

Hypergeometrisk fördelning

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #sannolikhetslära #fördelning

Kurs: S0008M Förkunskaper: Binomialfördelning

Den hypergeometrisk 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, $\xi \in \text{Hyp}(N, n, p)$, 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 $\text{Bin}(n, p)$ här — samma räkning som **dragning ur en urna**.

2. Väntevärde och varians

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

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 $\text{Bin}(n, K/N)$ – 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.

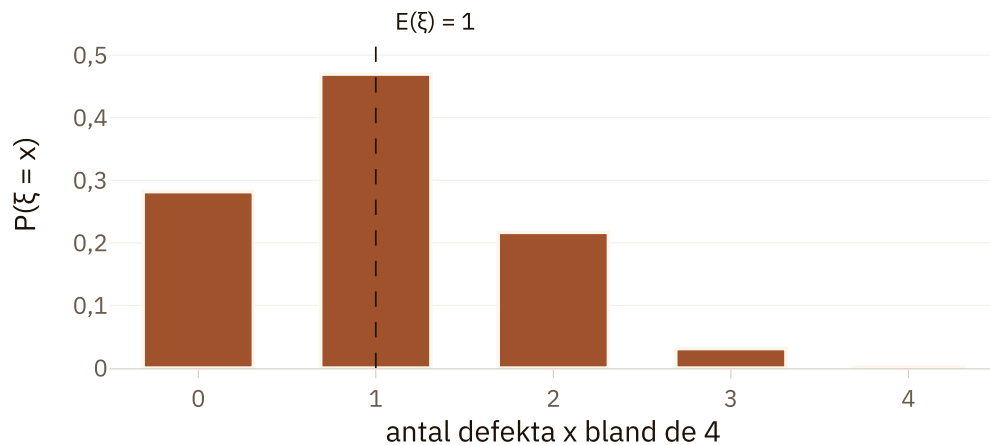
$$P(\xi = 1) = \frac{\binom{5}{1} \binom{15}{3}}{\binom{20}{4}} = \frac{5 \cdot 455}{4845} = \boxed{\frac{455}{969}} \approx 0,470$$

Rimlighetskontroll. 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.

Hyp(20, 4, 0,25): 4 dras ur 20, varav 5 defekta



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.** $V(\xi) = \frac{N-n}{N-1} np(1-p)$; 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 $p = 10/25 = 0,4$:

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

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: $\xi \in \text{Hyp}(3000, 100, 0,05)$.

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

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

Tröskeln. Vi söker minsta k med

$$P(\xi > k) < 0,05 \iff P(\xi \leq k) > 0,95$$

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

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
-
-