,}5)}{f_\eta(0{,}5)} = \boxed{1}, \qquad 0 < x < 1$$

en rektangelfördelning på $(0,\,1)$, likformig på sträckan $y = 0{,}5$ inuti triangeln.

Facit: $f_\xi(x)$ ser ut som en triangel, och den betingade fördelningen är likformig på den del av linjen $y = 0{,}5$ som ligger i triangeln. Facit ger bara formen; uttrycken är lösta två gånger oberoende av varandra och stämmer överens.

Kompendiet, övning 1.10 — Tre diskreta variabler

Definitionen av en diskret tredimensionell fördelning är analog med det tvådimensionella fallet. Antag att (ξ, η, ζ) har sannolikhetsfunktionen $p(x, y, z)$.

a) Ange uttryck för marginalfördelningarna för ξ, η och ζ.

b) Ange ett uttryck för den betingade sannolikhetsfunktionen för ξ givet $\eta = y$, och för den betingade sannolikhetsfunktionen för ξ givet $\eta = y$ och $\zeta = z$.

Lösning

a) Summera bort de variabler vi inte är intresserade av:

$$p_\xi(x) = \sum_y \sum_z p(x,y,z), \qquad p_\eta(y) = \sum_x \sum_z p(x,y,z), \qquad p_\zeta(z) = \sum_x \sum_y p(x,y,z)$$

b) Som i det tvådimensionella fallet: simultan sannolikhet delad med sannolikheten för villkoret. Med ett villkor behöver täljaren den simultana fördelningen för (ξ, η), där ζ är bortsummerad:

$$p_{\xi\,|\,\eta=y}(x) = \frac{\sum_z p(x,y,z)}{p_\eta(y)}$$

Med två villkor är täljaren $p(x, y, z)$ själv, och nämnaren är den simultana marginalen för (η, ζ):

$$p_{\xi\,|\,\eta=y,\,\zeta=z}(x) = \frac{p(x,y,z)}{\sum_{x'} p(x',y,z)}$$

Båda gäller för de y respektive (y,z) där nämnaren är positiv.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.11 — Betyget utan seminarier

Använd tabellen över tentamensbetyg ξ och antal seminarier η i kompendiets övning 1.1. Beräkna väntevärdet av betyget för en student som inte har gått på något seminarium, och jämför med väntevärdet innan vi fick den informationen.

Lösning

Utan villkor. Radsummorna är $p_\xi(2), \dots, p_\xi(5) = 0{,}30;\ 0{,}35;\ 0{,}25;\ 0{,}10$, så

$$E(\xi) = 2 \cdot 0{,}30 + 3 \cdot 0{,}35 + 4 \cdot 0{,}25 + 5 \cdot 0{,}10 = 3{,}15$$

Givet $\eta = 0$. Kolumnen $y = 0$ har värdena $0{,}18;\ 0{,}11;\ 0{,}06;\ 0{,}02$ med summan $p_\eta(0) = 0{,}37$. Den betingade fördelningen är kolumnen delad med $0{,}37$:

$$p_{\xi\,|\,\eta=0}(x) \approx 0{,}486;\ 0{,}297;\ 0{,}162;\ 0{,}054 \qquad x = 2,3,4,5$$

$$E(\xi \mid \eta = 0) = \frac{2 \cdot 0{,}18 + 3 \cdot 0{,}11 + 4 \cdot 0{,}06 + 5 \cdot 0{,}02}{0{,}37} = \frac{1{,}03}{0{,}37} \approx \boxed{2{,}78}$$

Jämförelse. Det förväntade betyget sjunker från $3{,}15$ till $2{,}78$ när vi vet att studenten inte gått på något seminarium. Nästan hälften av de studenterna blir underkända. Det säger inget om orsaken: seminarierna kan hjälpa, men flitiga studenter kan också vara de som går dit.

På räknaren. One-Variable Statistics med `{2,3,4,5}` och **Frequency List** `{0.18,0.11,0.06,0.02}/0.37`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.12 — Väntevärdet är den bästa prognosen

Låt ξ vara en stokastisk variabel med väntevärdet μ. Visa att funktionen $f(c) = E((\xi - c)^2)$ antar sitt minsta värde då $c = \mu$.

Lösning

Utveckla kvadraten och använd att väntevärdet är linjärt och att c är en konstant:

$$f(c) = E(\xi^2 - 2c\,\xi + c^2) = E(\xi^2) - 2c\,\mu + c^2$$

Det är ett andragradspolynom i c med positiv c^2-koefficient.

Kvadratkomplettera. Lägg till och dra ifrån μ^2:

$$f(c) = \underbrace{E(\xi^2) - \mu^2}_{V(\xi)} + (c - \mu)^2$$

Första termen beror inte på c, och andra är ≥ 0 med likhet precis då $c = \mu$. Alltså antar f sitt minsta värde $\boxed{f(\mu) = V(\xi)}$ i $c = \mu$.

Tolkning. Variansen är det minsta förväntade kvadratiska fel som går att uppnå med en fast prognos. Gör vi samma räkning i den betingade fördelningen givet $\eta = y$ följer att $E(\xi \mid \eta = y)$ är den bästa prognosen när η är känd.

Resonemangsuppgift — det finns inget facit att jämföra med.

Kompendiet, övning 1.13 (Vännman 2.25, i 2:a uppl. 2.24) — Livslängd för ett batteri ur centrallagret

Batterierna i lektion 2, övning 6 kommer från fabrik A (50 %), B (20 %) och C (30 %). Ett batteri från A räcker mer än 100 drifttimmar med sannolikheten $0{,}95$, från B med $0{,}97$ och från C med $0{,}98$. Antag att livslängderna är exponentialfördelade med parametrarna λ_A, λ_B och λ_C.

a) Beräkna λ_A, λ_B och λ_C.

b) Beräkna den förväntade livslängden för ett batteri som tas på måfå ur centrallagret.

Lösning

a) För $\xi \in \mathrm{Exp}(\lambda)$ är $P(\xi > 100) = e^{-100\lambda}$. Sätt det lika med den givna sannolikheten och lös ut λ:

$$\lambda = \frac{-\ln p}{100}: \qquad \lambda_A = \frac{-\ln 0{,}95}{100} \approx \boxed{5{,}13 \cdot 10^{-4}}, \quad \lambda_B \approx \boxed{3{,}05 \cdot 10^{-4}}, \quad \lambda_C \approx \boxed{2{,}02 \cdot 10}$$

per timme. Väntevärdena $1/\lambda$ blir 1950, 3283 och 4950 timmar.

b) Låt η vara fabriken. Givet fabriken är den förväntade livslängden $1/\lambda$ för den fabriken, och lagen om totalt väntevärde väger ihop:

$$E(\xi) = 0{,}5 \cdot \frac{1}{\lambda_A} + 0{,}2 \cdot \frac{1}{\lambda_B} + 0{,}3 \cdot \frac{1}{\lambda_C} \approx 0{,}5 \cdot 1950 + 0{,}2 \cdot 3283 + 0{,}3 \cdot 4950 \approx \boxed{3116\ \text{ti}}$$

Fallgrop. Det är väntevärdena som ska vägas ihop, inte parametrarna. $1/(0{,}5\lambda_A + 0{,}2\lambda_B + 0{,}3\lambda_C) \approx 2646$ timmar är fel: blandningen av tre exponentialfördelningar är ingen exponentialfördelning.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.14 — Summa med ett slumpmässigt antal termer

Antag att $\xi_1, \xi_2, \xi_3, \ldots$ är $R(0,1)$-fördelade och att $\eta \in \mathrm{Po}(10)$ är oberoende av dem. Låt $\xi = \sum_{i=1}^{\eta} \xi_i$. Beräkna $E(\xi)$.

Lösning

Betinga på antalet termer. Givet $\eta = y$ är ξ en summa av y termer med väntevärdet $1/2$ vardera, och eftersom η är oberoende av termerna ändras inte deras fördelning av villkoret:

$$E(\xi \mid \eta = y) = \frac{y}{2} \quad \Longrightarrow \quad E(\xi \mid \eta) = \frac{\eta}{2}$$

Lagen om totalt väntevärde.

$$E(\xi) = E(\frac{\eta}{2}) = \frac{E(\eta)}{2} = \frac{10}{2} = \boxed{5}$$

Samma resonemang som mjölkexemplet i avsnitt 3. Kompendiets eget exempel 1.14 har där ett tryckfel, $0{,}2 \cdot 0{,}4$ i stället för $0{,}2 \cdot 4$, men svaret $0{,}8$ stämmer.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Marginalfördelning
- Flerdimensionella fördelningar
- Betingad sannolikhet
- Lagen om total sannolikhet
- Väntevärde
- Tvådimensionell normalfördelning

Resurser

- Wikipedia: Conditional probability distribution
- Wikipedia: Law of total expectation

Kovarians

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #storhet

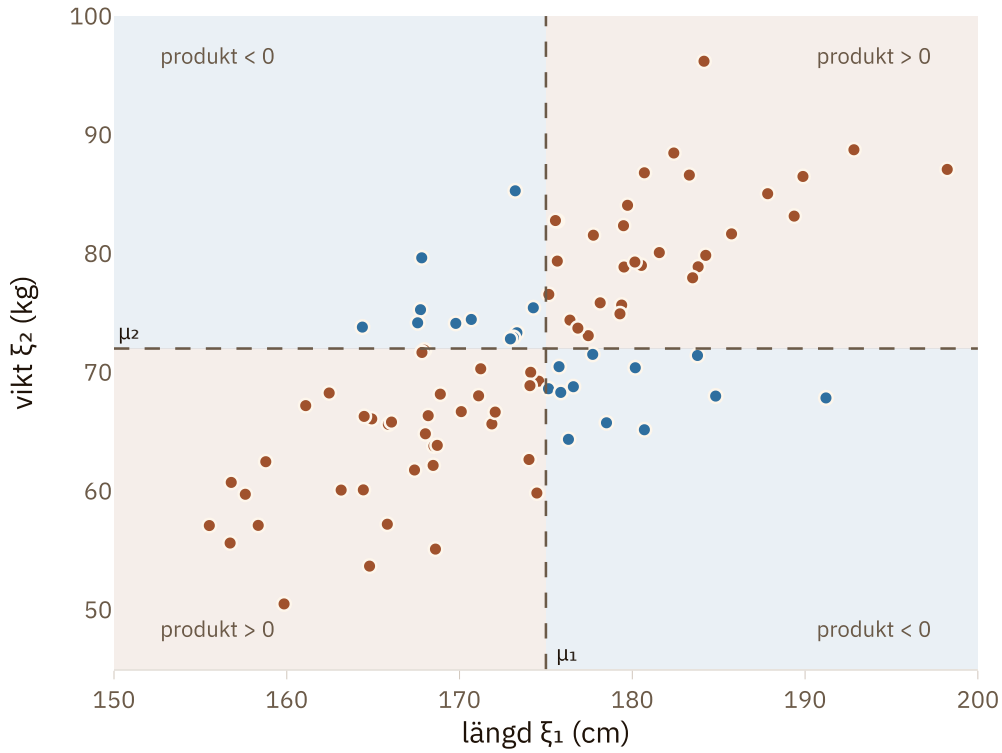
> **Kurs:** S0008M **Förkunskaper:** Flerdimensionella fördelningar, Varians och standardavvikelse

Kovariansen mäter hur två stokastiska variabler samvarierar: den är positiv om stora värden på den ena tenderar att åtföljas av stora värden på den andra, och negativ om de tenderar att gå åt olika håll. Den behövs för att räkna ut variansen av en summa av beroende variabler.

1. Idén

Låt ξ_1 och ξ_2 vara längd och vikt för en slumpvald person, med väntevärden μ_1 och μ_2. En person över medellängd väger oftast över medelvikten, och en person under medellängd oftast under. Avvikelserna $\xi_1 - \mu_1$ och $\xi_2 - \mu_2$ har alltså oftast samma tecken, så produkten $(\xi_1 - \mu_1)(\xi_2 - \mu_2)$ blir i genomsnitt positiv.

Vid ett negativt samband har avvikelserna oftast olika tecken och produkten blir i genomsnitt negativ. Saknas linjärt samband tar positiva och negativa produkter ut varandra.



Punkter i de rödbruna kvadranterna ger en positiv produkt $(\xi_1 - \mu_1)(\xi_2 - \mu_2)$, punkter i de blå en negativ. Här dominerar de rödbruna, så kovariansen är positiv. Håll musen över en punkt för att se produkten.

---

2. Definitionen

Definition 1.16 (kompendiet). **Kovariansen** mellan ξ_1 och ξ_2 är

$$\boxed{\operatorname{Cov}(\xi_1, \xi_2) = E((\xi_1 - \mu_1)(\xi_2 - \mu_2))}, \qquad \mu_i = E(\xi_i)$$

För räkning är omskrivningen

$$\boxed{\operatorname{Cov}(\xi_1, \xi_2) = E(\xi_1 \xi_2) - \mu_1 \mu_2}$$

oftast smidigast, där $E(\xi_1 \xi_2)$ räknas som en dubbelsumma eller dubbelintegral över den simultana fördelningen.

Två specialfall

- **Med sig själv:** $\operatorname{Cov}(\xi, \xi) = E((\xi - \mu)^2) = V(\xi)$. Variansen är kovariansen av en variabel med sig själv.
- **Oberoende:** om ξ_1 och ξ_2 är oberoende är $E(\xi_1 \xi_2) = \mu_1 \mu_2$, så $\operatorname{Cov}(\xi_1, \xi_2) = 0$.

---

3. Räkneregler

För konstanter a och b:

$$\begin{aligned}\operatorname{Cov}(\xi_1, \xi_2) &= \operatorname{Cov}(\xi_2, \xi_1)\\ \operatorname{Cov}(a\xi_1 + b,\ \xi_2) &= a\operatorname{Cov}(\xi_1, \xi_2)\\ \operatorname{Cov}(\xi_1 + \xi_2,\ \xi_3) &= \operatorname{Cov}(\xi_1, \xi_3) + \operatorname{Cov}(\xi_2, \xi_3)\end{aligned}$$

En additiv konstant flyttar variabeln men inte avvikelserna från väntevärdet, och därför försvinner b.

---

4. Variansen av en summa

Sats 1.17 (kompendiet).

$$\boxed{V(\xi_1 + \xi_2) = V(\xi_1) + V(\xi_2) + 2\operatorname{Cov}(\xi_1, \xi_2)}$$

och, eftersom $V(-\xi_2) = V(\xi_2)$ och $\operatorname{Cov}(\xi_1, -\xi_2) = -\operatorname{Cov}(\xi_1, \xi_2)$,

$$V(\xi_1 - \xi_2) = V(\xi_1) + V(\xi_2) - 2\operatorname{Cov}(\xi_1, \xi_2)$$

Varför. Utveckla kvadraten i definitionen av variansen:

$$V(\xi_1 + \xi_2) = E(((\xi_1 - \mu_1) + (\xi_2 - \mu_2))^2) = E((\xi_1 - \mu_1)^2) + E((\xi_2 - \mu_2)^2) + 2E((\xi_1 - \mu_1)(\xi_2 - \mu_2))$$

Är variablerna oberoende försvinner kovarianstermen och vi får tillbaka regeln att varianser adderas, se Linjära funktioner av stokastiska variabler.

Positivt beroende ökar spridningen i summan

Är kovariansen positiv drar variablerna åt samma håll och summan sprider mer än om de vore oberoende. För en differens är det tvärtom — det är därför stickprov i par ger snävare intervall: mätningarna i ett par är positivt korrelerade, och differensen tar bort den gemensamma variationen.

---

5. Skalberoende

Kovariansen har samma enhet som produkten av variablernas enheter. Mäter vi längd i meter i stället för centimeter blir kovariansen 100 gånger mindre, utan att sambandet har ändrats. För att få ett mått som inte beror på skalan dividerar vi med standardavvikelserna, och det ger korrelationskoefficienten:

$$\operatorname{Cov}(\xi_1, \xi_2) = \rho\,\sigma_1\sigma_2$$

Kovariansmatris

För flera variabler samlas varianser och kovarianser i en **kovariansmatris** C med elementen $c_{ij} = \operatorname{Cov}(\xi_i, \xi_j)$. I två dimensioner är

$$C = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Vanliga fel

- $\operatorname{Cov} = 0$ **betyder inte oberoende.** Oberoende ger kovariansen 0, men inte omvänt. Med $\xi \in R(-1, 1)$ och $\eta = \xi^2$ är $\operatorname{Cov}(\xi, \eta) = E(\xi^3) - 0 \cdot E(\xi^2) = 0$, fast η bestäms helt av ξ.
- **Glömd faktor** 2. $V(\xi_1 + \xi_2)$ har termen $2\operatorname{Cov}$, och för $a\xi_1 + b\xi_2$ blir den $2ab\operatorname{Cov}(\xi_1, \xi_2)$.
- **Tecknet vid differens.** I $V(\xi_1 - \xi_2)$ byter bara kovarianstermen tecken — varianserna adderas fortfarande.
- $E(\xi_1\xi_2) = \mu_1\mu_2$ **utan oberoende.** Produktregeln gäller bara för oberoende variabler; annars räknas $E(\xi_1\xi_2)$ som dubbelsumma eller dubbelintegral.

---

6. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 18, övning 3 — Hur stor kan variansen av en summa bli?

Låt ξ_1 och ξ_2 ha varianserna 9 respektive 25. Om de är beroende kan variansen av $\xi_1 + \xi_2$ vara antingen större eller mindre än 34, men inte hur stor eller liten som helst. Hur stor kan den högst vara? Hur liten kan den minst vara?

Lösning

Kompendiets sats 1.17 med $\operatorname{Cov}(\xi_1, \xi_2) = \rho\,\sigma_1\sigma_2 = 3 \cdot 5 \cdot \rho = 15\rho$:

$$V(\xi_1 + \xi_2) = 9 + 25 + 2 \cdot 15\rho = 34 + 30\rho$$

Korrelationen ligger alltid i $-1 \le \rho \le 1$, så

$$\boxed{4 \le V(\xi_1 + \xi_2) \le 64}$$

Maxvärdet $64 = (3+5)^2$ nås när $\rho = 1$ och variablerna drar exakt åt samma håll; minvärdet $4 = (5-3)^2$ när $\rho = -1$ och de delvis tar ut varandra.

Facit: $4 \le V(\xi_1 + \xi_2) \le 64$.

Kompendiet, övning 1.15 — Samma fråga med varianserna 4 och 9

Låt ξ_1 och ξ_2 ha varianserna 4 respektive 9. Hur stor kan variansen av $\xi_1 + \xi_2$ högst vara, och hur liten kan den minst vara?

Lösning

Nu är $\sigma_1 = 2$ och $\sigma_2 = 3$, så $\mathrm{Cov}(\xi_1, \xi_2) = 6\rho$ och

$$V(\xi_1 + \xi_2) = 4 + 9 + 12\rho = 13 + 12\rho$$

Med $-1 \le \rho \le 1$:

$$\boxed{1 \le V(\xi_1 + \xi_2) \le 25}$$

Gränserna är $(\sigma_2 - \sigma_1)^2$ och $(\sigma_1 + \sigma_2)^2$, precis som i lektionsuppgiften ovan.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Korrelation
- Flerdimensionella fördelningar
- Varians och standardavvikelse
- Linjära funktioner av stokastiska variabler
- Tvådimensionell normalfördelning

Resurser

- Khan Academy: Covariance and the regression line
- Wikipedia: Kovarians

Korrelation

19 okt. 2026, 6 min läsning

UTKAST

#matematik #statistik #storhet

> **Kurs:** S0008M **Förkunskaper:** Kovarians

Kovariansen mäter hur två stokastiska variabler samvarierar. Korrelationskoefficienten är den normerade varianten, som alltid ligger mellan -1 och 1 och mäter det linjära beroendet — korrelation noll betyder att linjärt samband saknas, men inte nödvändigtvis att variablerna är oberoende.

