SVERIGES LANTBRUKSUNIVERSITET
Institutionen för energi och teknik
Uwe Menzel
e-post: uwe.menzel@matstat.de

Formelsamling
Grundläggande matematiskt statistik
2018–08–22

Sannolikhetslära

Klassisk sannolikhetsdefinition:

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

g: antalet utfall som är gynsamma för A ; m: antalet möjliga utfall

Komplementsats:

$$P(A^*) = 1 - P(A)$$

Additionssats:

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

Betingade sannolikhet:

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

Oberoende händelser:

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{eller} \quad P(A \mid B) = P(A)$$

Lagen om total sannolikhet:

$$P(R) = \sum_i P(R \mid M_i) \cdot P(M_i) \qquad \sum_i P(M_i) = 1 \quad \text{(partition)}$$

Bayes sats:

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

Diskreta slumpvariabler

Sannolikhetsfunktion:

$$P(X = k) = p_X(k)$$

Normering:

$$\sum_{k \in \Omega} p_X(k) = 1$$

Sannolikhet för händelse A:

$$P(A) = \sum_{k \in A} p_X(k)$$

Fördelningsfunktion:

$$F_X(a) = P(X \le a) = \sum_{k \le a} p_X(k)$$

Sannolikhet för intervall:

$$P(a < X \le b) = F_X(b) - F_X(a)$$

Sammanhang $p_X(k)$ och $F_X(k)$:

$$p_X(k) = F_X(k) - F_X(k-1)$$

Komplementära händelser:

$$P(X > a) = 1 - P(X \le a) = 1 - F_X(a)$$

Monotoni:

$F_X(x)$ monotont växande från 0 till 1 (trappsteg)

Kontinuerliga slumpvariabler

Täthetsfunktion:

$$f_X(x) \quad \text{reell funktion}$$

Normering:

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

Fördelningsfunktion:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$$

Kvantiler:

$$x_\alpha \text{ defineras genom:} \quad F_X(x_\alpha) = 1 - \alpha$$

Sannolikhet för intervall:

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(x)\,dx$$

Sammanhang $f_X(x)$ och $F_X(x)$:

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$$

Komplementära händelser:

$$P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$$

Monotoni:

$$F_X(x) \, \text{monotont växande från 0 till 1}$$

Väntevärden

Diskreta s.v.:

$$\mu = E(X) = \sum_{k \in \Omega} k \cdot p_X(k)$$

$$E(g(X)) = \sum_{k \in \Omega} g(k) \cdot p_X(k)$$

Kontinuerliga s.v:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)\,dx$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x)\,dx$$

Varians, standardavvikelse

Diskreta s.v.:

$$V(X) = E\left[(X - \mu)^2\right] = E\left(X^2\right) - \mu^2 = \sum_{k \in \Omega} k^2 \cdot p_X(k) - \mu^2$$

$$D(X) = \sqrt{V(X)}$$

Kontinuerliga s.v:

$$V(X) = E\left[(X - \mu)^2\right] = E\left(X^2\right) - \mu^2 = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x)\,dx - \mu^2$$

$$D(X) = \sqrt{V(X)}$$

Linjärtransformation av s.v.:

$$\begin{aligned} E\left(a \cdot X+b\right) &= a \cdot E(X)+b = a \cdot \mu + b \\ V\left(a \cdot X+b\right) &= a^2 \cdot V(X) = a^2 \cdot \sigma^2 \\ D\left(a \cdot X+b\right) &= |a| \cdot D(X) = |a| \cdot \sigma \end{aligned}$$

Standardisering av s.v.:

$$E(X)=\mu \quad D(X)=\sigma \quad \Longrightarrow \quad Y=\frac{X-\mu}{\sigma} \qquad \text{har väntevärde 0 och std.avv. 1}$$

Väntevärden, standardavvikelse för summa/differens av s.v.:

$$E\left(a \cdot X+b \cdot Y\right) = a \cdot E(X) + b \cdot E(Y)$$

$$V\left(a \cdot X+b \cdot Y\right) = a^2 \cdot V(X) + b^2 \cdot V(Y) + 2ab \cdot C(X,Y)$$

$$V\left(a \cdot X+b \cdot Y\right) = a^2 \cdot V(X) + b^2 \cdot V(Y) \quad \text{för oberoende } X, Y \text{ dvs. } C(X,Y)=0$$

$$V(X+Y) = V(X)+V(Y) \quad \text{om X, Y oberoende}$$

$$V(X-Y) = V(X)+V(Y) \quad \text{om X, Y oberoende}$$

$$E\left(\sum X_i\right) = \sum E\left(X_i\right) \quad \text{gäller allmänt}$$

$$V\left(\sum X_i\right) = \sum V\left(X_i\right) \quad \text{om } X_i \text{ oberoende}$$

