Formelsamling i matematisk statistik

1. Sannolikhetsteori

1.1 Allmänt

Fördelningsfunktion: $F(x) = \mathrm{P}(X \le x)$.

Kvantil: Det tal x_α som uppfyller $F(x_\alpha) = 1 - \alpha$.

Diskret s.v.: X är diskret om den antar ett ändligt eller ett uppräkneligt oändligt antal värden $x_1, x_2, \ldots$ med sannolikheter $p(x_1), p(x_2), \ldots$, där $p(x)$ är *sannolikhetsfunktion* till X.

Kontinuerlig s.v.: X är kontinuerlig med *täthetsfunktion* $f(x)$ om X antar alla värden i ett intervall och $F(x) = \int_{-\infty}^{x} f(t)\,dt$ för alla x.

Följande gäller:

(1) $f(x) = F'(x) \ge 0$ för alla x där derivatan existerar,

(2) $\mathrm{P}(a < X < b) = \int_a^b f(x)dx$, (3) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Väntevärde:

$$\mathrm{E}(X) = \mu = \begin{cases} \sum_i x_i\, p(x_i) & (X \text{ diskret}), \\ \int_{-\infty}^{\infty} x\, f(x)\, dx & (X \text{ kontinuerlig}). \end{cases}$$

Väntevärdet för en funktion erhålls analogt som
$E(g(X)) = \sum g(x_i)p(x_i)$ eller $E(g(X)) = \int g(x)f(x)\,dx$.

Varians: $\mathrm{V}(X) = \sigma^2 = \mathrm{E}((X-\mu)^2) = \mathrm{E}(X^2) - \mu^2$.

Standardavvikelse: $\mathrm{D}(X) = \sigma = \sqrt{\mathrm{V}(X)}$.

Kovarians: $\mathrm{C}(X,Y) = \mathrm{E}((X-\mu_X)(Y-\mu_Y)) = \mathrm{E}(XY) - \mu_X \cdot \mu_Y$.

Korrelationskoefficient: $\rho(X,Y) = \dfrac{\mathrm{C}(X,Y)}{\mathrm{D}(X)\cdot\mathrm{D}(Y)}$.

1.2 Diskreta fördelningar

Binomialfördelning: $X \sim \mathrm{Bin}(n,p)$ om $p(k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, 2, \ldots, n$.

$0 \le p \le 1$, $q = 1-p$, $\mu = np$, $\sigma^2 = npq$.

För första gången-fördelning: $X \sim \mathrm{ffg}(p)$ om $p(k) = pq^{k-1}$, $k = 1, 2, 3, \ldots$

$0 < p \le 1$, $q = 1-p$, $\mu = \dfrac{1}{p}$, $\sigma^2 = \dfrac{q}{p^2}$.

Poissonfördelning: $X \sim \mathrm{Po}(m)$ om $p(k) = \dfrac{m^k}{k!} e^{-m}$, $k = 0, 1, 2, \ldots$

$m > 0$, $\mu = m$, $\sigma^2 = m$.

Hypergeometrisk fördelning: $X \sim \mathrm{Hyp}(N,n,p)$ om $p(k) = \dfrac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$,

$k = 0, 1, \ldots, n$, dock $k \le Np$ och $n - k \le Nq$.

$0 \le p \le 1$, $q = 1-p$, $\mu = np$, $\sigma^2 = npq \cdot \dfrac{N-n}{N-1}$.

1.3 Kontinuerliga fördelningar

Rektangelfördelning (Likformig förd.): $X \sim \mathrm{Re}(a, b)$ om $f(x) = \dfrac{1}{b-a}, \quad a \leq x \leq b.$

$\mu = \dfrac{a+b}{2}, \quad \sigma^2 = \dfrac{(b-a)^2}{12}.$

Γ*-fördelning:* $X \sim \Gamma(c, \lambda)$ om $f(x) = \dfrac{x^{c-1}\lambda^c}{\Gamma(c)} e^{-\lambda x}$, $x \geq 0$. $\lambda > 0$, $\mu = c/\lambda$, $\sigma^2 = c/\lambda^2$.

Exponentialfördelning: $X \sim \mathrm{Exp}(\lambda)$ om $X \sim \Gamma(1, \lambda)$, dvs. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$, $\mu = 1/\lambda$, $\sigma^2 = 1/\lambda^2$.