---

1. Definitionen

Definition 1.18 (kompendiet). **Korrelationskoefficienten** mellan ξ_1 och ξ_2 är

$$\boxed{\rho(\xi_1, \xi_2) = \frac{\operatorname{Cov}(\xi_1, \xi_2)}{\sqrt{V(\xi_1)\, V(\xi_2)}} = \frac{\operatorname{Cov}(\xi_1, \xi_2)}{\sigma_1 \, \sigma_2}}$$

Vi skriver ibland ρ_{12}, eller bara ρ när det är tydligt vilka variabler som avses. Divisionen med standardavvikelserna gör måttet dimensionslöst: det påverkas inte av om vi mäter i meter eller centimeter.

---

2. Gränserna

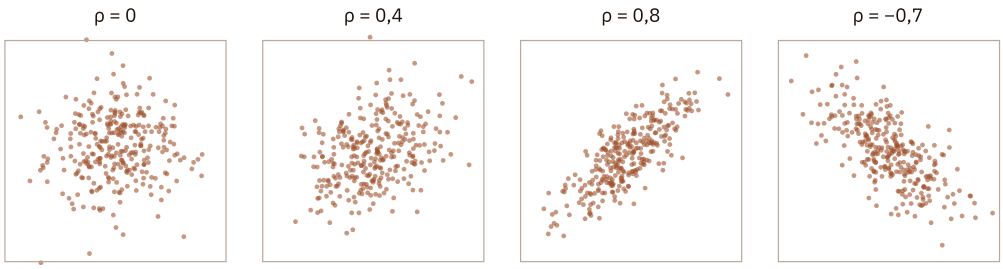
Det gäller alltid att

$$\boxed{-1 \le \rho(\xi_1, \xi_2) \le 1}$$

med likhet **om och endast om** ξ_2 är en linjär funktion av ξ_1, alltså $\xi_2 = a\xi_1 + b$. Tecknet på ρ är tecknet på a.

ρ	**Tolkning**
nära 1	starkt positivt linjärt samband

ρ	**Tolkning**
nära 0	inget linjärt samband — variablerna är **okorrelerade**
nära -1	starkt negativt linjärt samband



Molnen visar observationspar från fyra fördelningar. Ju närmare $|\rho|$ är 1, desto tätare ligger punkterna kring en linje.

3. Okorrelerad är inte oberoende

⚠ $\rho = 0$ **betyder inte oberoende**

Oberoende variabler har alltid $\mathrm{Cov} = 0$ och är alltså okorrelerade. Men det omvända gäller inte: ρ fångar bara **linjärt** beroende. Är (ξ, η) likformigt fördelad på en cirkelskiva är $\rho = 0$ av symmetriskäl, men variablerna är beroende — vet vi att ξ är nära 1 måste η vara nära 0.

Undantaget är den tvådimensionella normalfördelningen, där okorrelerade variabler också är oberoende.

4. Hur ρ påverkas av linjära omskrivningar

Ur räknereglerna för kovarians följer, för konstanter $a > 0$ och b:

$$\rho(a\xi_1 + b,\ \xi_2) = \rho(\xi_1, \xi_2), \qquad \rho(-\xi_1,\ \xi_2) = -\rho(\xi_1, \xi_2)$$

Att skala om eller flytta en variabel ändrar inte styrkan i sambandet, men att byta tecken vänder riktningen.

5. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 18, övning 1 — Kovarians och korrelation för omskrivna variabler

Slumpvariablerna ξ_1 och ξ_2 har väntevärden μ_1, μ_2 och standardavvikelser σ_1, σ_2. Låt $\rho = \rho(\xi_1, \xi_2)$. Använd definitionerna för att bestämma kovarians och korrelation nedan, och kommentera om korrelationen beter sig som ett mått på linjärt samband.

a) $\mathrm{Cov}(\xi_1, \xi_1)$ och $\rho(\xi_1, \xi_1)$

b) $\mathrm{Cov}(\xi_1, 2\xi_1)$ och $\rho(\xi_1, 2\xi_1)$

c) $\mathrm{Cov}(\xi_1 + a, \xi_2)$ och $\rho(\xi_1 + a, \xi_2)$, där a är en konstant

d) $\mathrm{Cov}(\xi_1, -\xi_2)$ och $\rho(\xi_1, -\xi_2)$

e) $\mathrm{Cov}(\xi_1 + \xi_2, \xi_3)$ och $\rho(\xi_1 + \xi_2, \xi_3)$, där ξ_3 har väntevärde μ_3 och standardavvikelse σ_3, och $\rho(\xi_1, \xi_3) = \rho_{13}$, $\rho(\xi_2, \xi_3) = \rho_{23}$

Lösning

a) $\mathrm{Cov}(\xi_1, \xi_1) = E((\xi_1 - \mu_1)^2) = \boxed{\sigma_1^2}$, och $\rho = \sigma_1^2/(\sigma_1\sigma_1) = \boxed{1}$. En variabel har perfekt linjärt samband med sig själv.

b) $2\xi_1$ har väntevärde $2\mu_1$ och standardavvikelse $2\sigma_1$:

$$\mathrm{Cov}(\xi_1, 2\xi_1) = E((\xi_1 - \mu_1)\cdot 2(\xi_1 - \mu_1)) = \boxed{2\sigma_1^2}, \qquad \rho = \frac{2\sigma_1^2}{\sigma_1 \cdot 2\sigma_1} = \boxed{1}$$

Kovariansen ändras med skalan, korrelationen inte.

c) $\xi_1 + a$ har väntevärde $\mu_1 + a$, så avvikelsen är oförändrad, $(\xi_1 + a) - (\mu_1 + a) = \xi_1 - \mu_1$:

$$\mathrm{Cov}(\xi_1 + a, \xi_2) = \boxed{\rho\,\sigma_1\sigma_2}, \qquad \rho(\xi_1 + a, \xi_2) = \boxed{\rho}$$

d) Avvikelsen för $-\xi_2$ är $-(\xi_2 - \mu_2)$ och standardavvikelsen fortfarande σ_2:

$$\operatorname{Cov}(\xi_1, -\xi_2) = \boxed{-\rho\,\sigma_1\sigma_2}, \qquad \rho(\xi_1, -\xi_2) = \boxed{-\rho}$$

Sambandet är lika starkt men går åt andra hållet.

e) Kovariansen är additiv i första argumentet:

$$\operatorname{Cov}(\xi_1 + \xi_2, \xi_3) = \operatorname{Cov}(\xi_1, \xi_3) + \operatorname{Cov}(\xi_2, \xi_3) = \boxed{\rho_{13}\,\sigma_1\sigma_3 + \rho_{23}\,\sigma_2\sigma_3}$$

Med kompendiets sats 1.17 är $V(\xi_1 + \xi_2) = \sigma_1^2 + \sigma_2^2 + 2\rho\,\sigma_1\sigma_2$, så

$$\rho(\xi_1 + \xi_2, \xi_3) = \boxed{\frac{\rho_{13}\,\sigma_1\sigma_3 + \rho_{23}\,\sigma_2\sigma_3}{\sigma_3\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\,\sigma_1\sigma_2}}}$$

Kommentar. I a)–d) beter sig ρ som väntat för ett mått på linjärt samband: ± 1 vid exakt linjärt samband, oförändrad vid flyttning och skalning, och teckenbyte när den ena variabeln byter tecken.

Facit: a) σ_1^2 resp. 1; b) $2\sigma_1^2$ resp. 1; c) $\rho\sigma_1\sigma_2$ resp. ρ; d) $-\rho\sigma_1\sigma_2$ resp. $-\rho$; e) som ovan.

Kompendiet, övning 1.16 — Räkneregler för kovarians och korrelation

Verifiera följande egenskaper, där a och b är konstanter:

a) $\operatorname{Cov}(a\xi_1 + b,\ \xi_2) = a\operatorname{Cov}(\xi_1, \xi_2)$

b) $\operatorname{Cov}(\xi_1 + \xi_2,\ \xi_3) = \operatorname{Cov}(\xi_1, \xi_3) + \operatorname{Cov}(\xi_2, \xi_3)$

c) $\rho(\xi_1,\ -\xi_2) = -\rho(\xi_1, \xi_2)$

d) $\rho(\xi_1,\ a\xi_1 + b)$ är lika med -1 eller 1 (för $a \neq 0$).

Lösning

Använd $\operatorname{Cov}(\xi_1, \xi_2) = E(\xi_1\xi_2) - \mu_1\mu_2$ och att väntevärdet är linjärt.

a) $E((a\xi_1 + b)\xi_2) = a\,E(\xi_1\xi_2) + b\,\mu_2$ och $E(a\xi_1 + b)\,E(\xi_2) = a\,\mu_1\mu_2 + b\,\mu_2$. Differensen är $a(E(\xi_1\xi_2) - \mu_1\mu_2) = a\,\mathrm{Cov}(\xi_1, \xi_2)$.

b) $E((\xi_1 + \xi_2)\xi_3) - (\mu_1 + \mu_2)\mu_3 = (E(\xi_1\xi_3) - \mu_1\mu_3) + (E(\xi_2\xi_3) - \mu_2\mu_3)$, vilket är högerledet.

c) Med a) och $a = -1, b = 0$ (rollerna ombytta, kovariansen är symmetrisk) är $\mathrm{Cov}(\xi_1, -\xi_2) = -\mathrm{Cov}(\xi_1, \xi_2)$. Standardavvikelsen för $-\xi_2$ är σ_2, så nämnaren i ρ är oförändrad och tecknet byts.

d) Med a) är $\mathrm{Cov}(\xi_1,\ a\xi_1 + b) = a\,V(\xi_1) = a\,\sigma_1^2$, och $D(a\xi_1 + b) = |a|\,\sigma_1$:

$$\rho(\xi_1,\ a\xi_1 + b) = \frac{a\,\sigma_1^2}{\sigma_1 \cdot |a|\,\sigma_1} = \frac{a}{|a|} = \boxed{\begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}}$$

Ett exakt linjärt samband ger alltså alltid $|\rho| = 1$, med tecknet efter lutningen.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.17 — Okorrelerade men beroende

Låt (ξ_1, ξ_2) vara likformigt fördelad på cirkelskivan med centrum i origo och radien 1. Bestäm $\mathrm{Cov}(\xi_1, \xi_2)$. Är ξ_1 och ξ_2 okorrelerade? Är de oberoende?

Lösning

Kovariansen. Arean är π, så $f(x_1, x_2) = 1/\pi$ på skivan. Skivan är symmetrisk kring båda axlarna, och integranderna x_1 och x_1x_2 är udda i x_1, så integralerna över skivan blir noll:

$$\mu_1 = \mu_2 = 0, \qquad E(\xi_1\xi_2) = \iint_{x_1^2 + x_2^2 \le 1} \frac{x_1 x_2}{\pi}\,dx_1\,dx_2 = 0$$

Alltså $\boxed{\mathrm{Cov}(\xi_1, \xi_2) = 0}$, och variablerna är **okorrelerade**.

Oberoende? Nej. Marginalerna är $f_{\xi_1}(x_1) = \frac{2}{\pi}\sqrt{1-x_1^2}$ på $[-1, 1]$, och likadant för ξ_2 (se Marginalfördelning). I punkten $(0{,}9;\ 0{,}9)$, som ligger utanför skivan, är $f = 0$ men produkten av marginalerna positiv. Alltså är $f(x_1, x_2) \neq f_{\xi_1}(x_1)\, f_{\xi_2}(x_2)$, och variablerna är **beroende**.

Det här är exemplet i avsnitt 3: vet vi att ξ_1 är nära 1 måste ξ_2 vara nära 0. Beroendet är inte linjärt, och därför syns det inte i ρ.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Kovarians
- Tvådimensionell normalfördelning
- Flerdimensionella fördelningar
- Oberoende stokastiska variabler

Resurser

- Khan Academy: Correlation coefficient
- Wikipedia: Korrelation

Tvådimensionell normalfördelning

19 okt. 2026, 9 min läsning

UTKAST

#matematik #sannolikhetslära #fördelning

> **Kurs:** S0008M **Förkunskaper:** Flerdimensionella fördelningar, Korrelation, Normalfördelning

---

Den tvådimensionella normalfördelningen beskriver två normalfördelade variabler tillsammans och bestäms helt av de två väntevärdena, de två standardavvikelserna och korrelationen. För just denna fördelning gäller att okorrelerade variabler också är oberoende.

---

1. Varför den behövs

Väljer vi två personer slumpmässigt och låter ξ_1 vara den enas längd och ξ_2 den andras vikt är de oberoende, och den simultana frekvensfunktionen är produkten av två normalfördelningar. Väljer vi i stället *en* person och mäter både längd och vikt är båda fortfarande normalfördelade var för sig, men de är inte oberoende. Vi behöver en modell där marginalerna är $N(\mu_1, \sigma_1)$ och $N(\mu_2, \sigma_2)$ men där variablerna får vara beroende.

---

2. Definitionen

Definition 1.20 (kompendiet). (ξ_1, ξ_2) har en **tvådimensionell normalfördelning** om den simultana frekvensfunktionen är

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right)$$

där $\sigma_1, \sigma_2 > 0$ och $-1 < \rho < 1$.

De fem parametrarna betyder: $\xi_1 \in N(\mu_1, \sigma_1)$, $\xi_2 \in N(\mu_2, \sigma_2)$ och $\rho = \rho(\xi_1, \xi_2)$ är korrelationen.

Okorrelerade ⇒ oberoende

Med $\rho = 0$ faller korstermen bort och $f(x_1, x_2)$ blir produkten av de två marginalerna. För tvådimensionellt normalfördelade variabler är okorrelerade alltså samma sak som oberoende — det gäller inte i allmänhet.

Nivåkurvorna är ellipser. Med $\rho > 0$ lutar de uppåt åt höger, och ju närmare $|\rho|$ är 1, desto smalare blir de.

---

3. Linjärkombinationer

Sats 1.21 (kompendiet). Om (ξ_1, ξ_2) har en tvådimensionell normalfördelning och c_1, c_2 är konstanter, inte båda noll, så är $c_1\xi_1 + c_2\xi_2$ (endimensionellt) normalfördelad.

Väntevärde och varians tar vi som vanligt, där kovariansen nu kommer med:

$$\boxed{\xi_1 \pm \xi_2 \in N(\mu_1 \pm \mu_2,\ \sqrt{\sigma_1^2 + \sigma_2^2 \pm 2\rho\,\sigma_1\sigma_2})}$$

Frågor om summor och differenser förs därmed över till den vanliga normalfördelningen och Φ-tabellen.

---

4. Betingad fördelning

Sats 1.23 (kompendiet). Den betingade fördelningen för ξ_2 givet $\xi_1 = x_1$ är $N(\mu_2', \sigma_2')$ med

$$\boxed{\mu_2' = \mu_2 + \rho\,\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \qquad \sigma_2' = \sqrt{1-\rho^2}\;\sigma_2}$$

- **Väntevärdet** flyttas längs en linje i x_1. Är $\rho > 0$ och $x_1 > \mu_1$ blir $\mu_2' > \mu_2$: ett stort värde på ξ_1 talar för ett stort värde på ξ_2.
- **Standardavvikelsen** minskar med faktorn $\sqrt{1-\rho^2}$ och beror inte på x_1. Information om ξ_1 minskar alltid osäkerheten om ξ_2 när variablerna är korrelerade.

Tvärsnittet genom ytan

Geometriskt är den betingade frekvensfunktionen ett lodrätt tvärsnitt genom $f(x_1, x_2)$ vid x_1, normerat så att arean blir 1. Tvärsnittet är en normalkurva med toppen på den streckade linjen i figuren i exemplet nedan.

5. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 18, övning 2 — Bertils tiokamp

De två första grenarna i tiokamp är 100 meter och längdhopp. Tiokamparen Bertil förväntar sig 950 respektive 1025 poäng. Resultaten (ξ_1, ξ_2) beskrivs med en tvådimensionell normalfördelning med väntevärden 950 och 1025, standardavvikelser $\sigma_1 = 120$ och $\sigma_2 = 190$ och korrelation $\rho_{12} = 0{,}55$.

a) Beräkna sannolikheten att det totala resultatet efter två grenar är minst 2100.

b) Beräkna sannolikheten att resultatet i längd blir bättre än resultatet på 100 m.

Resultatet på 100 meter blir ett personbästa på 11,34, vilket ger 1083 poäng.

c) Bertils personbästa i längd motsvarar 1240 poäng. Vad är den betingade sannolikheten att det blir personbästa även i längd?

d) Vad är det betingade väntevärdet av det totala resultatet efter två grenar?

e) I vilken mening ger svaret i d) den bästa prognosen för det totala resultatet, givet resultatet i första grenen?

Lösning

a) Enligt kompendiets sats 1.21 är summan normalfördelad:

$$\xi_1 + \xi_2 \in N(1975,\ \sqrt{120^2 + 190^2 + 2 \cdot 0{,}55 \cdot 120 \cdot 190}) = N(1975;\ 274{,}9)$$

$$P(\xi_1 + \xi_2 \geq 2100) = 1 - \Phi\left(\frac{2100 - 1975}{274{,}9}\right) = 1 - \Phi(0{,}455) \approx \boxed{0{,}325}$$

b) “Längd bättre än 100 m” är $\xi_2 - \xi_1 > 0$, och differensen är också normalfördelad:

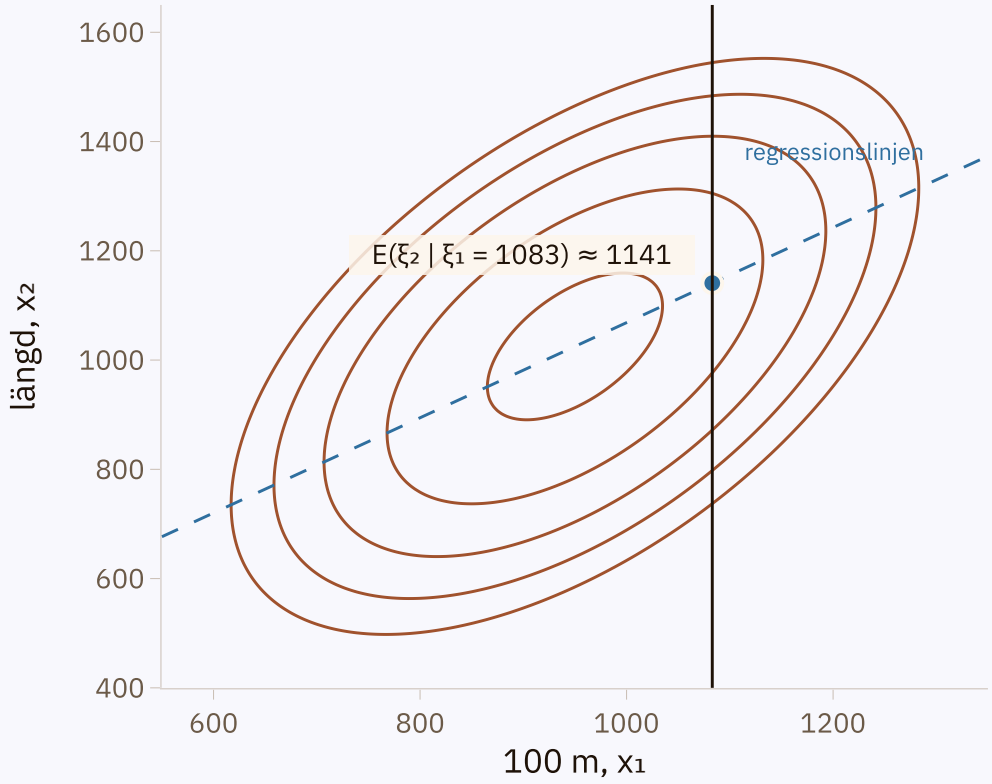
$$\xi_2 - \xi_1 \in N(75,\ \sqrt{120^2 + 190^2 - 2 \cdot 0{,}55 \cdot 120 \cdot 190}) = N(75;\ 159{,}4)$$

$$P(\xi_2 - \xi_1 > 0) = 1 - \Phi\left(\frac{0 - 75}{159{,}4}\right) = \Phi(0{,}470) \approx \boxed{0{,}681}$$

c) Kompendiets sats 1.23 med $x_1 = 1083$:

$$\mu_2' = 1025 + 0{,}55 \cdot \frac{190}{120} \cdot (1083 - 950) \approx 1140{,}8, \qquad \sigma_2' = \sqrt{1 - 0{,}55^2} \cdot 190 \approx 158{,}7$$

$$P(\xi_2 \geq 1240 \mid \xi_1 = 1083) = 1 - \Phi\left(\frac{1240 - 1140{,}8}{158{,}7}\right) = 1 - \Phi(0{,}625) \approx \boxed{0{,}266}$$



Figuren visar nivåkurvorna, linjen $x_1 = 1083$ och det betingade väntevärdet på den streckade linjen $\mu_2' = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$.

d) 100-metersresultatet är känt, så bara längden är slumpmässig:

$$E(\xi_1 + \xi_2 \mid \xi_1 = 1083) = 1083 + \mu_2' \approx \boxed{2223{,}8}$$

e) Det betingade väntevärdet är den prognos som minimerar det förväntade **kvadratiska** felet, givet informationen $\xi_1 = 1083$ — bästa prognosen i minsta-kvadrat-mening.