Diskreta fördelningar:

Likformig diskret fördelning:

$$\Omega_X = \{a, a+1, \ldots, b\}$$

$$p_X(k) = 1/N \quad \text{där} \quad N = b-a+1$$

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a+2) \cdot (b-a)}{12}$$

Binomialfördelning:

$$X \sim Bin(n,p)$$

$$\Omega_X = \{0, 1, 2, \ldots, n\}$$

$$p_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$E(X) = n \cdot p$$

$$V(X) = n \cdot p \cdot (1-p)$$

Poissonfördelning:

$$X \sim Po(\mu)$$

$$\Omega_X = \{0, 1, 2, \ldots, \infty\}$$

$$p_X(k) = \frac{\mu^k}{k!} \cdot e^{-\mu}$$

$$E(X) = \mu$$

$$V(X) = \mu$$

$$X \sim Po(\mu_1)\,;\, Y \sim Po(\mu_2) \quad \Longrightarrow \quad X+Y \sim Po(\mu_1+\mu_2)$$

Kontinuerliga fördelningar:

Likformig kontinuerlig fördelning:

$$X \sim U(a,b)$$
$$f_X(x) = 1/(b-a) \quad \text{för} \quad a \le x \le b \quad \text{annars } 0$$
$$F_X(x) = (x-a)/(b-a) \quad \text{för} \quad a \le x \le b \quad \text{annars } 0 \quad \text{respektive } 1$$
$$E(X) = \frac{a+b}{2}$$
$$V(X) = \frac{(b-a)^2}{12}$$

Exponentialfördelning:

$$X \sim Exp(\lambda)$$
$$f_X(x) = \lambda \cdot e^{-\lambda \cdot x} \quad \text{för} \quad x \ge 0 \quad \text{annars } 0$$
$$F_X(x) = 1 - e^{-\lambda \cdot x} \quad \text{för} \quad x \ge 0 \quad \text{annars } 0$$
$$E(X) = 1/\lambda$$
$$V(X) = 1/\lambda^2$$
$$x_\alpha = \frac{-\ln(\alpha)}{\lambda}$$

Normalfördelning:

$$X \sim N(\mu, \sigma)$$
$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad -\infty \le x \le +\infty$$
$$E(X) = \mu$$
$$V(X) = \sigma^2$$
$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{där } \Phi \text{ fördelningsfunktion för } N(0,1)$$
$$\Phi(-a) = 1 - \Phi(a)$$
$$P(a < X \le b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\begin{aligned} P(\mu - \sigma < X \le \mu + \sigma) &\approx 68\% \\ P(\mu - 2\sigma < X \le \mu + 2\sigma) &\approx 95\% \\ P(\mu - 3\sigma < X \le \mu + 3\sigma) &\approx 99.7\% \end{aligned}$$

$$\Phi(\lambda_\alpha) = 1 - \alpha \quad \text{kvantiler för } N(0,1)$$
$$P\left(-\lambda_{\alpha/2} < X \le \lambda_{\alpha/2}\right) = 1 - \alpha$$
$$X \sim N(\mu,\sigma) \quad \text{och} \quad Y = a \cdot X + b \quad \Longrightarrow \quad Y \sim N(a \cdot \mu + b, |a| \cdot \sigma)$$
$$X \sim N(\mu_X, \sigma_X) \quad \text{och} \quad Y \sim N(\mu_Y, \sigma_Y) \quad \Longrightarrow \quad X \pm Y \sim N\left(\mu_X \pm \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$
$$X_i \sim N(\mu,\sigma) \text{ (oberoende)} \quad \Longrightarrow \quad \sum_{i=1}^{n} X_i \sim N\left(n \cdot \mu, \sqrt{n} \cdot \sigma\right)$$
$$X_i \sim N(\mu,\sigma) \text{ (oberoende)} \quad \Longrightarrow \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

gäller också om $X_i \sim i.i.d(\mu,\sigma)$ och n stor (CGS)

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \Longrightarrow \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\bar{X} \sim N\left(\mu_X, \sigma_X/\sqrt{n}\right) \; ; \bar{Y} \sim N\left(\mu_Y, \sigma_Y/\sqrt{n}\right) \quad \Longrightarrow \quad \bar{X} \pm \bar{Y} \sim N\left(\mu_X \pm \mu_Y, \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right)$$

Approximationer:

$$X \sim Bin(n,p) \implies X \sim Po(n \cdot p) \text{ om } p < 0.1$$

$$X \sim Bin(n,p) \implies X \sim N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right) \text{ om } V(X) = n \cdot p \cdot (1-p) > 5$$

$$\text{detta innebär att } F_X^{Bin}(a) \approx \Phi\left(\frac{a + 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right) \text{ (med halvkorrektion)}$$

$$X \sim Po(\lambda) \implies X \sim N\left(\lambda, \sqrt{\lambda}\right) \text{ om } \lambda > 15$$

$$\text{detta innebär att } F_X^{Po}(a) \approx \Phi\left(\frac{a + 0.5 - \lambda}{\sqrt{\lambda}}\right) \text{ (med halvkorrektion)}$$

$$X \sim Hyp(N,n,m) \implies X \sim Bin(n, m/N) \text{ om } n/N < 0.1$$

Linjär regression:

Punktskattningar för intercept α och lutning β:

$$\beta^* = \frac{S_{xy}}{S_{xx}} \text{ och } \alpha^* = \bar{y} - \beta^* \bar{x}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}; \quad S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \quad (S_{yy} \text{ analogt})$$

Skattningar för varians σ^2 och standardavvikelse σ i $\epsilon \sim N(0, \sigma)$:

$$\left(\sigma^2\right)^* = s_r^2 \;;\; \sigma^* = s_r \text{ där } s_r^2 = \frac{Q_0}{n-2} \text{ och } Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

Ekvation för regressionslinjen:

$$\mu_0^* = \alpha^* + \beta^* \cdot x_0$$

KI för lutning β :

$$I_\beta = \beta^* \pm t_{\alpha/2}(n-2) \cdot \frac{s_r}{\sqrt{S_{xx}}}$$

KI för intercept α :

$$I_\alpha = \alpha^* \pm t_{\alpha/2}(n-2) \cdot s_r \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

KI för regressionslinjen μ_0 vid x_0 :

$$I_{\mu_0} = \mu_0^* \pm t_{\alpha/2}(n-2) \cdot s_r \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Förklaringsgrad:

$$R^2 = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}}$$

Pearsons korrelationskoefficient ρ:

$$\rho^2 = R^2 \text{ gäller för enkel linjär regression}$$

Intervallskattning:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right]$$

KI för μ i $N(\mu,\sigma)$; konfidensgrad $1-\alpha$:

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \quad \text{när } \sigma \text{ känd}$$

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2}\cdot\frac{s}{\sqrt{n}} \quad \sigma \text{ okänd } n \geq 30$$

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1)\cdot\frac{s}{\sqrt{n}} \quad \sigma \text{ okänd, litet stickprov}$$

Ensidiga KI för μ i $N(\mu,\sigma)$; konfidensgrad $1-\alpha$:

$$I_\mu = \left(\bar{x} - t_\alpha(n-1)\frac{s}{\sqrt{n}}\ ;\ +\infty\right) \quad \sigma \text{ okänd, litet stickprov}$$

$$I_\mu = \left(-\infty\ ;\ \bar{x} + t_\alpha(n-1)\frac{s}{\sqrt{n}}\right) \quad \sigma \text{ okänd, litet stickprov}$$

KI för varians σ^2 och standardavvikelse σ i $N(\mu,\sigma)$; konfidensgrad $1-\alpha$:

$$I_{\sigma^2} = \left(\frac{f\cdot s^2}{\chi^2_{\alpha/2}(f)};\frac{f\cdot s^2}{\chi^2_{1-\alpha/2}(f)}\right) \quad \text{där } f = n-1$$

$$I_\sigma = \left(\sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}}\cdot s;\sqrt{\frac{f}{\chi^2_{1-\alpha/2}(f)}}\cdot s\right) \quad \text{där } f = n-1$$

KI för $\Delta\mu = \mu_x - \mu_y$ för två normalfördelningar $N(\mu_x,\sigma_x)$ och $N(\mu_y,\sigma_y)$; konfidensgrad $1-\alpha$:

$$I_{\Delta\mu} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2}\cdot D \text{ med } D = \sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}} \text{ när } \sigma_{x,y} \text{ kända}$$

$$I_{\Delta\mu} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2}\cdot d \quad \text{med } d = \sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} \text{ när } \sigma_{x,y} \text{ okända men } n_x \text{ och } n_y \text{ stora } (\geq 25)$$

$$I_{\Delta\mu} = \bar{x} - \bar{y} \pm t_{\alpha/2}(f)\cdot d \quad \text{med } f = n_x + n_y - 2 \text{ när } \sigma_x = \sigma_y = \sigma \text{ okänd och små stickprov}$$

$$\text{och där } d = s_p\cdot\sqrt{\frac{1}{n_x}+\frac{1}{n_y}} \text{ med } s_p = \sqrt{\frac{(n_x-1)\cdot s_x^2 + (n_y-1)\cdot s_y^2}{f}}$$