Normalfördelning: $X \sim \mathrm{N}(\mu, \sigma^2)$ om $f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$, $\sigma > 0$.

För $\mathrm{N}(0, 1)$ betecknas fördelningsfunktionen med $\Phi(x)$ och kvantilerna med λ_α.

χ^2*-fördelning:* Om $X_1, X_2, \ldots, X_n$ är oberoende $\mathrm{N}(\mu, \sigma)$ så gäller:

(1) $\dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2(n)$,

(2) $\dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2(n-1)$,

(3) $\bar{X}$ och $\displaystyle\sum_{i=1}^{n} (X_i - \bar{X})^2$ är oberoende.

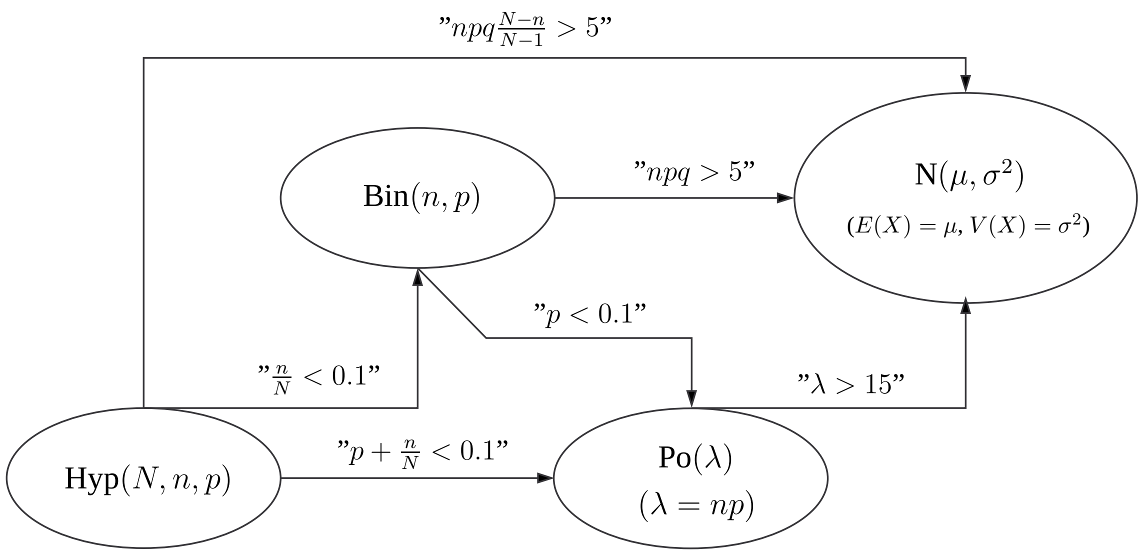
Vidare, om $U \sim \chi^2(f_1)$ och $V \sim \chi^2(f_2)$ är oberoende så är $U + V \sim \chi^2(f_1 + f_2)$.
Parametern f i $\chi^2(f)$-fördelningen kallas antalet *frihetsgrader* och kvantilerna betecknas $\chi^2_\alpha(f)$. $\mu = f$, $\sigma^2 = 2f$.

t*-fördelning:* Om $X \sim \mathrm{N}(0, 1)$ och $Y \sim \chi^2(f)$ är oberoende så gäller: $\dfrac{X}{\sqrt{Y/f}} \sim t(f)$.

t-fördelningen med f frihetsgrader är symmetrisk kring 0 med kvantiler $t_\alpha(f)$.

$\mu = 0$ (om $f > 1$), $\quad \sigma^2 = \dfrac{f}{f-2}$ (om $f > 2$).

Approximationer:



2. Statistikteori

2.1 Allmänt

Stickprov: Mätvärdena $x_1,\ x_2, \ldots,\ x_n$ utgör ett stickprov (slumpmässigt stickprov) på X om de är observationer av slumpvariabler (oberoende slumpvariabler) $X_1,\ X_2, \ldots,\ X_n$ med samma fördelning som X.

Stickprovsmedelvärde: $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$.