På räknaren. a) `normCdf(2100,∞,1975,274.9)`, b) `normCdf(0,∞,75,159.4)`, c) `normCdf(1240,∞,1140.8,158.7)`. Räknaren ger facits fyra decimaler direkt.

Facit: a) 0,3246; b) 0,6809; c) 0,2659; d) 2223,8; e) bästa prognosen i minsta-kvadrat-mening.

Kompendiet, övning 1.18 — Majas och Joels utgifter

Maja och Joel betraktar sina månatliga utgifter för kursmaterial (kr) som slumpmässiga. Majas utgifter har väntevärdet 545 och standardavvikelsen 60, Joels väntevärdet 495 och standardavvikelsen 55. Utgifterna har korrelationen $0{,}65$ och kan beskrivas med en tvådimensionell normalfördelning.

a) Vad är sannolikheten att deras sammanlagda utgifter under en månad överstiger 1100 kr?

b) Joel lägger 485 kr under en viss månad. Vad är sannolikheten att Majas utgifter samma månad överstiger 600 kr?

Lösning

Modell. ξ_1 = Majas utgifter, ξ_2 = Joels; (ξ_1, ξ_2) tvådimensionellt normalfördelad med $\mu_1 = 545$, $\sigma_1 = 60$, $\mu_2 = 495$, $\sigma_2 = 55$ och $\rho = 0{,}65$.

a) Summan är normalfördelad (kompendiets sats 1.21), med variansen från variansen av en summa:

$$D(\xi_1 + \xi_2) = \sqrt{60^2 + 55^2 + 2 \cdot 0{,}65 \cdot 60 \cdot 55} = \sqrt{10\,915} \approx 104{,}5$$

$$P(\xi_1 + \xi_2 > 1100) = 1 - \Phi\left(\frac{1100 - 1040}{104{,}5}\right) = 1 - \Phi(0{,}57) \approx \boxed{0{,}284}$$

Utan avrundning av z blir det $0{,}283$.

b) Kompendiets sats 1.23 med $x_2 = 485$ (rollerna ombytta, vi betingar på Joel):

$$\mu' = 545 + 0{,}65 \cdot \frac{60}{55}(485 - 495) \approx 537{,}9, \qquad \sigma' = 60\sqrt{1 - 0{,}65^2} \approx 45{,}6$$

$$P(\xi_1 > 600 \mid \xi_2 = 485) = 1 - \Phi\left(\frac{600 - 537{,}9}{45{,}6}\right) = 1 - \Phi(1{,}36) \approx \boxed{0{,}087}$$

Joel ligger under sitt väntevärde, och med positiv korrelation drar det ner förväntningen på Maja. Spridningen krymper också, från 60 till $45{,}6$.

På räknaren. a) `normCdf(1100,∞,1040,√(10915))`, b) `normCdf(600,∞,537.9,45.6)`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.19 — Bevis av den betingade fördelningen

Bevisa kompendiets sats 1.23: om (ξ_1, ξ_2) är tvådimensionellt normalfördelad är den betingade fördelningen för ξ_2 givet $\xi_1 = x_1$ en $N(\mu_2', \sigma_2')$-fördelning med $\mu_2' = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$ och $\sigma_2' = \sigma_2\sqrt{1 - \rho^2}$.

Lösning

Standardisera. Skriv $u = (x_1 - \mu_1)/\sigma_1$ och $v = (x_2 - \mu_2)/\sigma_2$. Hakparentesen i frekvensfunktionen är då $u^2 + v^2 - 2\rho uv$.

Kvadratkomplettera i v.

$$u^2 + v^2 - 2\rho uv = (v - \rho u)^2 + (1 - \rho^2)\,u^2$$

Frekvensfunktionen delar sig därför i två faktorer:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}\; e^{-u^2/2}\; e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}}$$

Dela med marginalen. $\xi_1 \in N(\mu_1, \sigma_1)$, så $f_{\xi_1}(x_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-u^2/2}$. Faktorn $e^{-u^2/2}$ och σ_1 försvinner:

$$f_{\xi_2\,|\,\xi_1=x_1}(x_2) = \frac{f(x_1,x_2)}{f_{\xi_1}(x_1)} = \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}}\; e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}}$$

Känn igen fördelningen. Multiplicera in σ_2 i täljaren och nämnaren i exponenten:

$$\frac{(v-\rho u)^2}{1-\rho^2} = \frac{(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1))^2}{\sigma_2^2(1-\rho^2)} = \frac{(x_2-\mu_2')^2}{(\sigma_2')^2}$$

Det är frekvensfunktionen för $\boxed{N(\mu_2',\ \sigma_2')}$ med μ_2' och σ_2' som i satsen. □

Resonemangsuppgift — det finns inget facit att jämföra med.

Kompendiet, övning 1.20 — Bygga en tvådimensionell normalfördelning

Antag att ζ_1 och ζ_2 är oberoende och $N(0,1)$. Finn konstanter $a_{11}, a_{12}, a_{21}, a_{22}$ så att $(a_{11}\zeta_1 + a_{12}\zeta_2,\ a_{21}\zeta_1 + a_{22}\zeta_2)$, flyttad till rätt väntevärden, har en tvådimensionell normalfördelning med $\mu_1 = 1, \mu_2 = 3, \sigma_1 = 3, \sigma_2 = 4$ och $\rho = 0{,}25$.

Lösning

Vad som krävs. Linjärkombinationerna har väntevärdet 0, så väntevärdena 1 och 3 måste läggas till som konstanter. Vi söker

$$V(\xi_1) = a_{11}^2 + a_{12}^2 = 9, \quad V(\xi_2) = a_{21}^2 + a_{22}^2 = 16, \quad \mathrm{Cov}(\xi_1,\xi_2) = a_{11}a_{21} + a_{12}a_{22} = 0{,}$$

Ett val. Låt ξ_1 bara bero av ζ_1: $a_{11} = 3, a_{12} = 0$. Kovariansvillkoret ger $3a_{21} = 3$, alltså $a_{21} = 1$, och variansvillkoret $a_{22} = \sqrt{15}$:

$$\boxed{\xi_1 = 1 + 3\zeta_1, \qquad \xi_2 = 3 + \zeta_1 + \sqrt{15}\,\zeta_2}$$

Kontroll: $V(\xi_2) = 1 + 15 = 16$ och $\rho = 3/(3\cdot 4) = 0{,}25$.

Varför det blir en tvådimensionell normalfördelning. Varje linjärkombination $c_1\xi_1 + c_2\xi_2$ är en linjärkombination av de oberoende normalfördelade ζ_1 och ζ_2 och alltså normalfördelad. Enligt omvändningen i kompendiets sats 1.21 är (ξ_1, ξ_2) då tvådimensionellt normalfördelad.

Valet är inte unikt: alla a_{ij} som uppfyller de tre villkoren fungerar.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.21 — Normalfördelade marginaler räcker inte

Låt $\xi_1 \in N(0,1)$ och $\xi_2 = \zeta\,\xi_1$, där ξ_1 och ζ är oberoende och $P(\zeta = -1) = P(\zeta = 1) = 0{,}5$.

a) Bestäm fördelningen för ξ_2.

b) Har (ξ_1, ξ_2) en tvådimensionell normalfördelning?

c) Har (ξ_1, ξ_2) en kontinuerlig tvådimensionell fördelning?

Lösning

a) Betinga på ζ med lagen om total sannolikhet och använd att $-\xi_1 \in N(0,1)$ av symmetri:

$$P(\xi_2 \le x) = 0{,}5\,P(\xi_1 \le x) + 0{,}5\,P(-\xi_1 \le x) = 0{,}5\,\Phi(x) + 0{,}5\,\Phi(x) = \Phi(x)$$

Alltså $\boxed{\xi_2 \in N(0,1)}$.

b) Nej. Med $\zeta = -1$ är $\xi_1 + \xi_2 = 0$, så $P(\xi_1 + \xi_2 = 0) = 0{,}5$. En normalfördelad variabel antar aldrig ett enskilt värde med positiv sannolikhet, så linjärkombinationen $\xi_1 + \xi_2$ är inte normalfördelad. Enligt kompendiets sats 1.21 är då (ξ_1, ξ_2) $\boxed{\text{inte tvådimensionellt normalfördelad}}$ — trots att båda marginalerna är $N(0,1)$.

c) Nej. Punkten (ξ_1, ξ_2) ligger alltid på någon av linjerna $x_2 = x_1$ och $x_2 = -x_1$. De har arean noll, och en dubbelintegral över ett område med arean noll är noll. Ingen frekvensfunktion kan därför ge linjerna sannolikheten 1, så $\boxed{\text{fördelningen är inte kontinuerlig}}$.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Normalfördelning
- Korrelation
- Kovarians
- Betingad fördelning och betingat väntevärde
- Stickprov i par

Resurser

- Wikipedia: Multivariate normal distribution — avsnittet om bivariata fallet

Icke-linjära funktioner av stokastiska variabler

19 okt. 2026, 10 min läsning

UTKAST

#matematik #sannolikhetslära #metod

> **Kurs:** S0008M **Förkunskaper:** Flerdimensionella fördelningar, Fördelningsfunktion, Linjära funktioner av stokastiska variabler

---

Fördelningen för en icke-linjär funktion $\zeta = g(\xi_1, \ldots, \xi_d)$ av stokastiska variabler tar vi fram i två steg: först fördelningsfunktionen $F_\zeta(z) = P(\zeta \leq z)$ som sannolikheten för ett område, sedan frekvensfunktionen genom derivering. När det inte går att räkna exakt kan vi i stället approximera väntevärde och varians genom linjärisering.

---

1. Fördelningsfunktionsmetoden

För linjära funktioner räcker räknereglerna för väntevärde och varians. För en godtycklig funktion $\zeta = g(\xi, \eta)$ behöver vi hela fördelningen, och den tar vi fram så här:

1. **Skriv händelsen som ett område.** $\{\zeta \leq z\} = \{(\xi, \eta) \in D_z\}$, där $D_z = \{(x, y) : g(x, y) \leq z\}$.
2. **Räkna sannolikheten.** $F_\zeta(z) = P((\xi, \eta) \in D_z)$ — en dubbelintegral av den simultana frekvensfunktionen, eller ett areaförhållande om fördelningen är likformig.
3. **Derivera.** $f_\zeta(z) = F_\zeta'(z)$.

> **Varför gå via F?**
>
> $P(\zeta \leq z)$ är sannolikheten för ett *område*, och sådana kan vi räkna. Frekvensfunktionen $f_\zeta(z)$ är ingen sannolikhet i sig, så den går inte att få fram direkt. Samma knep fungerar i en dimension: för $\eta = \xi^2$ är $F_\eta(y) = P(-\sqrt{y} \leq \xi \leq \sqrt{y})$.

---

2. Summor: faltningsformeln

För $\zeta = \xi + \eta$ är D_z halvplanet under linjen $y = z - x$:

$$F_\zeta(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x,y)\,dy\,dx$$

Är ξ och η oberoende är $f(x,y) = f_\xi(x)\,f_\eta(y)$, och derivering ger **faltningsformeln**

$$\boxed{f_\zeta(z) = \int_{-\infty}^{\infty} f_\xi(z-x)\,f_\eta(x)\,dx}$$

Summan av två oberoende rektangelfördelade variabler får på så sätt en triangelformad frekvensfunktion, och summan av n oberoende $\mathrm{Exp}(\lambda)$-variabler en Gammafördelning med frekvensfunktionen $\dfrac{\lambda^n}{(n-1)!} z^{n-1} e^{-\lambda z}$, $z > 0$.

---

3. Avstånd och kvadrater

Avstånd från origo. Är (ξ, η) likformigt fördelad på enhetscirkelskivan och $\zeta = \sqrt{\xi^2 + \eta^2}$ är $\{\zeta \le z\}$ en cirkelskiva med radie z, så

$$F_\zeta(z) = \frac{\pi z^2}{\pi} = z^2, \qquad f_\zeta(z) = 2z, \qquad 0 < z < 1$$

Frekvensfunktionen växer med z: det finns mer area långt ut än nära centrum.

Kvadrater. Är $\eta = \xi^2$ och ξ kontinuerlig ger steg 2 $F_\eta(y) = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y})$ för $y > 0$, och kedjeregeln

$$f_\eta(y) = \frac{1}{2\sqrt{y}} \left(f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})\right), \qquad y > 0$$

Med $\xi \in N(0,1)$ blir detta χ^2-fördelningen med 1 frihetsgrad. Summan av kvadraterna av k oberoende $N(0,1)$-variabler är χ^2-fördelad med k frihetsgrader.

Det minsta och största värdet av flera variabler är ett eget viktigt fall, se Max och min av stokastiska variabler.

---

4. Approximativt: Gauss approximationsformler

För en icke-linjär funktion av stokastiska variabler går väntevärde och varians i allmänhet inte att beräkna exakt. Vi linjäriserar då funktionen kring väntevärdena med en Taylorutveckling, vilket ger Gauss approximationsformler — grunden för felfortplantning.

För en variabel, med $\mu = E(\xi)$:

$$E(g(\xi)) \approx g(\mu), \qquad V(g(\xi)) \approx (g'(\mu))^2\, V(\xi)$$

och för oberoende $\xi_1, \ldots, \xi_n$:

$$V(g(\xi_1, \ldots, \xi_n)) \approx \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x_i}(\mu_1, \ldots, \mu_n)\right)^2 V(\xi_i)$$

Approximationen är bra när g är nästan linjär över det område där variablerna brukar hamna, alltså när standardavvikelserna är små.

⚠ Vanliga fel

- **Satt in direkt i frekvensfunktionen.** $f_\zeta(z)$ fås inte genom att sätta in g i f_ξ; gå via $F_\zeta(z) = P(\zeta \leq z)$ och derivera.
- **Intervallen.** Området D_z kan byta form med z — summan av två $R(0,1)$ har olika uttryck för $z < 1$ och $z > 1$. Ange också att $F_\zeta = 0$ under och 1 över värdemängden.
- **Bara ena roten.** För $\eta = \xi^2$ är $\{\eta \leq y\} = \{-\sqrt{y} \leq \xi \leq \sqrt{y}\}$ — båda rötterna ska med.
- **Gauss formler som exakta.** $E(g(\xi)) \approx g(\mu)$ är en approximation; likhet gäller bara när g är linjär.

---

5. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

☰ Lektion 19, övning 1 — Total väntetid för två bussar

En person ska ta två bussar. Väntetiderna är oberoende och $R(0, 10)$-fördelade. Låt ζ vara den totala väntetiden.

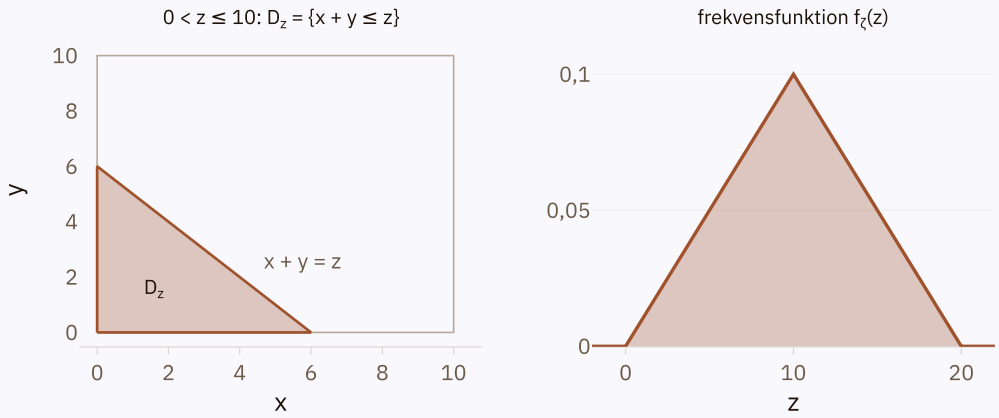
a) Beräkna fördelningsfunktionen för ζ.

b) Bestäm och skissa frekvensfunktionen för ζ.

✎ Lösning

Simultan fördelning. Med ξ och η som väntetiderna är $f(x,y) = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100}$ på kvadraten $K = [0,10] \times [0,10]$, alltså likformig fördelning med $|K| = 100$. Sannolikheter blir areor delat med 100.

a) $\{\zeta \le z\}$ är delen av kvadraten under linjen $x + y = z$.



- $0 < z \le 10$: området är en triangel med kateterna z, arean $z^2/2$.
- $10 < z \le 20$: komplementet är en triangel uppe till höger med kateterna $20 - z$.

$$F_\zeta(z) = \begin{cases} 0, & z \le 0 \\ \dfrac{z^2}{200}, & 0 < z \le 10 \\ 1 - \dfrac{(20-z)^2}{200}, & 10 < z \le 20 \\ 1, & z > 20 \end{cases}$$

b) Derivera bit för bit:

$$f_\zeta(z) = \begin{cases} \dfrac{z}{100}, & 0 < z \le 10 \\ \dfrac{20-z}{100}, & 10 < z \le 20 \\ 0, & \text{annars} \end{cases}$$

En triangel med toppen vid $z = 10$ (högra panelen). Två oberoende likformiga väntetider ger alltså **inte** en likformig summa: värden nära mitten kan uppstå på många sätt, värden nära 0 eller 20 bara på ett.

Rimlighetskoll. Triangelns area är $\frac{1}{2} \cdot 20 \cdot \frac{10}{100} = 1$.

Facit: $F_\zeta(z) = z^2/200$ *för* $0 < z \le 10$ *och* $1 - (20-z)^2/200$ *för* $10 < z \le 20$; f_ζ *är triangelformad med toppen i* $z = 10$.

Lektion 19, övning 3 — Partikel i ett klot

Positionen för en partikel relativt origo är likformigt fördelad på klotet med centrum i origo och radie $R > 0$.

a) Antag $R = 1$. Vad är sannolikheten att partikeln befinner sig högst $0{,}35$ från origo?

b) Låt ζ vara partikelns avstånd från origo. Diskutera hur frekvensfunktionen för ζ borde se ut.

c) Beräkna och skissa frekvensfunktionen för ζ för godtyckligt R. Bestäm även väntevärde och median för ζ.

