Formelblad – Statistik och sannolikhetslära (LMA120)

Bayes formel: Låt A och D vara två händelser. Då gäller

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)}.$$

Diskreta fördelningar

Binomialfördelning: $\xi \sim Bin(n, p)$

- sannolikhetsfunktionen $P(\xi = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$
- väntevärdet np och variansen $np(1-p)$

Poissonfördelning: $\xi \sim Poi(\lambda)$

- sannolikhetsfunktionen $P(\xi = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, ...$
- väntevärdet λ och variansen λ

Hypergeometrisk fördelning: $\xi \sim HypGeom(n, N, m)$

- sannolikhetsfunktionen $P(\xi = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, ..., n$
- väntevärdet $\frac{nm}{N}$ och variansen $\frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right)$

Kontinuerliga fördelningar

Rektangelfördelning (likformig fördelning): $\xi \sim Lik(a, b)$

- frekvensfunktionen $f(x) = \frac{1}{b-a}, \quad a < x < b$
- väntevärdet $\frac{1}{2}(a+b)$ och variansen $\frac{1}{12}(b-a)^2$

Normalfördelning: $\xi \sim N(\mu, \sigma)$

- frekvensfunktionen $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad -\infty < x < \infty$
- väntevärdet μ och variansen σ^2

Standardiserad normalfördelning: $Z \sim N(0, 1)$

- frekvensfunktionen $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$

Exponentialfördelning: $\xi \sim Exp(\lambda)$

- frekvensfunktionen $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$
- väntevärdet $\frac{1}{\lambda}$ och variansen $\frac{1}{\lambda^2}$

Summor av oberoende stokastiska variabler

- $\xi_1 \sim Bin(n_1, p)$ och $\xi_2 \sim Bin(n_2, p)$: $\xi_1 + \xi_2 \sim Bin(n_1 + n_2, p)$
- $\xi_1 \sim Poi(\lambda_1)$ och $\xi_2 \sim Poi(\lambda_2)$: $\xi_1 + \xi_2 \sim Pois(\lambda_1 + \lambda_2)$
- $\xi_1 \sim N(\mu_1, \sigma_1)$ och $\xi_2 \sim N(\mu_2, \sigma_2)$: $\xi_1 + \xi_2 \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

Centrala gränsvärdessatsen: Låt $\xi_i, i = 1, ..., n$, vara oberoende och lika fördelade stokastiska variabler med väntevärdet μ och standardavvikelsen σ. Om n är tillräckligt stort (> 30) är

$$\sum_{i=1}^{n} \xi_i \approx N(n\mu, \sqrt{n}\sigma)$$

och

$$\frac{\sum\limits_{i=1} \xi - n\mu}{\sqrt{n}\sigma} \approx N(0,1)$$

Statistikor

- **Ett stickprov**: Låt $\xi_1, \xi_2, \ldots, \xi_n$ vara oberoende normalfördelade stokastiska variabler med väntevärde μ och varians σ^2.
  - *Stickprovsväntevärde*: $\bar{\xi} = \frac{1}{n} \sum\limits_{i=1}^{n} \xi_i$ är en väntevärdesriktig skattning för μ, och
  $$\bar{\xi} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$
  - *Stickprovsvarians*: $S^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (\xi_i - \bar{\xi})^2 = \frac{1}{n-1} \left(\sum\limits_{i=1}^{n} \xi_i^2 - n\bar{\xi} \right)$ är en väntevärdesriktig skattning för σ^2.
  - Ett två-sidigt $100(1-\alpha)$%-igt konfidensintervall för μ, när σ är känt:
  $$\bar{\xi} \pm z_{\alpha/2}\sigma/\sqrt{n},$$
  där $z_{\alpha/2}$ är sådant att $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$ och $Z \sim N(0,1)$
  - *Ett stickprov t-statistika*:
  $$\frac{\bar{\xi} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$
  - Ett två-sidigt $100(1-\alpha)$%-igt konfidensintervall för μ, när σ är okänt:
  $$\bar{\xi} \pm t_{\alpha/2}^{(n-1)} S/\sqrt{n},$$
  där $t_{\alpha/2}^{(n-1)}$ är sådant att $P(T \leq t_{\alpha/2}^{(n-1)}) = 1 - \alpha/2$ och $T \sim T_{n-1}$
- **Två stickprov, lika varians**: Låt $\xi_1, \xi_2, \ldots, \xi_n$ och $\nu_1, \nu_2, \ldots, \nu_m$ vara två oberoende stickprov från en $N(\mu_1, \sigma)$- respektive $N(\mu_2, \sigma)$-fördelning. Låt S_1^2 respektive S_2^2 beteckna stickprovsvariansen för vardera stickprov.
  - *Sammanvägd stickprovsvarians*: $S_p^2 = \frac{(n-1)S_1^2 + (m-1)s_2^2}{n+m-2}$ är en väntevärdesriktig skattare för σ^2.
  - *Två stickprov t-statistika*:
  $$\frac{\bar{\xi} - \bar{\nu} - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T_{n+m-2}$$

- Ett två-sidigt $100(1-\alpha)$%-igt konfidensintervall för $\mu_1 - \mu_2$, när $\sigma_1 = \sigma_2 = \sigma$ och σ är känt:

$$\bar{\xi} - \bar{\nu} \pm z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}},$$

  där $z_{\alpha/2}$ är sådant att $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$ och $Z \sim N(0,1)$;

- Ett två-sidigt $100(1-\alpha)$%-igt konfidensintervall för $\mu_1 - \mu_2$, när $\sigma_1 = \sigma_2 = \sigma$ och σ är okänt:

$$\bar{\xi} - \bar{\nu} \pm t_{\alpha/2}^{(n+m-2)} S_p \sqrt{\frac{1}{n} + \frac{1}{m}},$$

  där $t_{\alpha/2}^{(n+m-2)}$ är sådant att $P(T \leq t_{\alpha/2}^{(n+m-2)}) = 1 - \alpha/2$ och $T \sim T_{n+m-2}$

- **Två stickprov, parvisa observationer**: Låt $(\xi_1, \nu_1), (\xi_2, \nu_2), \ldots, (\xi_n, \nu_n)$ vara ett stickprov av parvisa observationer sådana att ξ_i och ν_i är normalfördelade med parametrar μ_1 och σ_1, respektive μ_2 och σ_2. Sätt $D_i = \xi_i - \nu_i$ och låt S_D^2 vara stickprovsvariansen för $D_1, D_2, \ldots, D_n$. Då är

  - *t-statistika*:

$$\frac{\bar{D} - (\mu_1 - \mu_2)}{S_D/\sqrt{n}} \sim T_{n-1}$$