Stickprovsvarians: $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 = \dfrac{S_{xx}}{n-1}$, där

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n} = \sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2.$$

Standardavvikelse i ett stickprov: $s = \sqrt{s^2}$.

2.2 Inferens vid normalfördelning

2.2.1 Ett stickprov

$x_1, x_2, \ldots, x_n$ slumpmässigt stickprov från $\mathrm{N}(\mu, \sigma^2)$.

$$\text{Referensvariabler:} \quad \begin{aligned} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} &\sim \mathrm{N}(0,1) \quad (\sigma \text{ känd}), \\ \frac{\bar{X} - \mu}{s/\sqrt{n}} &\sim t(n-1) \quad (\sigma \text{ okänd}), \\ \frac{(n-1)\cdot s^2}{\sigma^2} &\sim \chi^2(n-1). \end{aligned}$$

2.2.2 Två stickprov

$x_1, x_2, \ldots, x_{n_1}$ slumpmässigt stickprov från $\mathrm{N}(\mu_1, \sigma_1^2)$.
$y_1,\ y_2, \ldots, y_{n_2}$ slumpmässigt stickprov från $\mathrm{N}(\mu_2, \sigma_2^2)$.

$$\text{Referensvariabler:} \quad \begin{aligned} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} &\sim \mathrm{N}(0,1) \quad (\sigma_1 \text{ och } \sigma_2 \text{ kända}), \\ \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} &\sim t(n_1 + n_2 - 2) \quad (\sigma_1 = \sigma_2 = \sigma \text{ okänd}), \\ &\text{där} \quad s_p^2 = \frac{S_{xx} + S_{yy}}{n_1 + n_2 - 2}, \\ \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} &\approx \mathrm{N}(0,1) \quad (\sigma_1 \text{ och } \sigma_2 \text{ okända}), \\ &\text{för någorlunda stora stickprov.} \end{aligned}$$

2.2.3 Approximationer vid stora stickprov

Om $\theta^* \approx \mathrm{N}(\theta, \mathrm{D}^2(\theta^*))$, med medelfel $d(\theta^*) = (\mathrm{D}(\theta^*))^*$ så gäller:

$$\frac{\theta^* - \theta}{\mathrm{D}(\theta^*)} \approx \mathrm{N}(0,1) \quad \text{(om } \mathrm{D}(\theta^*) \text{ känd)},$$

Referensvariabler:

$$\frac{\theta^* - \theta}{d(\theta^*)} \approx \mathrm{N}(0,1) \quad \text{(om } \mathrm{D}(\theta^*) \text{ okänd)},$$

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \approx t(n-1).$$

2.3 Regressionsanalys

Modell: Talparen $(x_1, y_1), \ldots, (x_n, y_n)$ har observerats. Här är $x_1, \ldots, x_n$ fixa tal och y_i observation av $Y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, 2, \ldots, n$, där ε_i är oberoende $\mathrm{N}(0, \sigma^2)$.

Låt $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)/n = \sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y}$.

$\beta^* = \frac{S_{xy}}{S_{xx}}$ är observation från $\mathrm{N}\Big(\beta, \frac{\sigma^2}{S_{xx}}\Big)$,

$\alpha^* = \bar{y} - \beta^* \bar{x}$ är observation från $\mathrm{N}\Big(\alpha, \sigma^2\big(\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}\big)\Big)$,

$\mu_0 = \alpha + \beta x_0$ skattas med $\bar{y} + \beta^*(x_0 - \bar{x})$, som är observation från $\mathrm{N}\Big(\mu_0, \sigma^2\big(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\big)\Big)$,

$\sigma^* = s_r$, där $s_r^2 = \frac{Q_0}{n-2}$ med $Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$ och $\frac{Q_0}{\sigma^2} \sim \chi^2(n-2)$.

Vidare är $\bar{y}$, β^* och Q_0 oberoende.

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

Referensvariabler:

$$\frac{\alpha^* - \alpha}{s_r\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}}} \sim t(n-2),$$

$$\frac{\beta^* - \beta}{s_r/\sqrt{S_{xx}}} \sim t(n-2),$$

$$\frac{\mu_0^* - \mu_0}{s_r\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t(n-2).$$

3 Några användbara summor

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2},$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6},$$

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}, \quad \text{om } |a| < 1.$$