Lösning

a) Likformig fördelning i tre dimensioner: sannolikheten är ett volymförhållande. $\{\zeta \le 0{,}35\}$ är ett klot med radie $0{,}35$:

$$P(\zeta \le 0{,}35) = \frac{\frac{4}{3}\pi \cdot 0{,}35^3}{\frac{4}{3}\pi \cdot 1^3} = 0{,}35^3 \approx \boxed{0{,}043}$$

b) Sannolikheten att hamna i en liten kub är ungefär $f\,dx_1\,dx_2\,dx_3$, och f är konstant. Ett tunt skal på avståndet z har volymen ungefär $4\pi z^2\,dz$, som växer med z. Frekvensfunktionen för ζ borde därför vara noll vid origo och växa ungefär som z^2 ut till R.

c) För $0 \le z \le R$ är $\{\zeta \le z\}$ ett klot med radie z:

$$F_\zeta(z) = \frac{\frac{4}{3}\pi z^3}{\frac{4}{3}\pi R^3} = \frac{z^3}{R^3}, \qquad f_\zeta(z) = F_\zeta'(z) = \boxed{\frac{3z^2}{R^3}}, \qquad 0 \le z \le R$$

Väntevärde:

$$E(\zeta) = \int_0^R z \cdot \frac{3z^2}{R^3}\,dz = \frac{3}{R^3} \cdot \frac{R^4}{4} = \boxed{0{,}750\,R}$$

Median: $F_\zeta(m) = 0{,}5$ ger $m^3/R^3 = 0{,}5$, alltså

$$m = \sqrt[3]{0{,}5}\,R \approx \boxed{0{,}794\,R}$$

Medianen ligger över väntevärdet eftersom fördelningen är vänsterskev: massan är samlad långt ut, med en lång svans in mot origo.

Facit: a) 0,042 — vi får 0,043, eftersom $0{,}35^3 = 0{,}042875$ *och facit har kortat bort decimalerna i stället för att avrunda; c)* $f_\zeta(z) = 3z^2/R^3$, $0 \le z \le R$, *väntevärdet* $0{,}750R$, *median* $0{,}794R$.

Kompendiet, övning 1.4 — Högst 12 minuters väntan

En person ska ta två bussar. Väntetiderna är oberoende och $R(0, 10)$-fördelade (minuter). Beräkna sannolikheten att den totala väntetiden blir högst 12 minuter.

Lösning

Det är fördelningsfunktionen från lektion 19, övning 1 ovan, i $z = 12$. Området $\{x + y \le 12\}$ är hela kvadraten utom en triangel uppe till höger med kateterna $20 - 12 = 8$:

$$P(\zeta \le 12) = 1 - \frac{8^2/2}{100} = 1 - 0{,}32 = \boxed{0{,}68}$$

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.5 — Summan av två $R(0, 1)$

Antag att $\xi_1 \in R(0, 1)$ och $\xi_2 \in R(0, 1)$ är oberoende. Beräkna fördelningsfunktionen och frekvensfunktionen för $\xi_1 + \xi_2$.

Lösning

Samma räkning som bussexemplet, med kvadraten $[0, 1] \times [0, 1]$ i stället för $[0, 10] \times [0, 10]$. Arean är 1, så sannolikheten är arean själv:

$$F(z) = \begin{cases} 0, & z \le 0 \\ \dfrac{z^2}{2}, & 0 < z \le 1 \\ 1 - \dfrac{(2-z)^2}{2}, & 1 < z \le 2 \\ 1, & z > 2 \end{cases} \qquad f(z) = \begin{cases} z, & 0 < z \le 1 \\ 2 - z, & 1 < z \le 2 \\ 0, & \text{annars} \end{cases}$$

Frekvensfunktionen är en triangel på $(0, 2)$ med toppen 1 i $z = 1$.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.7 — Den ovane pilkastaren

En ovan pilkastare träffar en cirkulär tavla med radien 1, och träffpunkten är likformigt fördelad på tavlan. Låt ζ vara avståndet från träffpunkten till tavlans centrum.

a) Bestäm variansen för ζ.

b) Pilkastaren får $10 - 10\zeta$ poäng. Beräkna det förväntade antalet poäng på ett kast.

Lösning

Fördelningen för ζ är avståndet från origo i avsnitt 3: $f_\zeta(z) = 2z$ för $0 < z < 1$.

a)

$$E(\zeta) = \int_0^1 z \cdot 2z \, dz = \frac{2}{3}, \qquad E(\zeta^2) = \int_0^1 z^2 \cdot 2z \, dz = \frac{1}{2}$$

$$V(\zeta) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \boxed{\frac{1}{18}} \approx 0{,}056$$

b) Poängen är en linjär funktion av ζ:

$$E(10 - 10\zeta) = 10 - 10 \cdot \frac{2}{3} = \boxed{\frac{10}{3}} \approx 3{,}3 \text{ poäng}$$

Det förväntade avståndet $2/3$ ligger närmare kanten än centrum, eftersom det finns mer area långt ut.

På räknaren. `∫(z*2*z,z,0,1)` och `∫(z^2*2*z,z,0,1)` i en Calculator-sida.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 1.9 — Avstånd i tre dimensioner

Låt (ξ_1, ξ_2, ξ_3) ha en tredimensionell kontinuerlig fördelning (som i två dimensioner, men med trippelintegraler). Bestäm väntevärdet av $\zeta = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}$ om

a) (ξ_1, ξ_2, ξ_3) är likformigt fördelad på klotet med centrum i origo och radien 1,

b) frekvensfunktionen är

$$f(x_1, x_2, x_3) = \frac{1}{\pi a_0^3}\, e^{-2\sqrt{x_1^2+x_2^2+x_3^2}/a_0}, \qquad a_0 = 5{,}2917720859 \cdot 10^{-11}$$

Den beskriver var väteatomens elektron befinner sig i förhållande till kärnan i grundtillståndet.

Lösning

a) Enligt lektion 19, övning 3 ovan, med $R = 1$, är $f_\zeta(z) = 3z^2$ på $(0, 1)$ och

$$E(\zeta) = \int_0^1 z \cdot 3z^2\, dz = \boxed{\frac{3}{4}}$$

b) Frekvensfunktionen beror bara på avståndet r till origo. Ett tunt skal på avståndet r har volymen $4\pi r^2\, dr$ (se Sfäriska koordinater), så

$$f_\zeta(r) = 4\pi r^2 \cdot \frac{1}{\pi a_0^3}\, e^{-2r/a_0} = \frac{4}{a_0^3}\, r^2 e^{-2r/a_0}, \qquad r > 0$$

Med substitutionen $t = 2r/a_0$ och $\int_0^\infty t^n e^{-t}\, dt = n!$:

$$E(\zeta) = \int_0^\infty r \cdot \frac{4}{a_0^3} r^2 e^{-2r/a_0}\, dr = \frac{4}{a_0^3}\left(\frac{a_0}{2}\right)^4 \cdot 3! = \boxed{\frac{3}{2} a_0} \approx 7{,}94 \cdot 10^{-11}$$

Samma teknik ger $\int_0^\infty f_\zeta(r)\, dr = \frac{4}{a_0^3}\left(\frac{a_0}{2}\right)^3 \cdot 2! = 1$, en kontroll av att f verkligen är en frekvensfunktion. Elektronen befinner sig i genomsnitt $1{,}5$ Bohrradier från kärnan.

På räknaren. b) `∫(r*4/a^3*r^2*e^(-2*r/a),r,0,∞)` med `a` odefinierad ger `3*a/2`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 2.4 — Brownsk rörelse

Låt ξ_1, ξ_2, ξ_3 vara oberoende, var och en $N(0, \sqrt{t})$. I en diffusionsmodell (Brownsk rörelse) är de koordinaterna för en mikroskopisk partikel vid tiden t. Beräkna sannolikheten att partikeln vid tiden $t = 1$ befinner sig högst $0{,}341$ från origo.

Lösning

Till χ^2. Vid $t = 1$ är $\xi_i \in N(0, 1)$, och enligt avsnitt 3 är $\xi_1^2 + \xi_2^2 + \xi_3^2$ då χ^2-fördelad med 3 frihetsgrader. Avståndet är högst $0{,}341$ precis när kvadratsumman är högst $0{,}341^2$:

$$P\left(\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2} \le 0{,}341\right) = P\left(\chi^2(3) \le 0{,}1163\right)$$

Tabellen. χ^2-tabellen ger att 3 frihetsgrader har $1\ \%$ av massan under $0{,}115$. Vårt $0{,}1163$ ligger alldeles ovanför, så

$$P \approx \boxed{0{,}01}$$

Med räknaren blir det $0{,}0102$.

På räknaren. `χ²Cdf(0,0.341^2,3)`, under `menu` → **Statistics** → **Distributions** → **χ² Cdf**.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Max och min av stokastiska variabler
- Linjära funktioner av stokastiska variabler
- Flerdimensionella fördelningar
- Fördelningsfunktion
- Sfäriska koordinater

Resurser

- Wikipedia: Convolution of probability distributions
- Wikipedia: Propagation of uncertainty — Gauss approximationsformler

Max och min av stokastiska variabler

19 okt. 2026, 5 min läsning

UTKAST

#matematik #sannolikhetslära #metod

> **Kurs:** S0008M **Förkunskaper:** Fördelningsfunktion, Oberoende stokastiska variabler

---

Det minsta och det största värdet av flera oberoende stokastiska variabler har enkla fördelningsfunktioner: $1-(1-F)^n$ respektive F^n. De beskriver till exempel livslängden hos en serie- respektive parallellkopplad krets.

---

1. Seriekoppling och parallellkoppling

Låt $\xi_1, \dots, \xi_n$ vara livslängderna hos n komponenter som går sönder oberoende av varandra.

- **Seriekopplad** krets fungerar tills *någon* komponent går sönder: livslängden är $\min(\xi_1, \dots, \xi_n)$.
- **Parallellkopplad** krets fungerar tills *alla* komponenter har gått sönder: livslängden är $\max(\xi_1, \dots, \xi_n)$.

Samma idé som för serie- och parallellsystem av händelser, men nu med tiden som stokastisk variabel.

---

2. Största värdet

$\max(\xi_1, \dots, \xi_n) \le x$ inträffar precis när **alla** $\xi_i \le x$. Med oberoende blir sannolikheten en produkt:

$$F_{\max}(x) = P(\xi_1 \le x) \cdots P(\xi_n \le x)$$

och är variablerna dessutom likafördelade med fördelningsfunktionen F:

$$\boxed{F_{\max}(x) = (F(x))^n}$$

---

3. Minsta värdet

För minimum är det lättare att räkna med komplementet: $\min(\xi_1, \dots, \xi_n) > x$ inträffar precis när **alla** $\xi_i > x$. Alltså

$$P(\min > x) = (1 - F(x))^n, \qquad \text{så} \qquad \boxed{F_{\min}(x) = 1 - (1 - F(x))^n}$$

Minimum av exponentialfördelningar

Är $\xi_i \in \text{Exp}(\lambda)$ är $P(\xi_i > x) = e^{-\lambda x}$, och

$$P(\min > x) = (e^{-\lambda x})^n = e^{-n\lambda x}$$

Minimum är alltså självt exponentialfördelat, $\min(\xi_1, \dots, \xi_n) \in \text{Exp}(n\lambda)$, med väntevärdet $\frac{1}{n\lambda}$. En seriekopplad krets med n komponenter håller i genomsnitt n gånger kortare tid än en enskild komponent.

Maximum av exponentialfördelningar är däremot ingen fördelning vi känner igen; med två komponenter och $\lambda = 1$ är $F_{\max}(x) = (1 - e^{-x})^2$.

Frekvensfunktionerna får vi genom derivering, som i fördelningsfunktionsmetoden.

---

4. Lösta lektionsuppgifter

Övningarna märkta *Kompendiet* kommer ur kompendiet om flerdimensionella fördelningar. Det saknar facit, så de är lösta två gånger oberoende av varandra.

Lektion 19, övning 2 — Serie- och parallellkopplad krets

En krets byggs av tio komponenter vars livslängder är oberoende och exponentialfördelade med väntevärde 100 timmar. Seriekopplad fungerar kretsen tills någon komponent går sönder; parallellkopplad tills alla har gått sönder.

a) Bestäm sannolikheten att kretsen fungerar i minst 150 timmar om komponenterna parallellkopplas.

b) Bestäm sannolikheten att kretsen fungerar i minst 150 timmar om komponenterna seriekopplas. Bestäm även fördelningen för ζ = tiden då kretsen slutar fungera. Känner du igen fördelningen?

Lösning

Väntevärdet 100 timmar ger $\lambda = 1/100 = 0{,}01$, och för en komponent är

$$P(\xi_i > 150) = e^{-0{,}01 \cdot 150} = e^{-1{,}5} \approx 0{,}2231$$

a) Parallellkopplad krets går sönder före 150 timmar bara om alla tio gör det:

$$P(\max > 150) = 1 - (1 - e^{-1{,}5})^{10} = 1 - 0{,}7769^{10} \approx \boxed{0{,}92}$$

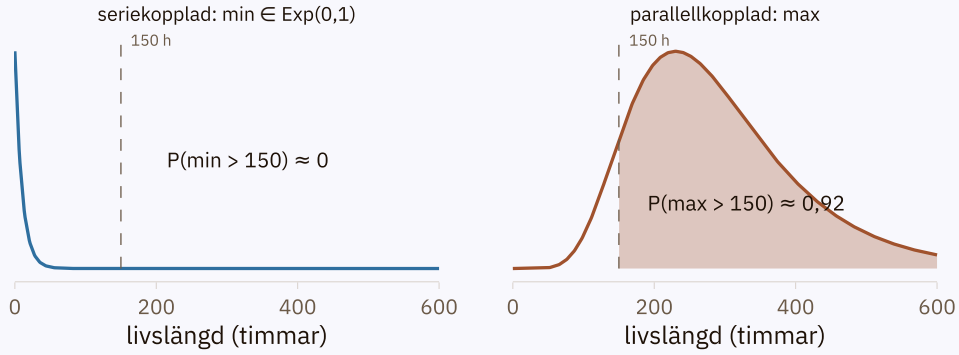
b) Seriekopplad krets kräver att alla tio håller:

$$P(\min > 150) = (e^{-1{,}5})^{10} = e^{-15} \approx \boxed{3 \cdot 10^{-7} \approx 0}$$

Fördelningen för $\zeta = \min(\xi_1, \ldots, \xi_{10})$: $P(\zeta > z) = e^{-10 \cdot 0{,}01\, z} = e^{-0{,}1z}$ för $z > 0$, alltså

$$\boxed{\zeta \in \text{Exp}(0{,}1)}$$

en exponentialfördelning med väntevärde 10 timmar.



Figuren visar livslängdens frekvensfunktion för de två kopplingarna. Den seriekopplade kretsen är nästan säkert trasig långt före 150 timmar, medan den parallellkopplade oftast håller betydligt längre.

På räknaren. a) `1-(1-e^(-1.5))^10`, b) `e^(-15)`.

Facit: a) parallellkoppling, sannolikheten $0{,}92$*; b) seriekoppling, sannolikheten* ≈ 0*,* $\zeta \in \text{Exp}(0{,}1)$.

Kompendiet, övning 2.1 — Fem seriekopplade komponenter

Beräkna den förväntade livslängden för en krets av 5 seriekopplade komponenter vars livslängder är oberoende och exponentialfördelade med väntevärdet 10 timmar.

Lösning

Varje komponent har $\lambda = 1/10$. Kretsens livslängd är minimum av fem, och enligt rutan i avsnitt 3 är $\min \in \text{Exp}(5 \cdot \frac{1}{10}) = \text{Exp}(0{,}5)$:

$$E(\min) = \frac{1}{0{,}5} = \boxed{2 \text{ timmar}}$$

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 2.2 — Agnes längsta väntetid

Agnes tar tunnelbanan till jobbet. Tågen går med 10 minuters intervall, men hon har varken tidtabell eller klocka, så hon betraktar väntetiderna olika dagar som oberoende stokastiska variabler. Beräkna sannolikheten att hennes längsta väntetid under en 10-dagarsperiod blir större än 8 minuter.

Lösning

Modell. Utan klocka kommer hon till perrongen vid en slumpmässig tidpunkt, så väntetiden $\xi_i \in R(0, 10)$ med $F(x) = x/10$ på $[0, 10]$, oberoende för $i = 1, \ldots, 10$.

Största värdet. $\max \leq 8$ kräver att alla tio väntetider är högst 8:

$$P(\max > 8) = 1 - F(8)^{10} = 1 - 0{,}8^{10} \approx \boxed{0{,}89}$$

En enskild dag är $P(\xi_i > 8) = 0{,}2$ liten, men på tio dagar är det nästan säkert att det händer minst en gång.

På räknaren. `1-0.8^10`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Kompendiet, övning 2.3 — Väntevärde och varians för ett maximum

Låt $\xi_1, \ldots, \xi_{10}$ vara oberoende och $R(0, b)$-fördelade, och $\zeta = \max(\xi_1, \ldots, \xi_{10})$. Beräkna $E(\zeta)$ och $V(\zeta)$.

Lösning

Fördelningen. $F(x) = x/b$ på $[0, b]$, så enligt avsnitt 2

$$F_\zeta(z) = \left(\frac{z}{b}\right)^{10}, \qquad f_\zeta(z) = \frac{10\,z^9}{b^{10}}, \qquad 0 \le z \le b$$

Väntevärdet.

$$E(\zeta) = \int_0^b z \cdot \frac{10\,z^9}{b^{10}}\,dz = \frac{10}{11}\,b, \qquad E(\zeta^2) = \int_0^b z^2 \cdot \frac{10\,z^9}{b^{10}}\,dz = \frac{10}{12}\,b^2$$

$$\boxed{E(\zeta) = \frac{10}{11}\,b \approx 0{,}909\,b}$$

Variansen.

$$V(\zeta) = \frac{10}{12}\,b^2 - \left(\frac{10}{11}\,b\right)^2 = \boxed{\frac{5}{726}\,b^2} \approx 0{,}0069\,b^2$$

Standardavvikelsen är $\approx 0{,}083\,b$: med tio observationer hamnar det största värdet nästan alltid nära b. Därför kan $\frac{11}{10}\max(\xi_i)$ användas som skattning av en okänd övre gräns b.

På räknaren. `∫(z*10*z^9/b^10,z,0,b)` med `b` odefinierad ger `10*b/11`.

Ej kontrollerat mot facit — två oberoende lösningar stämmer överens.

Se även

- Icke-linjära funktioner av stokastiska variabler
- Exponentialfördelning
- Oberoende händelser
- Fördelningsfunktion
- Oberoende stokastiska variabler

Resurser

- Wikipedia: Order statistic

S0008M – sammanfattning inför tentan

19 okt. 2026, 31 min läsning

UTKAST

#statistik #sannolikhetslära #sammanfattning #tenta

Kurs: S0008M — Sannolikhetslära och statistik

Grundtanken

Del 1 på tentan består av ungefär tolv korta uppgifter, och de kommer i nästan samma ordning varje gång. Den här texten går igenom dem i den ordningen. För varje uppgiftstyp får du veta vad den handlar om, hur du känner igen den, vad du gör steg för steg och vilka fällor som brukar kosta poäng. Texten är skriven för att lyssnas på: orden bär förklaringen, och formlerna står med för att du ska kunna slå upp dem efteråt.

1. Hur tentan ser ut och hur du använder texten

Tentan har två delar. Del 1 har ungefär tolv uppgifter som tillsammans ger tjugofem poäng, och för betyget tre behöver du sjutton av dem. Del 2 rättas bara om del 1 är godkänd. Den här sammanfattningen handlar om del 1, eftersom det är den som avgör om du klarar kursen.

Det viktigaste att veta om del 1 är att den är förutsägbar. Uppgift ett handlar nästan alltid om händelser, uppgift två om Bayes sats, uppgift tre till fem om en enskild fördelning, uppgift sex och sju om centrala gränsvärdessatsen, uppgift åtta till tio om statistisk inferens och uppgift elva och tolv om två slumpvariabler samtidigt. Figuren nedan visar den kartan. Om du vet var du är på kartan vet du också ungefär vilket verktyg som ska fram.

