

Formelsamling

Diskreta fördelningar

Fördelning	Sannolikhetsfunktion	Väntevärde	Varians	Användning
Bernoulli $Ber(p)$	$p_X(k) = \begin{cases} p & , k = 1 \\ 1 - p & , k = 0 \end{cases}$	p	$p(1 - p)$	Ett försök som antingen lyckas ($k = 1$) eller misslyckas ($k = 0$)
Binomial $Bin(n, p)$	$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k},$ $k = 0, 1, \dots, n$	np	$np(1 - p)$	Antalet lyckade, då n stycken likadana (oberoende) försök utförs
Hypergeometrisk $Hyp(N, n, p)$	$p_X(k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}},$ $0 \leq k \leq Np,$ $0 \leq n - k \leq N(1 - p)$	np	$np(1 - p) \frac{N-n}{N-1}$	Vi drar n stycken utan återläggning från en ändlig population av storlek N , där Np stycken har en viss egenskap och räknar antalet med denna egenskap som finns bland de dragna
Likformig $U(N)$	$p_X(k) = \frac{1}{N}, \quad 1 \leq k \leq N$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$	
för-första-gången $ffg(p)$	$p_X(k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Antalet försök t.o.m. det första lyckade, alla försök är lika och har samma slh p att lyckas
Poisson $Po(\lambda)$	$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$	λ	λ	Används när något sker med en viss frekvens under ett givet intervall. T.ex. tryckfel per sida

Kontinuerliga fördelningar

Fördelning	Täthetsfunktionsfunktion	Väntevärde	Varians	Användning
Likformig $U(a, b)$	$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Exponential $Exp(\beta)$	$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \quad x \geq 0$	β	β^2	Livslängder
Normal $N(\mu, \sigma)$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	μ	σ^2	Längd, vikt, o.s.v.

För $X \sim N(E(X) = \mu, D(X) = \sigma)$ gäller

$$\mathbb{P}(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

där $\Phi(\cdot)$ är fördelningsfunktionen för standard normal fördelning (d.v.s. $\mu = 0$ och $\sigma = 1$).
 $\Phi(-x) = 1 - \Phi(x)$.

Räkneregler

För godtyckliga mängder $A, B, C \subset \Omega$ gäller

- $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup \Omega = \Omega$, $A \cap \Omega = A$
- $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Punktskattning

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\ s_{xy}^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right) \\ s_p^2 &= \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}\end{aligned}$$

Korrelation och regressionslinjen $y = a + bx$:

$$\hat{\rho} = r = \frac{s_{xy}^2}{s_x s_y} \quad \text{och} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}, \quad \hat{b} = \frac{s_{xy}^2}{s_x^2}$$

Centrala Gränsvärdessatsen

Låt $X_1, X_2, X_3, \dots$ vara en följd av oberoende slumpvariabler med samma sannolikhetsfördelning där väntevärdet är $\mathbb{E}(X) = \mu$ och $D(X) = \sigma$. Då gäller

$$\mathbb{P} \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x \right) \longrightarrow \Phi(x) \quad \text{då } n \longrightarrow \infty.$$

Approximationer

$$X \sim \text{Bin}(n, p) \approx \begin{cases} N(np, \sqrt{np(1-p)}) & , \text{ om } np(1-p) > 10 \\ Po(np) & , \text{ om } p < 0.1, np(1-p) \approx np \end{cases}$$

$$X \sim \text{Hyp}(N, n, p) \approx N(np, \sqrt{np(1-p)\frac{N-n}{N-1}}) \quad , \text{ om } np(1-p)\frac{N-n}{N-1} > 10$$

Konfidensintervall

Låt $(x_1, \dots, x_n)$ vara ett stickprov av tillräckligt stort stickprovstorlek n och därmed utfall för en (approx.) $N(\mu, \sigma)$ -fördelad stokastisk variabel.

- $100(1 - \alpha)\%$ **konfidensintervall för andelar p**

- i en oändlig population

$$I_p = \hat{p} \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

- i en ändlig population

$$I_p = \hat{p} \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} \frac{N - n}{N - 1}}$$

- för andelsskillnader

- i en oändlig population

$$I_{p_2 - p_1} = \hat{p}_2 - \hat{p}_1 \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

- i en ändlig population

$$I_{p_2 - p_1} = \hat{p}_2 - \hat{p}_1 \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} \frac{N_1 - n_1}{N_1 - 1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2} \frac{N_2 - n_2}{N_2 - 1}}$$

- $100(1 - \alpha)\%$ **konfidensintervall för väntevärden μ**

- känd σ

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- okänd σ

$$I_\mu = \bar{x} \pm t(n - 1)_{\alpha/2} \frac{s}{\sqrt{n}}$$

- för väntevärdesskillnad

- samma känd varians

$$I_{\mu_x - \mu_y} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \sigma \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

- samma okänd varians

$$I_{\mu_x - \mu_y} = \bar{x} - \bar{y} \pm t(n_x + n_y - 2)_{\alpha/2} s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

- för stickprov-i-par $z_i = x_i - y_i$

$$I_\Delta = \bar{z} \pm t(n - 1)_{\alpha/2} \frac{s_z}{\sqrt{n}}$$