KI för förskjutning mellan väntevärden $\Delta\mu$ för 2 parade stickprov från N ; konfidensgrad $1-\alpha$:

$$I_{\Delta\mu} = \bar{z} \pm t_{\alpha/2}(n-1)\cdot\frac{s_z}{\sqrt{n}} \quad \text{med } z_i = y_i - x_i$$

$$\text{och där } s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})^2$$

KI för skattning av en proportion (för p i Bin(n, p)):

$$I_p = \frac{x}{n} \pm \lambda_{\alpha/2}\cdot d \quad \text{där } d = \sqrt{\frac{\frac{x}{n}\cdot(1-\frac{x}{n})}{n}}$$

KI för skattning av skillnaden mellan två proportioner (två Bin)):

$$I_{\Delta p} = \frac{x_1}{n_1} - \frac{x_2}{n_2} \pm \lambda_{\alpha/2}\cdot d \quad \text{där } d = \sqrt{\frac{\frac{x_1}{n_1}\cdot(1-\frac{x_1}{n_1})}{n_1}+\frac{\frac{x_2}{n_2}\cdot(1-\frac{x_2}{n_2})}{n_2}}$$

Hypotestest:

Z-test ; ett stickprov ; $X_i \sim N(\mu, \sigma)$; X_i oberoende ; σ känd:

$$H_0 : \mu = \mu_0 \qquad Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \text{om } H_0 \text{ är sann}$$

Kritiska områden:

$$\begin{aligned}
\Omega_\alpha &= \{z > \lambda_\alpha\} \;\text{ för }\; H_a : \mu > \mu_0 \\
\Omega_\alpha &= \{z < -\lambda_\alpha\} \;\text{ för }\; H_a : \mu < \mu_0 \\
\Omega_\alpha &= \left\{|z| > \lambda_{\alpha/2}\right\} \;\text{ för }\; H_a : \mu \neq \mu_0
\end{aligned}$$

T-test ; ett stickprov ; $X_i \sim N(\mu, \sigma)$; X_i oberoende ; σ okänd:

$$H_0 : \mu = \mu_0 \qquad T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \quad \text{om } H_0 \text{ är sann}$$

Kritiska områden:

$$\begin{aligned}
\Omega_\alpha &= \{t > t_\alpha(n-1)\} \;\text{ för }\; H_a : \mu > \mu_0 \\
\Omega_\alpha &= \{t < -t_\alpha(n-1)\} \;\text{ för }\; H_a : \mu < \mu_0 \\
\Omega_\alpha &= \left\{|t| > t_{\alpha/2}(n-1)\right\} \;\text{ för }\; H_a : \mu \neq \mu_0
\end{aligned}$$

T-test ; två oberoende stickprov ; $X_i \sim N(\mu_x, \sigma)$; $Y_i \sim N(\mu_y, \sigma)$; (samma σ som är dock okänd):

$$H_0 : \mu_x - \mu_y = \Delta\mu_0 \qquad T = \frac{\bar{X} - \bar{Y} - \Delta\mu_0}{S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(f) \quad \text{om } H_0 \text{ är sann}$$

$$f = n_x + n_y - 2 \quad ; \quad S_p = \sqrt{\frac{(n_x - 1)\,S_x^2 + (n_y - 1)\,S_y^2}{f}} \quad ; \quad \text{oftast } \Delta\mu_0 = 0$$

Kritiska områden:

$$\begin{aligned}
\Omega_\alpha &= \{t > t_\alpha(f)\} \;\text{ för }\; H_a : \Delta\mu > \Delta\mu_0 \\
\Omega_\alpha &= \{t < -t_\alpha(f)\} \;\text{ för }\; H_a : \Delta\mu < \Delta\mu_0 \\
\Omega_\alpha &= \left\{|t| > t_{\alpha/2}(f)\right\} \;\text{ för }\; H_a : \Delta\mu \neq \Delta\mu_0
\end{aligned}$$

T-test ; två parade stickprov ; $X_i \sim N(\mu_i, \sigma_x)$; $Y_i \sim N(\mu_i + \Delta\mu, \sigma_y)$ $\Delta\mu =$ systemat. förskjutning

$$H_0 : \Delta\mu = \Delta\mu_0 \qquad T = \frac{\bar{Z} - \Delta\mu_0}{S_z/\sqrt{n}} \sim t(n-1) \quad \text{om } H_0 \text{ är sann}$$

$$z_i = x_i - y_i \quad ; \quad s_z^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2 \quad ; \quad \text{oftast } \Delta\mu_0 = 0$$

Kritiska områden:

$$\begin{aligned}
\Omega_\alpha &= \{t > t_\alpha(n-1)\} \;\text{ för }\; H_a : \Delta\mu > \Delta\mu_0 \\
\Omega_\alpha &= \{t < -t_\alpha(n-1)\} \;\text{ för }\; H_a : \Delta\mu < \Delta\mu_0 \\
\Omega_\alpha &= \left\{|t| > t_{\alpha/2}(n-1)\right\} \;\text{ för }\; H_a : \Delta\mu \neq \Delta\mu_0
\end{aligned}$$