1	händelser: union, snitt, komplement	
2	Bayes sats, total sannolikhet	
3	en fördelning: binomial, Poisson …	
4	en fördelning: exponential, rektangel …	sannolikhetslära
5	normal ur kvantiler, väntevärde och varians	
6	centrala gränsvärdessatsen	
7	centrala gränsvärdessatsen	
8	test: kritisk gräns, P-värde	
9	konfidensintervall, oftast stickprov i par	statistisk inferens
10	styrka, teckenintervall, stickprovsstorlek	
11	två variabler: area, kovarians, 2D-normal	två variabler
12	två variabler: area, kovarians, 2D-normal	

Lyssna gärna på ett avsnitt i taget och försök sedan själv med en gammal tentauppgift av samma typ. Exemplen i texten är hopfällda, så att du kan läsa uppgiften först och öppna lösningen efteråt.

2. Lösningsmallen som ger delpoäng

Innan vi går in på uppgiftstyperna: varje lösning i del 1 bör följa samma fem steg. Läraren har sagt att den vanligaste orsaken till att delpoäng uteblir är att slumpvariabeln inte är definierad i ord. Ett rätt svar ger full poäng, men ett fel svar utan uppställning ger noll. Med en tydlig uppställning kan du få poäng även när du räknat fel.

Steg ett är att definiera slumpvariabeln i ord. Skriv till exempel: ξ är antalet trasiga muttrar i urvalet. Eller: ξ är tiden tills nästa kund kommer, mätt i minuter. Meningen ska säga vad som räknas eller mäts och i vilken enhet.

Steg två är att ange fördelningen med parametrar, till exempel $\xi \in \mathrm{Bin}(20;\ 0{,}05)$. I kursen skriver man tecknet “tillhör”, alltså $\in$, mellan slumpvariabeln och fördelningen. När en parameter är ett decimaltal skiljer man parametrarna med semikolon, så att decimalkommat inte förväxlas med skiljetecknet.

Steg tre är att skriva vad som ska beräknas, till exempel $P(\xi \leq 3)$.

Steg fyra är att räkna. Det räcker att skriva: jag beräknade $P(\xi \leq 3)$ med räknaren och fick $0{,}02275$. Du behöver inte skriva vilket räknarkommando du använde, men uttrycket måste stå där.

Steg fem är att svara exakt så som uppgiften ber om. Står det “svara i procent med två decimaler” ska svaret vara till exempel $55{,}43\ \%$, inte $55{,}4\ \%$. Avrunda bara i det sista steget, aldrig i mellanleden.

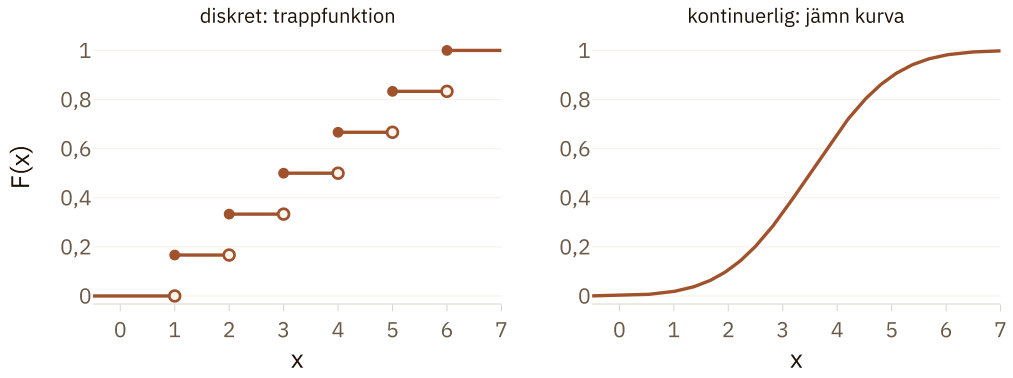
> **Lämna alltid in lösningen**
>
> Även i del 1, där bara svaret rättas, ger en uppställning med definierad slumpvariabel chansen till delpoäng.

3. Grundbegreppen som allt vilar på

Innan uppgiftstyperna kommer några begrepp som återkommer i nästan varje uppgift. Om de sitter blir resten mycket lättare.

En slumpvariabel är ett tal som bestäms av ett slumpmässigt försök. Antalet krona vid tio slantsinglingar är en slumpvariabel, liksom tiden tills nästa buss kommer. I kursen betecknas slumpvariabler med grekiska bokstäver, oftast ξ, som uttalas ksi, och η, som uttalas eta. Ett konkret värde som variabeln antar skrivs med vanlig bokstav, till exempel x.

Slumpvariabler är antingen diskreta eller kontinuerliga. En diskret variabel antar enskilda värden som du kan räkna upp, typiskt heltal. Den beskrivs av en sannolikhetsfunktion, som för varje värde anger sannolikheten att variabeln antar just det värdet. Summan av alla sannolikheterna är ett. En kontinuerlig variabel kan anta alla värden i ett intervall, och sannolikheten för ett enskilt exakt värde är noll. Den beskrivs i stället av en frekvensfunktion, där sannolikheten för ett intervall är arean under kurvan. Hela arean är ett.



Fördelningsfunktionen $F_\xi(x)$ finns för båda sorterna. Den anger sannolikheten att variabeln är mindre än eller lika med x. Den börjar på noll längst till vänster, växer och slutar på ett. Det är den som räknarens kumulativa kommandon räknar ut, och därför vill räknaren alltid ha “mindre än eller lika med”.

Väntevärdet $E(\xi)$ är det värde variabeln antar i genomsnitt i det långa loppet. Föreställ dig att försöket upprepas tusentals gånger och att du tar medelvärdet av alla utfall. För en diskret variabel är väntevärdet summan av varje värde gånger dess sannolikhet. För en kontinuerlig variabel är det motsvarande integral.

Variansen $V(\xi)$ mäter hur mycket värdena sprider sig kring väntevärdet. Den är väntevärdet av den kvadrerade avvikelsen från väntevärdet. Eftersom variansen har enheten i kvadrat använder man ofta standardavvikelsen $D(\xi)$, som är roten ur variansen och har samma enhet som variabeln själv. I kursen skrivs väntevärdet E, variansen V och standardavvikelsen D.

Två slumpvariabler är oberoende om utfallet av den ena inte säger något om den andra. Oberoende är ett antagande som uppgiften måste ge dig, med ord som “oberoende av varandra” eller ett sammanhang där det är uppenbart, till exempel olika tärningskast. Många av formlerna i kursen gäller bara vid oberoende, och det är en vanlig felkälla att använda dem ändå.

4. Räknaren på tentan

Du får använda TI-Nspire på tentan, och den gör nästan alla beräkningar åt dig. Det viktiga är att veta vilket kommando som hör till vilken fördelning och i vilken ordning argumenten ska skrivas. Fördelningarna finns i menyn Statistics och sedan Distributions.

För binomialfördelningen finns ett kommando för sannolikheten för exakt ett värde, med argumenten antal försök, sannolikhet och värdet. Det finns också ett kumulativt kommando som tar en undre och en övre gräns. Poissonfördelningen fungerar likadant, men med λ i stället för antal försök och sannolikhet. För normalfördelningen anger du undre gräns, övre gräns, väntevärde och standardavvikelse, och oändligheten finns i symbolpaletten. Den inversa normalfördelningen och den inversa t-fördelningen tar arean till vänster, och t-fördelningen dessutom antalet frihetsgrader.

För mätdata matar du in värdena i en lista och kör One-Variable Statistics. Utskriften ger medelvärdet och två standardavvikelser. Den som kallas sx delar med $n-1$ och är den som kursen kallar s. Den som kallas σx delar med n och ska inte användas. Vid stickprov i par skapar du en tredje kolumn med differenserna och kör statistiken på den.

5. Uppgift 1: händelser

Uppgift ett ger dig några sannolikheter för händelser, oftast två händelser A och B, och frågar efter sannolikheten för en kombination av dem. Det är den enklaste uppgiften på tentan, och den kommer varje gång.

Grundverktygen är tre. Det första är komplementet: sannolikheten att A inte inträffar, som skrivs $P(A^c)$, är ett minus sannolikheten att A inträffar. Det andra är unionen, alltså att minst en av händelserna inträffar. Den är summan av de två sannolikheterna minus snittet, eftersom snittet annars räknas två gånger:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

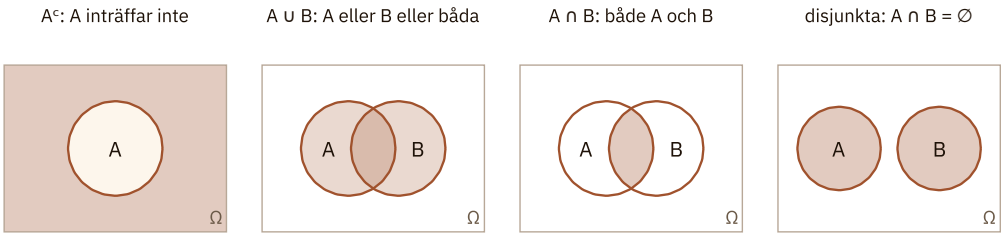
Det tredje är betingning: sannolikheten för A givet att B har inträffat är snittet delat med sannolikheten för B. Man krymper världen till B och frågar hur stor del av den som också är A.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Nästan alla varianter av uppgiften är kombinationer av de tre. “Varken A eller B” betyder att ingen av dem inträffar, och det är komplementet till unionen. “Exakt en av dem” är unionen minus snittet, alltså $P(A) + P(B) - 2P(A \cap B)$. “Minst en av n oberoende händelser” är lättast att räkna baklänges: ett minus sannolikheten att ingen inträffar, och om händelserna är oberoende är det produkten av komplementen.

Om händelserna är oberoende får du multiplicera: $P(A \cap B) = P(A)\,P(B)$. Det gäller bara vid oberoende, och uppgiften säger det i så fall uttryckligen. Om händelserna är beroende måste snittet ges i uppgiften, eller räknas ut via betingning: $P(A \cap B) = P(B)\,P(A \mid B)$.

Om alla utfall är lika troliga, som för tärningar och mynt, gäller den klassiska definitionen: antalet gynnsamma utfall delat med antalet möjliga. Två tärningar har trettiosex möjliga utfall.



Exempel 1 — varken A eller B (tentan 2025-03-17, uppgift 1)

För händelserna A och B gäller att $P(A) = 0{,}3$, $P(B) = 0{,}56$ och $P(A \cap B) = 0{,}14$. Beräkna $P(A^c \cap B^c)$ i procent med en decimal.

Lösning

“Varken A eller B” är komplementet till “minst en av dem”. Först unionen: $P(A \cup B) = 0{,}3 + 0{,}56 - 0{,}14 = 0{,}72$ Sedan komplementet: $P(A^c \cap B^c) = 1 - 0{,}72 = 0{,}28$, alltså $\boxed{28{,}0\ \%}$.

Fällor i uppgift 1

Glöm inte att dra bort snittet i unionen. Multiplicera inte sannolikheter om uppgiften inte säger att händelserna är oberoende. Och läs noga om frågan gäller “givet att”, för då ska du dela med sannolikheten för villkoret.

6. Uppgift 2: Bayes sats och total sannolikhet

Uppgift två handlar om att vända på en betingning. Typiskt vet du hur stor andel av produkterna som kommer från olika fabriker, och hur stor andel av varje fabriks produkter som är defekta. Frågan är sedan: givet att en produkt är defekt, hur troligt är det att den kommer från fabrik A? Samma struktur finns i medicinska test, där du vet hur ofta testet slår ut hos sjuka och friska, och frågan är hur troligt det är att en person med positivt test faktiskt är sjuk.

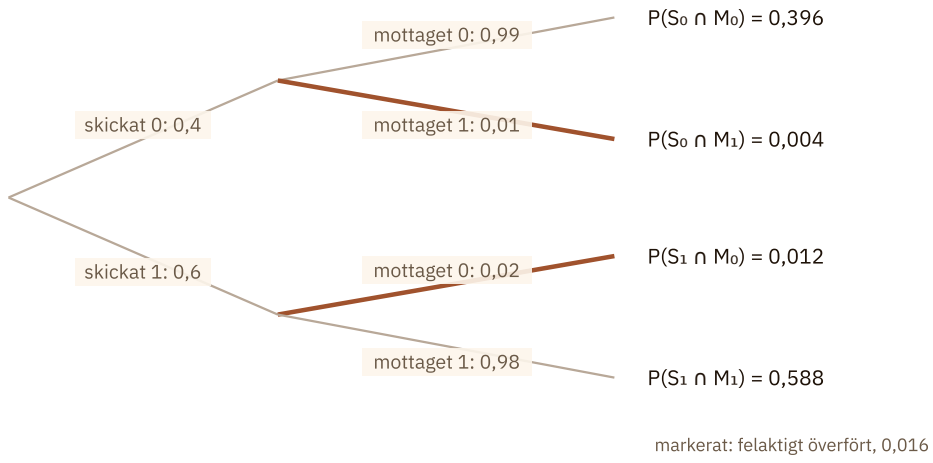
Du löser den i två steg. Först räknar du ut den totala sannolikheten för det du har observerat, till exempel att en produkt är defekt. Det gör du genom att gå igenom varje källa: sannolikheten att produkten kommer från källan gånger sannolikheten att den är defekt givet källan, summerat över alla källor. Det är lagen om total sannolikhet:

$$P(A) = P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2) + \cdots + P(B_n)\,P(A \mid B_n)$$

Sedan använder du Bayes sats: den del av den totala sannolikheten som kommer från just den källa du frågar om, delat med hela den totala sannolikheten.

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{P(A)}$$

Ett träddiagram är det säkraste sättet att hålla ordning. Den första förgreningen är källorna, den andra är utfallet. Varje gren längst ut har en sannolikhet som är produkten längs vägen. Täljaren i Bayes sats är en enda sådan gren, nämnaren är summan av alla grenar som slutar i det observerade utfallet.



Ett annat bra sätt att tänka är med naturliga frekvenser. Föreställ dig tiotusen personer och räkna hur många som hamnar i varje gren. Då blir det uppenbart varför ett test med hög träffsäkerhet ändå kan ge många falska larm när sjukdomen är ovanlig: de friska är så många fler att även en liten andel felaktiga positiva svar blir stor i antal.



Exempel 2 — vilken signal skickades? (tentan 2025-03-17, uppgift 2)

Ett utsänt tecken 0 tas emot som 1 med sannolikheten $0{,}02$. Ett utsänt tecken 1 tas emot som 0 med sannolikheten $0{,}035$. Sextio procent av alla utsända tecken är ettor. Om en etta tagits emot, vad är sannolikheten att en etta skickades? Svara i procent med två decimaler.

Lösning

Låt S_1 vara att en etta skickades och M_1 att en etta togs emot. Total sannolikhet: $P(M_1) = 0{,}6 \cdot 0{,}965 + 0{,}4 \cdot 0{,}02 = 0{,}579 + 0{,}008 = 0{,}587$ Bayes sats: $P(S_1 \mid M_1) = \dfrac{0{,}579}{0{,}587} = 0{,}98637 \approx \boxed{98{,}64\ \%}$

Fällor i uppgift 2

Betinga åt rätt håll: uppgiften ger dig $P(\text{defekt} \mid \text{fabrik})$ men frågar efter $P(\text{fabrik} \mid \text{defekt})$. Kontrollera att källornas sannolikheter summerar till ett. Och räkna inte med tal som inte behövs, uppgiften innehåller ibland överflödig information.

Ibland dyker Bayes upp inne i en fördelning, till exempel: givet att en komponent har hållit i minst fyra år, vad är sannolikheten att den håller i minst sju? Det är samma betingningsformel. Snittet av “minst sju” och “minst fyra” är bara “minst sju”, så kvoten blir sannolikheten för minst sju delat med sannolikheten för minst fyra.

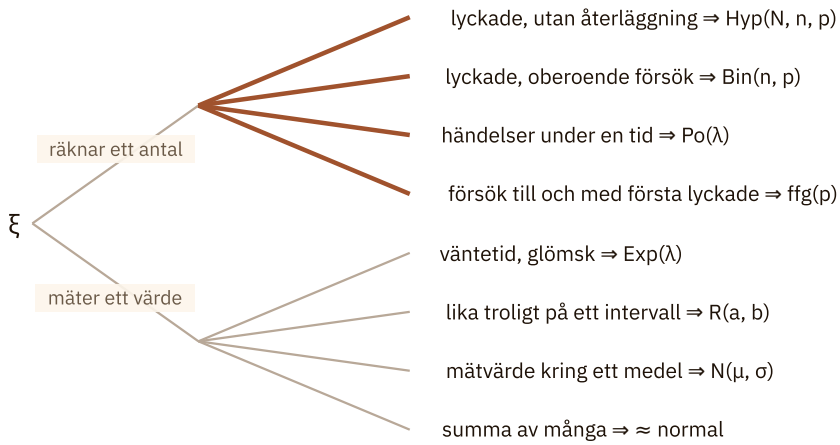
7. Uppgift 3 till 5: välj rätt fördelning

Uppgift tre, fyra och fem handlar oftast om en enda slumpvariabel och en känd fördelning. Det svåra är sällan räkningen, eftersom räknaren gör den. Det svåra är att känna igen vilken fördelning det är. Ställ dig två frågor.

Den första frågan är: räknar jag ett antal, eller mäter jag ett värde? Ett antal är ett heltal som noll, ett, två och så vidare. Ett värde är en mätning som kan anta alla tal i ett intervall, till exempel en tid, en längd eller en vikt.

Om du räknar ett antal finns fyra kandidater. Räknar du hur många av n oberoende försök som lyckas, och sannolikheten är densamma varje gång, är det binomialfördelningen. Drar du utan att lägga tillbaka ur en liten hög, så att sannolikheten ändras efter varje dragning, är det den hypergeometriska fördelningen. Räknar du hur många händelser som inträffar under en viss tid, till exempel samtal per timme, är det Poissonfördelningen. Räknar du hur många försök som behövs till och med det första lyckade, är det för-första-gången-fördelningen, som i kursen kallas ffg-fördelningen.

Om du mäter ett värde finns också fyra kandidater. En väntetid utan minne, där det inte spelar någon roll hur länge man redan väntat, är exponentialfördelad. Ett värde som är lika troligt överallt på ett intervall är rektangelfördelat. Ett mätvärde som sprider sig symmetriskt kring ett medelvärde är normalfördelat. Och en summa eller ett medelvärde av många oberoende variabler är ungefär normalfördelat enligt centrala gränsvärdessatsen.



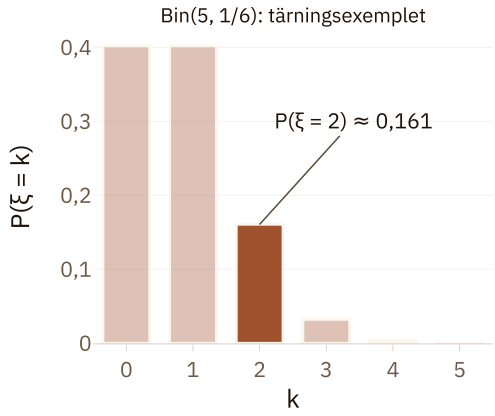
Väntevärde och varians för alla de vanliga fördelningarna står på formelbladet, så dem behöver du inte kunna utantill. Det du behöver kunna är att välja fördelning och ställa upp vad som ska räknas.

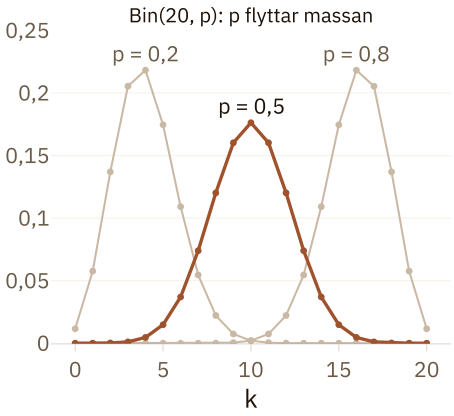
8. Binomial och hypergeometrisk fördelning

Binomialfördelningen beskriver antalet lyckade försök bland n försök, när varje försök lyckas med samma sannolikhet p och försöken är oberoende. Skriv: ξ är antalet lyckade bland de n försöken, $\xi \in \mathrm{Bin}(n, p)$.

På räknaren finns två kommandon. Det ena ger sannolikheten för exakt ett värde, till exempel exakt tre lyckade. Det andra ger summan av sannolikheterna mellan två gränser. Räknaren räknar bara “mindre än eller lika med”, så “minst k” måste vändas: sannolikheten för minst tre är ett minus sannolikheten för högst två. Det vanligaste felet är att skriva högst tre i stället för högst två.

$$P(\xi \geq k) = 1 - P(\xi \leq k - 1)$$



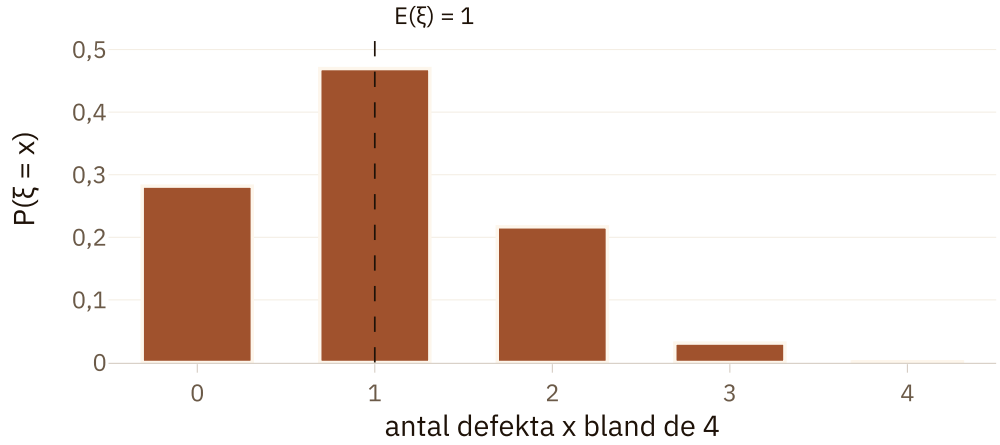


Ibland står inte p direkt i uppgiften. Då räknar du först ut p som sannolikheten att ett enskilt försök lyckas, och använder den sedan i binomialfördelningen. Det kallas två steg. Ett lyckat försök kan till exempel vara att två oberoende egenskaper inträffar samtidigt, och då är p produkten av de två sannolikheterna. Eller så är ett lyckat försök att en exponentialfördelad livslängd överstiger en viss tid, och då är $p = e^{-\lambda t}$.

Det mest sannolika värdet för en binomialfördelad variabel är heltalsdelen av $(n + 1)p$. För $\text{Bin}(11;\ 0{,}15)$ blir det heltalsdelen av $1{,}8$, alltså ett.

Den hypergeometriska fördelningen används när du drar utan återläggning ur en ändlig mängd, till exempel kort ur en kortlek eller bollar ur en låda. Den har tre parametrar: hur många som finns totalt, hur många du drar, och andelen lyckade från början. Sannolikhetsfunktionen, väntevärdet och variansen står på formelbladet, och på räknaren räknar du binomialkoefficienterna med kommandot nCr. Om du drar en liten del av en stor mängd, mindre än en tiondel, kan du approximera med binomialfördelningen.

Hyp(20, 4, 0,25): 4 dras ur 20, varav 5 defekta



Exempel 3 — exakt tre av tolv (två steg)

Ett jordbruk odlar ekologiskt med sannolikheten $0{,}61$ och har mjölkkor med sannolikheten $0{,}87$, oberoende av varandra. Man tittar på tolv jordbruk. Vad är sannolikheten att exakt tre av dem både odlar ekologiskt och har mjölkkor? Svara i procent med två decimaler.

Lösning

Ett lyckat jordbruk gör båda sakerna, så $p = 0{,}61 \cdot 0{,}87 = 0{,}5307$. Låt ξ vara antalet sådana jordbruk av de tolv, $\xi \in \text{Bin}(12;\ 0{,}5307)$. $P(\xi = 3) = \binom{12}{3} 0{,}5307^3\,(1 - 0{,}5307)^9 \approx 0{,}03631 = \boxed{3{,}63\ \%}$ På räknaren: `binomPdf(12,0.5307,3)`. Att de övriga nio "kombinerar på något annat sätt" ingår redan: de är de misslyckade försöken.

Exempel 4 — röda bollar utan återläggning (tentan 2025-03-17, uppgift 5)

I en låda finns 29 bollar, varav 8 är röda. Man drar 4 utan återläggning. Vad är standardavvikelsen för antalet röda bollar? Svara med två decimaler.

Lösning

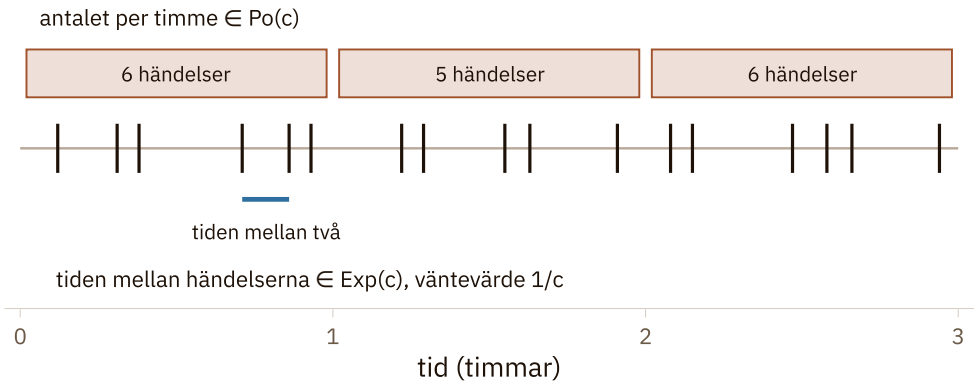
ξ är antalet röda bland de fyra, $\xi \in \text{Hyp}(29;\ 4;\ 8/29)$. Formelbladets varians: $V(\xi) = n\,p(1-p)\,\dfrac{N-n}{N-1} = 4 \cdot \dfrac{8}{29} \cdot \dfrac{21}{29} \cdot \dfrac{25}{28} \approx 0{,}713$ Standardavvikelsen är roten: $D(\xi) \approx \boxed{0{,}84}$.

9. Poisson och exponential: två sidor av samma process

Poissonfördelningen beskriver hur många händelser som inträffar under en viss tid, när händelserna inträffar oberoende av varandra och med en jämn intensitet. Parametern λ är det förväntade antalet under den tid du tittar på, alltså intensiteten gånger tiden. Om det i medel kommer sex kunder per timme och du tittar på en halvtimme är $\lambda = 3$. Väntevärdet och variansen är båda λ, så om uppgiften ger standardavvikelsen är λ standardavvikelsen i kvadrat.

En summa av oberoende Poissonvariabler är också Poissonfördelad, och parametrarna adderas. Det betyder att du ofta kan räkna exakt med räknarens Poissonkommando även när det handlar om en summa av många variabler. När λ är stor går det också att approximera med normalfördelningen, där både väntevärde och varians är λ. Villkoren för approximationen står på formelbladet. Facit har räknat på båda sätten på olika tentor, så räkna helst exakt och skriv gärna att du gjort det.

Exponentialfördelningen beskriver väntetiden till nästa händelse i samma slags process. Det är kopplingen som är lätt att missa. Om händelserna kommer med intensiteten c per timme, så att antalet per timme är $\mathrm{Po}(c)$, är tiden mellan två händelser exponentialfördelad med parametern c och väntevärdet $1/c$ timmar. Sex block per timme betyder alltså i genomsnitt tio minuter mellan blocken.



För exponentialfördelningen gäller att sannolikheten att vänta längre än x är $e^{-\lambda x}$. Det är den formel du använder mest. Fördelningsfunktionen är ett minus den. Medianen, alltså den tid som hälften av alla väntetider understiger, är $\ln 2$ delat med λ, och den är alltid kortare än väntevärdet. Exponentialfördelningen är glömsk: om du redan väntat i fem minuter är sannolikheten att vänta minst fem minuter till precis densamma som från början.

Exempel 5 — från Poisson till exponential

Block skapas i medel sex gånger per timme. Vad är sannolikheten att det dröjer mer än tjugo minuter till nästa block?

Lösning

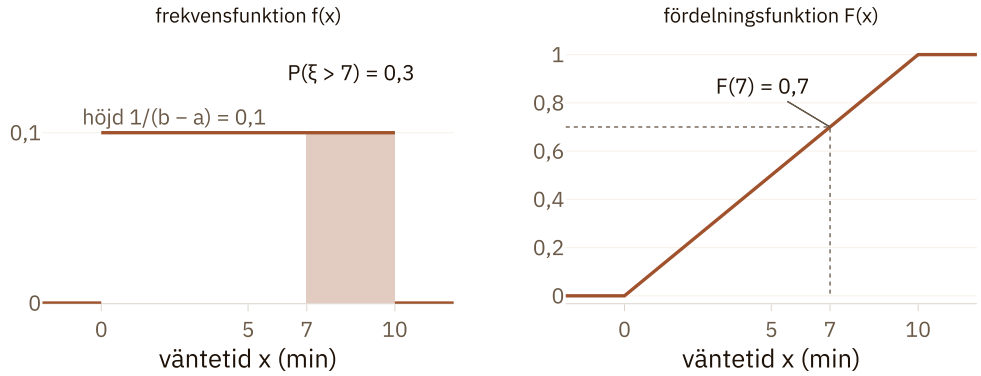
Låt ξ vara tiden till nästa block i timmar. Intensiteten sex per timme ger $\xi \in \mathrm{Exp}(6)$. Tjugo minuter är en tredjedels timme: $P\left(\xi > \tfrac{1}{3}\right) = e^{-6\cdot 1/3} = e^{-2} \approx \boxed{0{,}135}$

Fällor med Poisson och exponential

Kontrollera tidsenheten. Om intensiteten anges per timme och frågan gäller minuter måste du räkna om. “Fler än fem” betyder ett minus sannolikheten för högst fem.

10. Rektangel, ffg och frekvensfunktioner

En rektangelfördelad variabel är lika trolig överallt på ett intervall från a till b. Sannolikheten att den hamnar i ett delintervall är delintervallets längd delat med hela intervallets längd. Det enda du måste komma ihåg är att först klippa delintervallet mot $[a, b]$: den del som ligger utanför har sannolikheten noll.



Ffg-fördelningen räknar antalet försök till och med det första lyckade. Sannolikheten att det behövs fler än k försök är att de k första misslyckas, alltså $(1-p)^k$. “Minst fyra försök” betyder att de tre första misslyckas. Väntevärdet är ett delat med p.

Exempel 6 — minst fyra försök

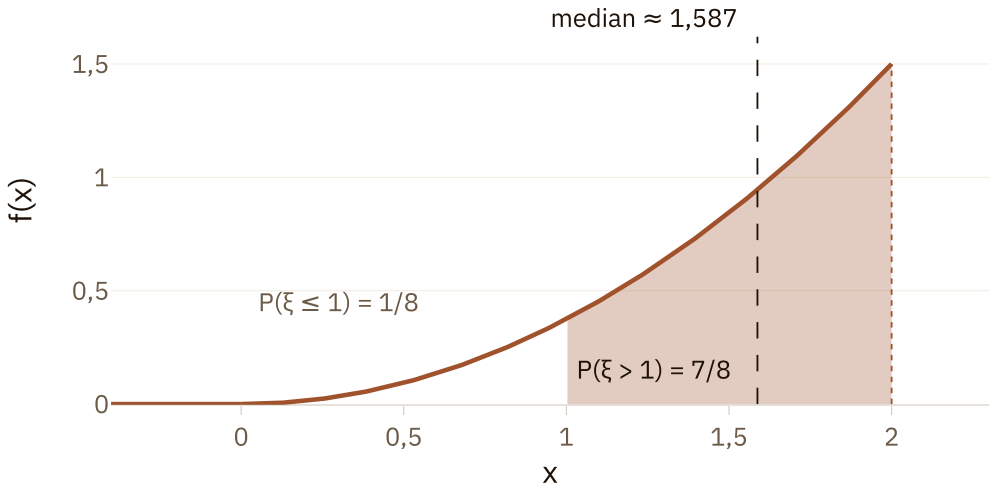
Ett försök lyckas med sannolikheten $0{,}3$, oberoende av tidigare försök. Vad är sannolikheten att det behövs minst fyra försök?

Lösning

Låt ξ vara antalet försök till och med det första lyckade, $\xi \in \text{ffg}(0{,}3)$. Minst fyra försök betyder att de tre första misslyckas: $P(\xi \geq 4) = P(\xi > 3) = 0{,}7^3 = \boxed{0{,}343}$

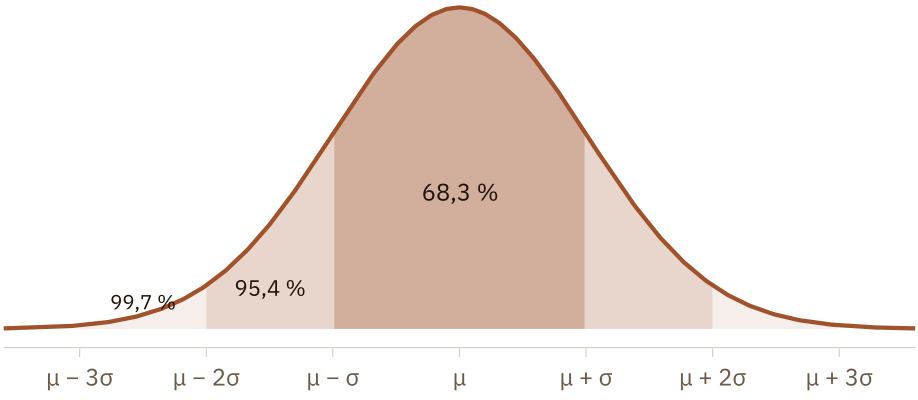
En del uppgifter ger dig en frekvensfunktion, alltså en funktion $f_\xi(x)$ som beskriver hur tätt sannolikheten ligger. Sannolikheten för ett intervall är arean under kurvan över intervallet, alltså en integral. Om funktionen innehåller en okänd konstant bestämmer du den genom att hela arean ska vara ett. Väntevärdet är integralen av x gånger frekvensfunktionen. Medianen är det värde där fördelningsfunktionen, alltså arean till vänster, är en halv.

f(x) = 3x²/8 på [0, 2]: sannolikheten är arean



11. Normalfördelningen och kvantiler

Normalfördelningen beskrivs av sitt väntevärde μ och sin standardavvikelse σ. I kursen skrivs den $N(\mu, \sigma)$ med standardavvikelsen, inte variansen, som andra parameter. Ungefär två tredjedelar av värdena ligger inom en standardavvikelse från medelvärdet, och ungefär nittiofem procent inom två.



Allt räknas genom att standardisera: dra bort väntevärdet och dela med standardavvikelsen. Då får du ett värde på standardnormalfördelningen, och sannolikheten får du från Φ, eller direkt med räknarens kommando för normalfördelningen.

En vanlig tentauppgift går åt andra hållet: uppgiften ger två andelar, till exempel att tio procent väger under ett visst värde och fem procent över ett annat, och du ska bestämma μ och σ. Du gör det i två steg. Först översätter du varje andel till ett standardiserat värde med räknarens inversa normalfördelning. Då får du två ekvationer av formen värdet minus μ delat med σ är lika med det standardiserade värdet. Sedan subtraherar du ekvationerna, så att μ försvinner, och löser ut σ. Därefter får du μ ur någon av ekvationerna.

Räknarens inversa normalfördelning vill ha arean till vänster. Om uppgiften säger “fem procent väger mer än” är arean till vänster nittiofem procent. Det är den vanligaste felkällan i den här uppgiften.

Exempel 7 — normalfördelning ur två andelar

Tio procent av alla förpackningar väger under $4{,}8$ kg och fem procent väger över $5{,}6$ kg. Vikten är normalfördelad. Bestäm μ och σ.

Lösning

Arean till vänster är $0{,}10$ respektive $0{,}95$, så de standardiserade värdena är $z_1 = -1{,}2816$ och $z_2 = 1{,}6449$: $\dfrac{4{,}8-\mu}{\sigma} = -1{,}2816$, $\dfrac{5{,}6-\mu}{\sigma} = 1{,}6449$ Subtrahera: $\sigma = \dfrac{5{,}6-4{,}8}{1{,}6449+1{,}2816} \approx 0{,}273$. Sedan $\mu = 5{,}6 - 1{,}6449 \cdot 0{,}273 \approx 5{,}15$. Svar: $\boxed{\mu \approx 5{,}15,\ \sigma \approx 0{,}273}$.

12. Väntevärde och varians av en funktion

Många uppgifter gäller en storhet som beror på en slumpvariabel, till exempel en vinst som är tre kronor per såld enhet minus en fast kostnad. Väntevärdet följer med linjära funktioner: väntevärdet av $a\xi + b$ är a gånger väntevärdet plus b. Standardavvikelsen påverkas bara av faktorn, inte av tillägget, eftersom att flytta alla värden lika mycket inte ändrar spridningen.

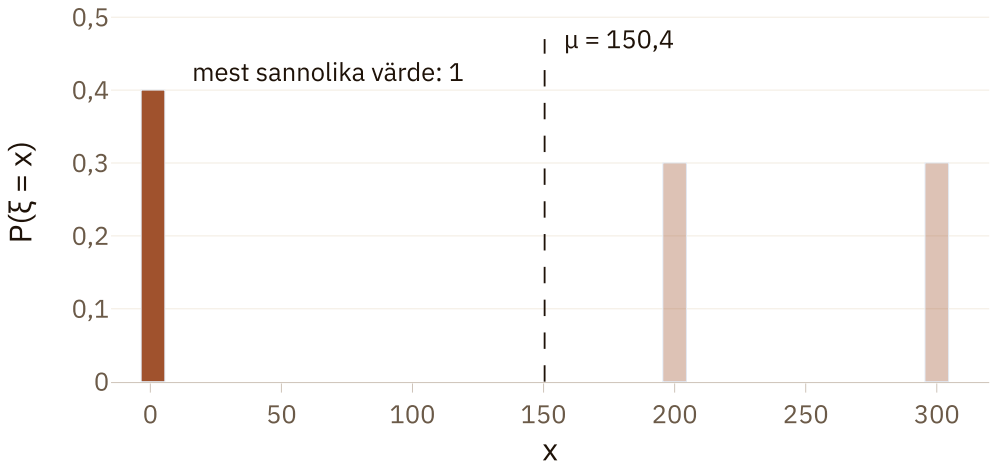
$$E(a\xi + b) = aE(\xi) + b, \qquad D(a\xi + b) = |a|\, D(\xi)$$

För en summa av två variabler adderas variansen, men om variablerna är beroende tillkommer en term med kovariansen:

$$V(a\xi + b\eta) = a^2 V(\xi) + b^2 V(\eta) + 2ab\,\mathrm{Cov}(\xi, \eta)$$

Är variablerna oberoende är kovariansen noll och den sista termen försvinner.

Om funktionen inte är linjär, till exempel en energi som beror på hastigheten i kvadrat, fungerar inte reglerna ovan. Då behöver du väntevärdet av kvadraten, och det får du ur variansen: väntevärdet av ξ^2 är variansen plus väntevärdet i kvadrat. För en diskret variabel som ges i en tabell gör du en ny tabell för den nya storheten och räknar väntevärdet direkt.



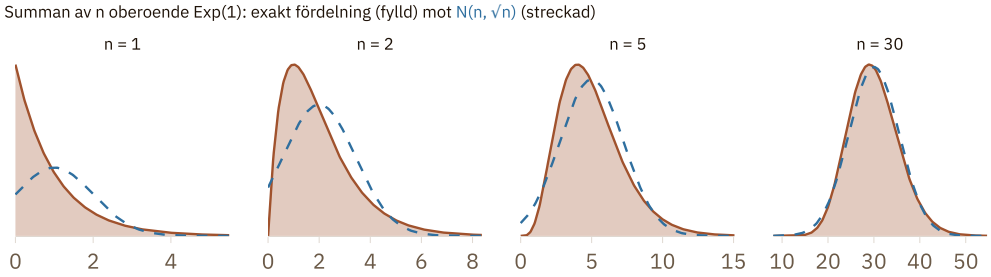
> ⚠ **Fällor med väntevärde och varians**
>
> Variansen kan aldrig vara negativ. Om uppgiften frågar efter standardavvikelsen, glöm inte att dra roten. Och räkna varianser, inte standardavvikelser, när du adderar.

13. Uppgift 6 och 7: centrala gränsvärdessatsen

Uppgift sex och sju handlar nästan alltid om centrala gränsvärdessatsen. Satsen säger att en summa av många oberoende och likafördelade variabler är ungefär normalfördelad, oavsett vilken fördelning varje enskild variabel har. Detsamma gäller för medelvärdet av dem.

Arbetsgången har tre steg. Först tar du fram väntevärdet μ och standardavvikelsen σ för en enda variabel, ofta från formelbladet. Sedan skriver du upp summans eller medelvärdets fördelning. Summans väntevärde är $n\mu$ och dess standardavvikelse är $\sigma\sqrt{n}$. Medelvärdets väntevärde är μ och dess standardavvikelse är $\sigma/\sqrt{n}$. Lägg märke till att spridningen växer med roten ur n, inte med n. Till sist räknar du sannolikheten med normalfördelningen.

$$\xi_1 + \cdots + \xi_n \approx N(n\mu,\ \sigma\sqrt{n}), \qquad \bar{\xi} \approx N(\mu,\ \sigma/\sqrt{n})$$



En vanlig variant frågar efter ett gränsvärde i stället för en sannolikhet, till exempel: hur länge räcker ett lager med nittiofem procents sannolikhet? Då söker du det värde T som summan överstiger med sannolikheten nittiofem procent, och det ligger $1{,}6449$ standardavvikelser under

summans väntevärde.

Exempel 8 — hur länge räcker lagret? (tentan 2025-03-17, uppgift 4)

Livslängden för ett elektronrör är exponentialfördelad med väntevärdet 220 timmar. Lagret innehåller 45 rör, och ett trasigt rör byts genast. Bestäm en tid T som lagret räcker åtminstone med 95 procents sannolikhet. Svara med en decimal.

Lösning

Låt ξ_i vara livslängden för rör nummer i. För exponentialfördelningen är $\mu = \sigma = 220$. Den totala tiden är summan av 45 livslängder: $\eta \approx N(45 \cdot 220,\ 220\sqrt{45}) = N(9900;\ 1475{,}8)$ Vi söker T med $P(\eta \geq T) = 0{,}95$, alltså $T = 9900 - 1{,}6449 \cdot 1475{,}8 \approx \boxed{7472{,}5\ \text{h}}$.

Fällor med centrala gränsvärdessatsen

Använd μ och σ för en enda variabel, inte för summan, när du börjar. Facit använder ingen halvkorrektion om uppgiften inte säger det. Och om variablerna är Poissonfördelade går det ofta att räkna exakt i stället, eftersom summan då själv är Poissonfördelad.

14. Från sannolikhetslära till statistik

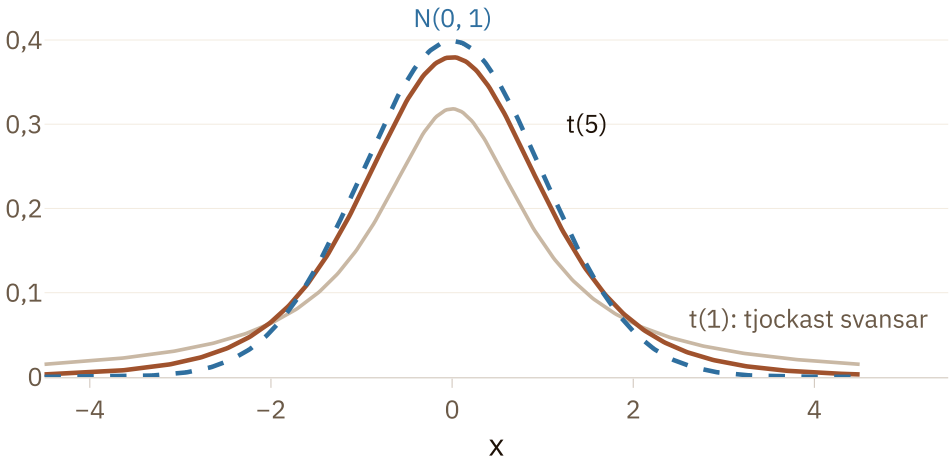
Fram till nu har vi utgått från att vi känner fördelningen och frågat hur troliga olika utfall är. Statistisk inferens vänder på frågan. Nu har vi observerat ett stickprov, och vi vill dra slutsatser om den okända fördelningen som stickprovet kommer ifrån, oftast om dess väntevärde.

Stickprovets medelvärde $\bar{x}$ är den naturliga skattningen av väntevärdet μ, och stickprovets standardavvikelse s är skattningen av σ. Medelvärdet är självt en slumpvariabel: tar man ett nytt stickprov blir det ett annat medelvärde. Dess spridning är $\sigma/\sqrt{n}$. Det är därför större stickprov ger säkrare slutsatser: spridningen hos medelvärdet minskar med roten ur stickprovets storlek. För att halvera osäkerheten behöver du fyra gånger så många observationer.



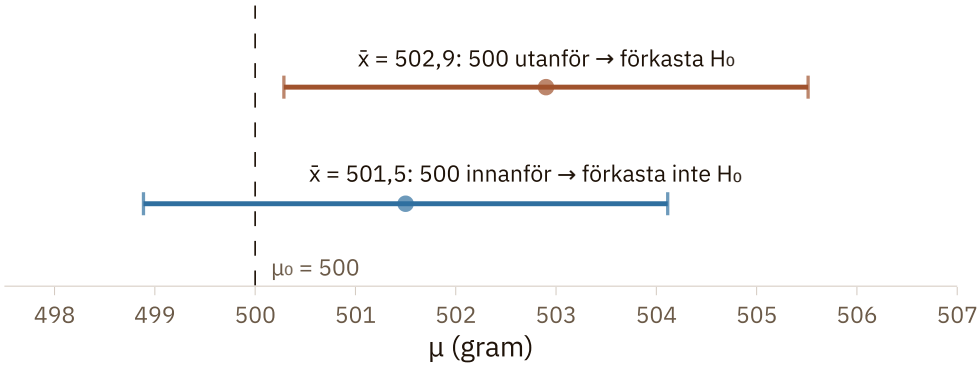
inferens: slutsats om μ och σ, med ett mått på osäkerheten

När σ är känd standardiserar man med den, och då gäller normalfördelningen. När σ är okänd, vilket är det vanliga, skattar man den med s. Det tillför extra osäkerhet, och därför används t-fördelningen i stället. Den liknar normalfördelningen men har tjockare svansar, och tjockleken beror på antalet frihetsgrader. För en mätserie är antalet frihetsgrader $n - 1$. Med många observationer blir t-fördelningen nästan identisk med normalfördelningen.



Vid hypotesprövning kan man göra två sorters fel. Det första är att förkasta nollhypotesen fast den är sann. Sannolikheten för det felet är signifikansnivån α, och den väljer man själv, ofta fem procent. Det andra felet är att inte förkasta nollhypotesen fast den är falsk. Sannolikheten för det är β, och styrkan är ett minus β. Man kan inte minska båda felen samtidigt utan att öka stickprovets storlek. Sänker man signifikansnivån flyttas den kritiska gränsen längre ut, och då blir det svårare att upptäcka en verklig skillnad.

Konfidensintervall och hypotesprövning hänger ihop. Ett tvåsidigt test på nivån fem procent förkastar nollhypotesen precis när värdet μ_0 ligger utanför det nittiofemprocentiga konfidensintervallet. Det är samma räkning sedd från två håll.

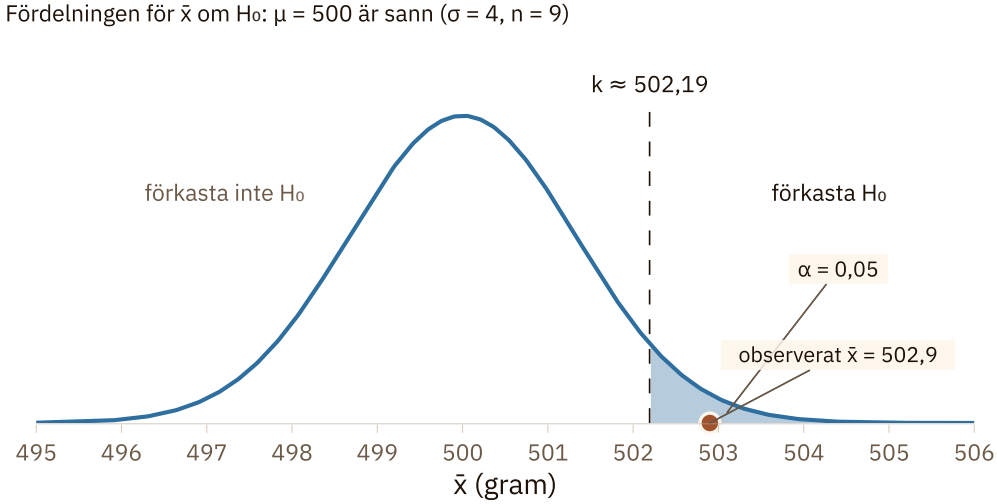


15. Uppgift 8: hypotesprövning, kritisk gräns och P-värde

Från uppgift åtta handlar tentan om statistisk inferens: att dra slutsatser om en okänd parameter utifrån ett stickprov. Hypotesprövning börjar med två hypoteser. Nollhypotesen H_0 är det man utgår från, oftast att ingenting har ändrats. Mothypotesen H_1 är det man vill visa. Tumregeln är att det man vill visa ska stå i H_1.

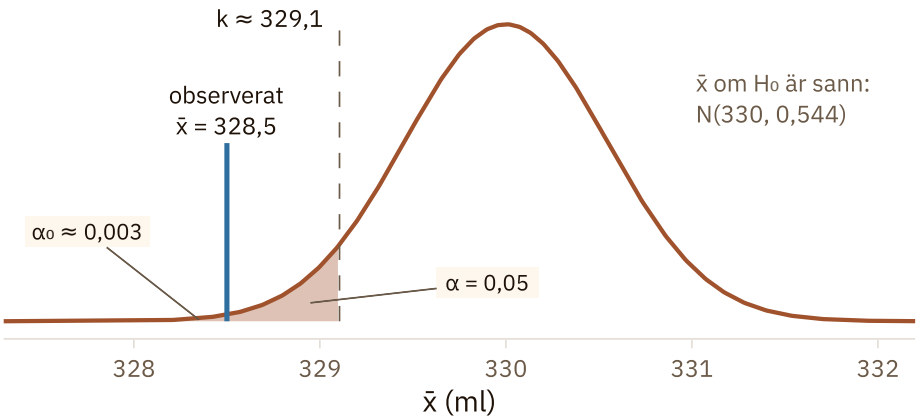
Testet bygger på en kritisk gräns k. Om medelvärdet i stickprovet hamnar bortom gränsen förkastar man nollhypotesen. Gränsen väljs så att risken att förkasta en sann nollhypotes, signifikansnivån α, blir till exempel fem procent. När standardavvikelsen är känd ligger gränsen λ_α standardavvikelser för medelvärdet ovanför μ_0, och medelvärdets standardavvikelse är $\sigma/\sqrt{n}$.

$$k = \mu_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}$$



Om standardavvikelsen är okänd skattas den med stickprovets standardavvikelse s, och då används t-fördelningen med $n - 1$ frihetsgrader i stället för normalfördelningen. På räknaren får du gränsen med den inversa t-fördelningen.

Ett annat sätt att testa är direktmetoden. Där räknar du P-värdet, som skrivs α_0: sannolikheten att få ett resultat som är minst lika extremt som det observerade, om nollhypotesen är sann. Är P-värdet mindre än signifikansnivån förkastar du nollhypotesen. Vid ett tvåsidigt test, där mothypotesen bara säger att parametern är skild från μ_0, ska P-värdet multipliceras med två.



Exempel 9 — kritisk gräns

Ett stickprov med 16 observationer tas från $N(\mu;\ 2)$. Man testar $H_0 : \mu = 10$ mot $H_1 : \mu > 10$ på nivån fem procent. Var ligger den kritiska gränsen?

Lösning

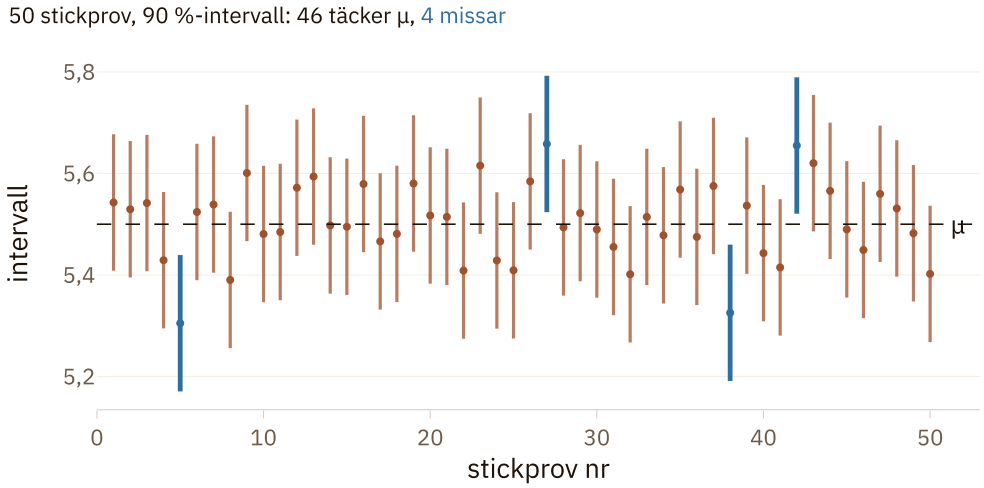
Medelvärdets standardavvikelse är $2/\sqrt{16} = 0{,}5$, och $\lambda_{0{,}05} = 1{,}6449$: $k = 10 + 1{,}6449 \cdot 0{,}5 \approx \boxed{10{,}82}$ Nollhypotesen förkastas om medelvärdet i stickprovet är större än $10{,}82$.

Fällor i uppgift 8

Ett vänstersidigt test, där mothypotesen säger att μ är mindre än μ_0, ger en gräns under μ_0. Räknarens inversa funktioner vill ha arean till vänster: för ett högersidigt test på fem procent är det nittiofem procent.

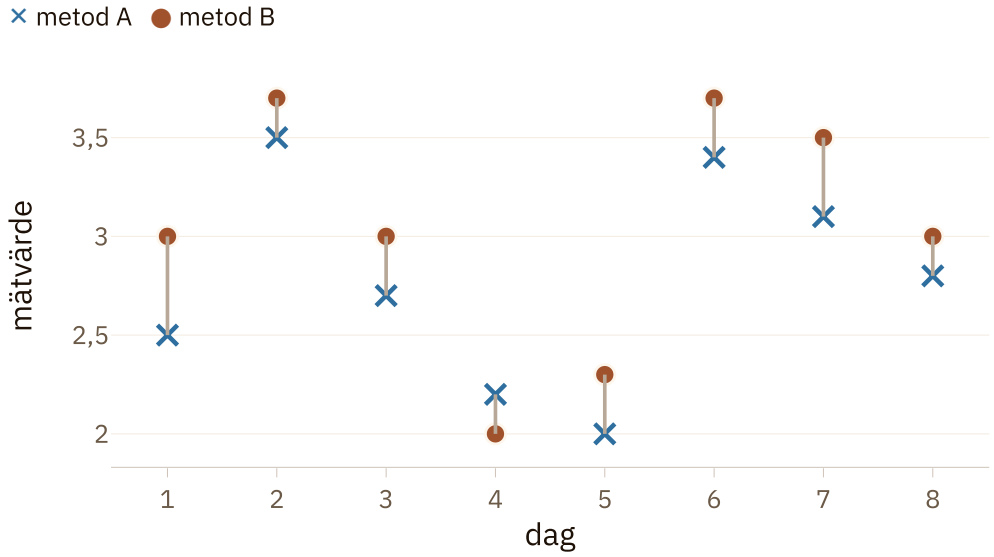
16. Uppgift 9: konfidensintervall

Ett konfidensintervall är ett intervall som med en viss sannolikhet, konfidensgraden, täcker den okända parametern. Ett nittiofemprocentigt intervall för ett väntevärde är medelvärdet plus minus ett t-värde gånger stickprovets standardavvikelse delat med roten ur n. För nittiofem procent används $t_{0{,}025}$, eftersom de fem procenten delas lika på båda sidorna.



Den viktigaste frågan i uppgift nio är om det är stickprov i par eller två oberoende stickprov. Stickprov i par betyder att mätningarna hör ihop två och två: samma bil mätt före och efter, samma dag på två platser, samma person med två metoder. Då bildar du differenserna för varje par, och sedan är det ett vanligt intervall för en enda mätserie med $n - 1$ frihetsgrader, där n är antalet par.

$$I_\Delta = \bar{z} \mp t_{\alpha/2}(n-1)\frac{s_z}{\sqrt{n}}$$



Två oberoende stickprov betyder att grupperna inte hör ihop, till exempel två laboratorier som mäter flera gånger var. Då används en gemensam standardavvikelse, som står på formelbladet, och $n_1 + n_2 - 2$ frihetsgrader. En klassisk fälla är två laboratorier som mäter samma prov flera gånger: det är ändå två oberoende stickprov, eftersom mätning nummer ett i det ena labbet inte hör ihop med mätning nummer ett i det andra.

Om noll ligger i ett intervall för en differens kan nollhypotesen att det inte finns någon skillnad inte förkastas. Ligger noll utanför kan den förkastas.

Exempel 10 — stickprov i par

Sex bilar mäts före och efter en justering. Differenserna har medelvärdet $1{,}2$ och standardavvikelsen $0{,}8$. Bestäm ett nittiofemprocentigt konfidensintervall för den genomsnittliga förändringen.

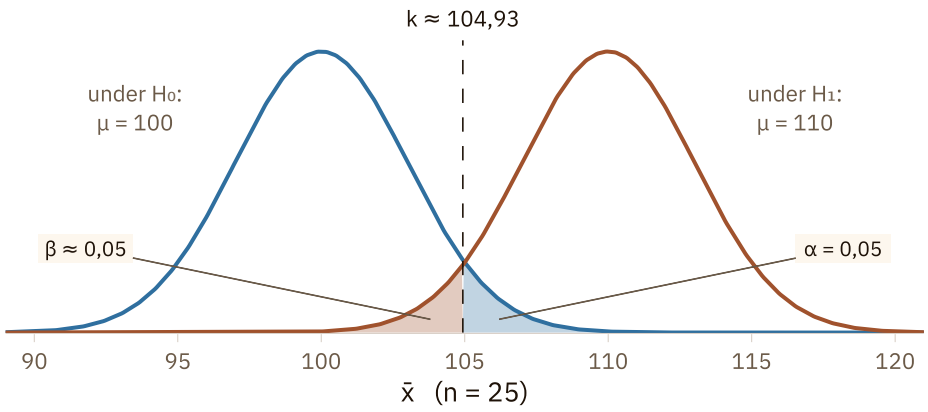
Lösning

Det är stickprov i par med sex par, så $f = 5$ och $t_{0{,}025}(5) = 2{,}5706$: $1{,}2 \mp 2{,}5706 \cdot \frac{0{,}8}{\sqrt{6}} = 1{,}2 \mp 0{,}84$ Intervallet blir $\boxed{[0{,}36;\ 2{,}04]}$. Noll ligger utanför, så förändringen är signifikant.

17. Uppgift 10: styrka, stickprovsstorlek och teckenintervall

Uppgift tio varierar mer än de andra. De tre vanligaste typerna är styrka, stickprovsstorlek och teckenintervall.

Styrkan är sannolikheten att testet förkastar nollhypotesen när den är falsk, alltså när det sanna väntevärdet är μ_1. Arbetsgången har två steg. Först bestämmer du den kritiska gränsen under nollhypotesen, som i uppgift åtta. Sedan räknar du sannolikheten att hamna bortom gränsen, men nu med det sanna väntevärdet μ_1. För ett högersidigt test är det ett minus Φ av gränsen minus μ_1, delat med medelvärdets standardavvikelse. För ett vänstersidigt test är det Φ av samma uttryck. Styrkan ska alltid ligga mellan signifikansnivån och ett.



Uppgiften kan också gå baklänges: du får en önskad styrka och ska bestämma signifikansnivån. Då lägger du gränsen så att den önskade styrkan uppnås under μ_1, och räknar sedan hur stor del av fördelningen under μ_0 som ligger bortom gränsen.

Exempel 11 — styrka

Fortsätt exempel 9, där gränsen är $10{,}82$ och medelvärdets standardavvikelse $0{,}5$. Vad är styrkan om det sanna väntevärdet är 11?

Lösning

$$S(11) = 1 - \Phi\left(\frac{10{,}82 - 11}{0{,}5}\right) = 1 - \Phi(-0{,}355) \approx \boxed{0{,}64}$$

Exempel 12 — styrka baklänges (tentan 2025-03-17, uppgift 8)

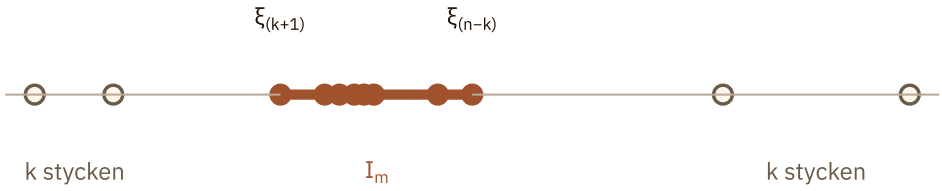
Ett stickprov med 25 observationer tas från $N(\mu;\ 4{,}15)$. Man testar $H_0 : \mu = 10$ mot $H_1 : \mu = 12{,}5$ och vill att styrkan ska vara 90 procent. Vad blir signifikansnivån? Svara i procent med två decimaler.

Lösning

Medelvärdets standardavvikelse är $4{,}15/5 = 0{,}83$. Styrkan 90 procent under $\mu_1 = 12{,}5$ ger gränsen $k = 12{,}5 - 1{,}2816 \cdot 0{,}83 \approx 11{,}436$ Signifikansnivån är sannolikheten att hamna över gränsen när $\mu = 10$: $\alpha = 1 - \Phi\left(\frac{11{,}436 - 10}{0{,}83}\right) \approx \boxed{4{,}18\ \%}$

Stickprovsstorleken bestäms antingen av hur brett ett intervall får vara eller av vilken styrka testet ska ha. Formlerna löser ut n ur uttrycket för bredden eller för styrkan, och svaret avrundas alltid uppåt, eftersom ett mindre stickprov inte räcker.

Teckenintervallet används när man vill skatta medianen och inte kan anta normalfördelning. Idén är enkel: varje observation hamnar under medianen med sannolikheten en halv, så antalet observationer under medianen är binomialfördelat med p lika med en halv. Om du sorterar observationerna och tar intervallet från den $(k+1)$:a minsta till den $(k+1)$:a största, är konfidensgraden ett minus två gånger sannolikheten att högst k observationer hamnar under medianen.



Exempel 13 — konfidensgrad för ett teckenintervall (tentan 2025-03-17, uppgift 7)

Ett stickprov på åtta observationer kommer från en fördelning som inte är normal. Beräkna konfidensgraden för intervallet från den näst minsta till den näst största observationen. Svara i procent med två decimaler.

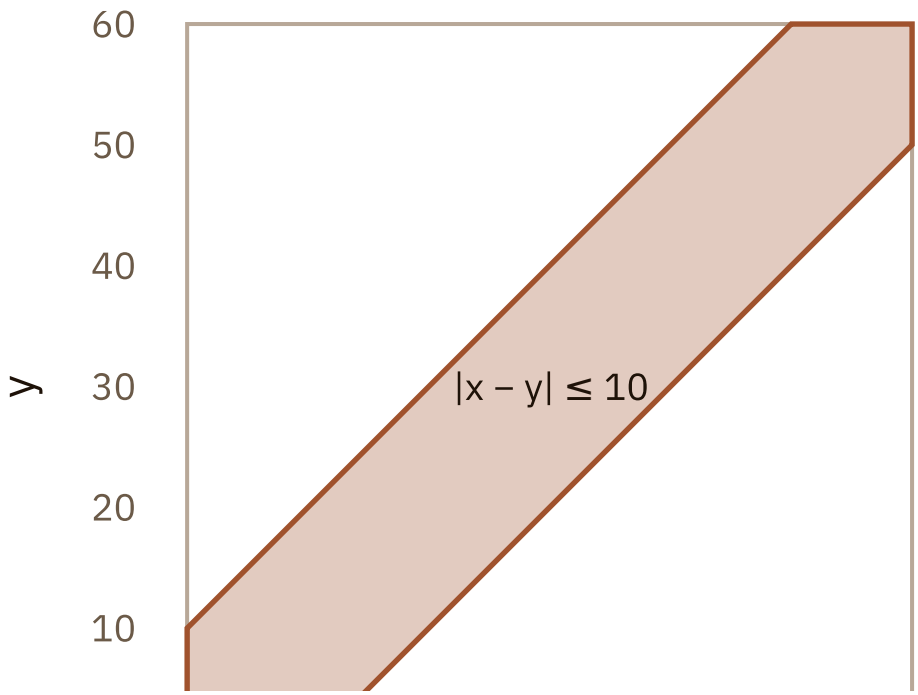
Lösning

Låt η vara antalet observationer under medianen, $\eta \in \mathrm{Bin}(8;\ 0{,}5)$. Intervallet missar medianen om högst en observation hamnar på ena sidan: $1 - 2P(\eta \leq 1) = 1 - 2 \cdot \frac{1+8}{256} = \frac{238}{256} \approx \boxed{92{,}97\ \%}$

18. Uppgift 11 och 12: två slumpvariabler

De två sista uppgifterna handlar om två slumpvariabler samtidigt. Fyra typer återkommer: area, kovarians, tvådimensionell normalfördelning och en blandning av simultana fördelningar, slumpvandring och max eller min.

Areauppgifterna är ofta de mest tacksamma. Om ett par (ξ, η) är likformigt fördelat över ett område är sannolikheten för en händelse arean av händelsens del av området delat med hela områdets area. Rita alltid. Ett klassiskt exempel är två personer som kommer slumpmässigt under en timme och väntar en viss tid på varandra: de träffas om skillnaden mellan ankomsttiderna är mindre än väntetiden, och det är ett band längs diagonalen i en kvadrat. Det enklaste är att räkna arean av de två trianglarna utanför bandet och dra bort den från ett.



Exempel 14 — möte inom tio minuter

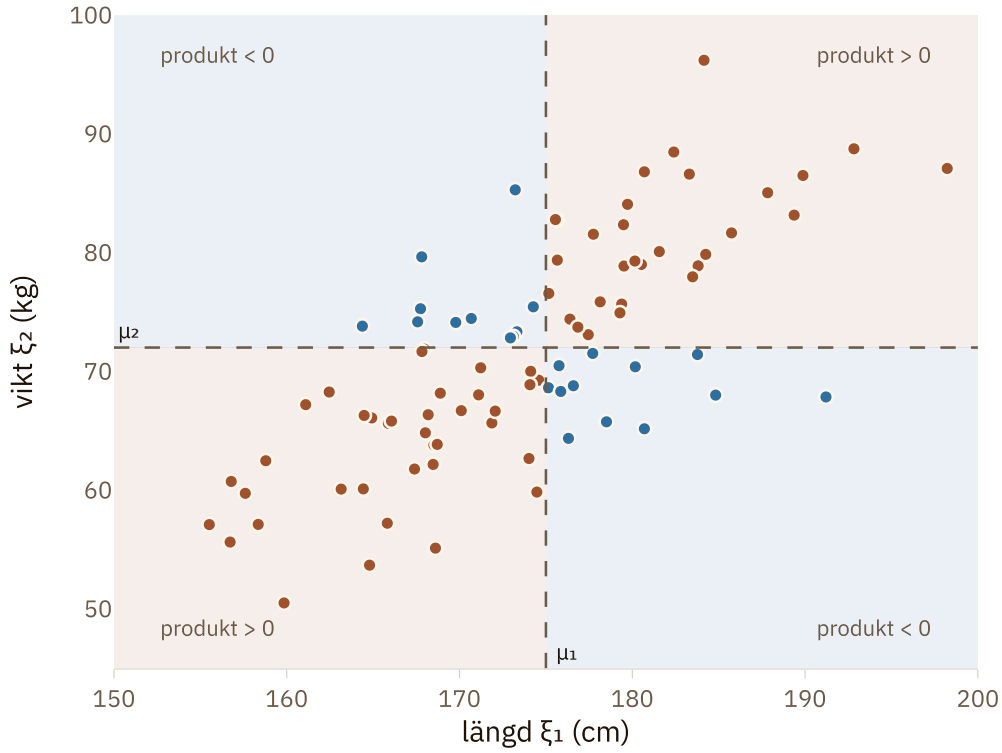
Två personer kommer oberoende av varandra vid en slumpmässig tidpunkt under samma timme, och var och en väntar tio minuter. Vad är sannolikheten att de träffas?

Lösning

Ankomsttiderna är likformigt fördelade på kvadraten $[0, 60] \times [0, 60]$. De träffas om skillnaden är högst tio minuter. Utanför bandet ligger två trianglar med kateten 50: $P = 1 - \left(\frac{50}{60}\right)^2 = 1 - \frac{25}{36} = \frac{11}{36} \approx \boxed{30{,}6\ \%}$

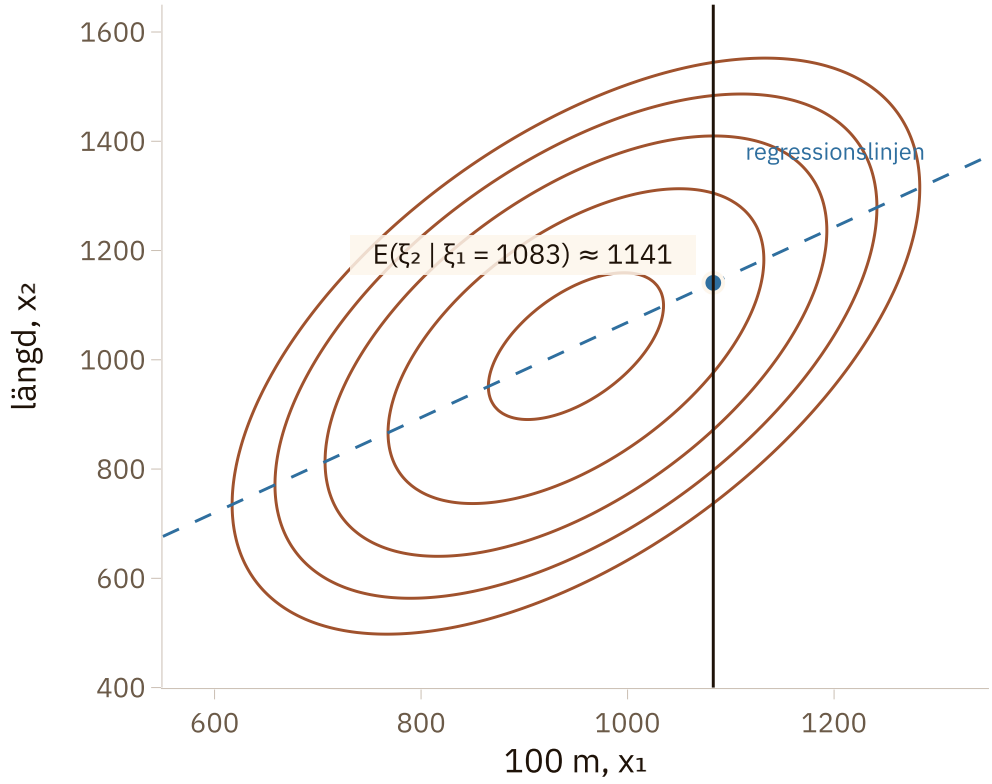
Kovariansen mäter hur två variabler samvarierar. Den är positiv när stora värden på den ena brukar gå ihop med stora värden på den andra. Korrelationskoefficienten ρ är kovariansen delad med produkten av standardavvikelserna, och den ligger alltid mellan minus ett och ett. Oberoende variabler har kovariansen noll. Kovariansen kan delas upp över summor, och en faktor framför en variabel kommer med som en faktor framför kovariansen.

$$\mathrm{Cov}(\xi,\eta) = E(\xi\eta) - E(\xi)E(\eta) = \rho\, D(\xi)\, D(\eta)$$



Den tvådimensionella normalfördelningen dyker upp i uppgifter som gäller två mått som hör ihop, till exempel längden hos far och son, eller två idrottares resultat. Tricket är nästan alltid att bilda skillnaden eller summan mellan dem. Skillnaden mellan två normalfördelade variabler är normalfördelad, med väntevärdet lika med skillnaden mellan väntevärdena. Variansen är summan av varianserna minus två gånger kovariansen. Frågan “vem blir längst?” blir då frågan om skillnaden är större än noll.

$$D(\xi_1 - \xi_2) = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\sigma_2}$$



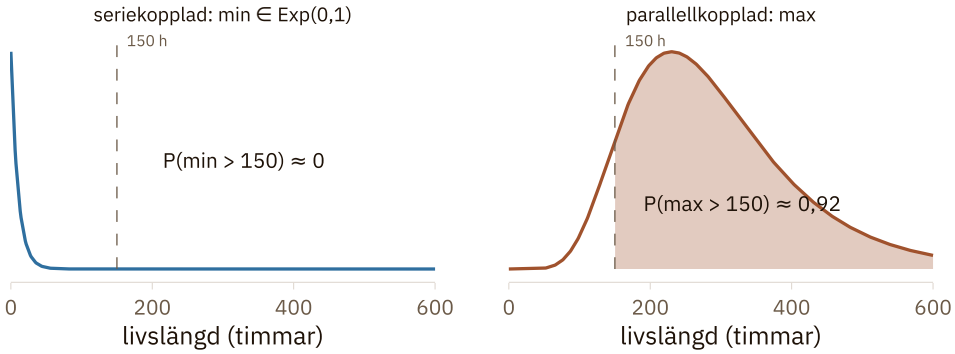
Exempel 15 — blir sonen längre än fadern?

Faderns längd är $N(178;\ 7)$ och sonens $N(180;\ 7)$, med korrelationen $\rho = 0{,}5$. Vad är sannolikheten att fadern är längre än sonen?

Lösning

Låt ζ vara faderns längd minus sonens. Då är $E(\zeta) = -2$ och $D(\zeta) = \sqrt{49 + 49 - 2 \cdot 0{,}5 \cdot 7 \cdot 7} = 7$ Sannolikheten är $P(\zeta > 0) = 1 - \Phi(2/7) \approx \boxed{0{,}39}$.

Max och min dyker upp i uppgifter om system av komponenter. Ett seriekopplat system fungerar bara så länge alla komponenter fungerar, så dess livslängd är den kortaste komponentens livslängd, alltså minimum. Sannolikheten att systemet håller längre än en viss tid är produkten av sannolikheterna att varje komponent gör det. Ett parallellkopplat system fungerar så länge minst en komponent fungerar, så dess livslängd är maximum, och sannolikheten att det har gått sönder före en viss tid är produkten av komponenternas fördelningsfunktioner.



En slumpvandring är ett antal steg åt höger eller vänster. Om ξ är antalet steg åt höger av n är ξ binomialfördelad, och positionen efter n steg är $2\xi - n$. En fråga om positionen översätts därmed till en fråga om ξ, som räknaren klarar.

Exempel 16 — slumpvandring

En partikel tar tio steg, var och ett åt höger eller vänster med samma sannolikhet. Vad är sannolikheten att den hamnar minst fyra steg till höger om startpunkten?

Lösning

Låt ξ vara antalet steg åt höger, $\xi \in \mathrm{Bin}(10;\ 0{,}5)$. Positionen är $2\xi - 10 \geq 4$, alltså $\xi \geq 7$:

$$P(\xi \geq 7) = 1 - P(\xi \leq 6) = \frac{176}{1024} \approx \boxed{0{,}172}$$

19. De vanligaste felen, samlade

Den här listan samlar de fel som oftast kostar poäng på del 1. Lyssna på den kvällen före tentan.

Det första felet gäller “minst”. Sannolikheten för minst k är ett minus sannolikheten för högst $k - 1$, inte högst k. Det gäller binomial-, Poisson- och ffg-fördelningen och allt annat som räknar heltal.

Det andra felet gäller arean till vänster. Räknarens inversa funktioner vill alltid ha arean till vänster om värdet. Om uppgiften säger “de fem procent tyngsta” eller “fem procent över” är arean till vänster nittiofem procent.

Det tredje felet gäller standardavvikelse och varians. Kursen skriver normalfördelningen med standardavvikelsen, inte variansen. När du adderar spridningar adderar du varianser och drar roten på slutet, aldrig standardavvikelserna direkt.

Det fjärde felet gäller tidsenheten. En intensitet per timme måste räknas om om frågan gäller minuter, och en exponentialfördelnings λ ska ha samma tidsenhet som frågan.

Det femte felet gäller par eller två stickprov. Stickprov i par kräver att mätningarna hör ihop två och två. Två laboratorier eller två grupper är två oberoende stickprov, med en gemensam standardavvikelse och $n_1 + n_2 - 2$ frihetsgrader.

Det sjätte felet gäller ensidigt och tvåsidigt. Vid ett tvåsidigt test delas signifikansnivån på båda sidorna, och P-värdet multipliceras med två. Vid ett konfidensintervall används alltid $\alpha/2$ i varje svans.

Det sjunde felet gäller avrundning. Avrunda bara i det sista steget, och svara med exakt det antal decimaler som efterfrågas. En stickprovsstorlek avrundas alltid uppåt.

Det åttonde felet gäller slumpvariabeln. Definiera den i ord, med vad den räknar eller mäter och i vilken enhet. Det är den vanligaste orsaken till uteblivna delpoäng.

20. Strategi på tentan

Läs igenom alla tolv uppgifter först och börja med dem du känner igen direkt. Uppgift ett, två och sex brukar vara de säkraste poängen. Kartan i början visar var de brukar ligga.

För varje uppgift: definiera slumpvariabeln i ord, ange fördelningen, skriv vad som ska beräknas, räkna och svara exakt som uppgiften ber om. Kontrollera rimligheten innan du går vidare. En sannolikhet ligger mellan noll och ett. En varians är aldrig negativ. Ett antal är ett heltal mellan noll och n. Medianen i en exponentialfördelning är mindre än väntevärdet. Styrkan ligger mellan signifikansnivån och ett.

Om du kör fast, skriv ändå upp slumpvariabeln och fördelningen. Det kan ge delpoäng, och ofta lossnar uppgiften när uppställningen står på papperet.

Se även

- S0008M
- Betingad sannolikhet · Bayes sats · Lagen om total sannolikhet
- Binomialfördelning · Hypergeometrisk fördelning · Poissonfördelning · Exponentialfördelning · ffg-fördelning · Likformig fördelning
- Normalfördelning · Median och kvantiler · Linjära funktioner av stokastiska variabler
- Centrala gränsvärdessatsen · Normalapproximation
- Hypotesprövning · Direktmetoden · Konfidensintervall · Stickprov i par · Två stickprov
- Signifikansnivå och styrka · Styrkefunktion · Teckenintervall
- Flerdimensionella fördelningar · Kovarians · Korrelation · Tvådimensionell normalfördelning · Max och min av stokastiska variabler