Formelsamling för TMS063

Händelser (events):

$\cup$= union, $\cap$= snitt, A' = komplement.

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

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

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

Disjunkta händelser:

$E_1 \cap E_2 = \phi$ och $P(E_1 \cap E_2) = 0$

Totala sannolikhetslagen:

För disjunkta $E_1, E_2, \dots, E_k$ där $\bigcup_{i=1}^{k} E_i = S$ och S är hela utfallsrummet.

$$P(B) = P(B \cap E_1) + \cdots + P(B \cap E_k)$$
$$= P(B|E_1)P(E_1) + \cdots P(B|E_k)P(E_k)$$

Oberoende händelser A och B:

$P(A \cap B) = P(A)P(B)$ och $P(A|B) = P(A)$.

Bayes sats:

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ för $P(B) > 0$

Räknetekniker (kombinatorik)

Multiplikationsregeln:

Om n_i alternativ i steg $i = 1, \dots, r$:

$n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_r$

Antal permutationer bland n objekt:

$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1$

Permutationer av liknande objekt:

Om $n = n_1 + n_2 + \cdots + n_r$:

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}$$

Välja r objekt bland n när:

- Ordningen spelar roll, utan återläggning
$$\frac{n!}{(n-r)!}$$
- Ordningen spelar roll, med återläggning
$$n^r$$
- Ordningen kvittar, utan återläggning
$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$
- Ordningen kvittar, med återläggning
$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!}$$

Diskreta stokastiska variabler

Frekvensfunktion:

(1) $f(x) \geq 0$ för alla x och $\sum_{i=1}^{n} f(x_i) = 1$

(2) $f(x) = P(X = x)$

Kumulativ fördelningsfunktion:

(1) $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$

(2) $0 \leq F(x) \leq 1$

(3) Om $x \leq y$ så gäller $F(x) \leq F(y)$

Väntevärde och varians:

$$\mu = E[X] = \sum_x x\, f(x)$$
$$\sigma^2 = V(X) = E[(X-\mu)^2]$$
$$= \sum_x (x-\mu)^2 f(x) = E[X^2] - (E[X])^2$$

För en funktion $h(X)$: $E[h(X)] = \sum_x h(x) f(x)$

Diskreta fördelningar

Binomialfördelning $\boldsymbol{Bin(n,p)}$

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$

$E[X] = np, V(X) = np(1-p)$

Geometrisk fördelning $\boldsymbol{Geo(p)}$

$f(x) = (1-p)^{x-1} p$

$E[X] = 1/p, V(X) = (1-p)/p^2$

Poissonfördelning $\boldsymbol{Poi(\lambda)}$

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\dots$$

$\mu = E[X] = \lambda, \sigma^2 = V(X) = \lambda$

Kontinuerliga stokastiska variabler

Täthetsfunktion:

(1) $f(x) \geq 0$ för alla x och $\int_{-\infty}^{\infty} f(x)dx = 1$

(2) $P(a \leq X \leq b) = \int_a^b f(x)dx$

(3) $P(X = a) = 0$

Kumulativ fördelningsfunktion:

(1) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$

(2) $0 \leq F(x) \leq 1$

(3) Om $x \leq y$ så gäller $F(x) \leq F(y)$

(4) $P(a \leq X \leq b) = F(b) - F(a)$

(5) $P(a \leq X \leq b) = P(a < X < b)$

Väntevärde och varians:

$$\mu = E[X] = \int_{-\infty}^{\infty} x\, f(x)dx$$

$\sigma^2 = V(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$

För en funktion $h(X)$:

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

Kontinuerliga fördelningar

Likformig fördelning $U(a, b)$

$f(x) = \frac{1}{b-a}, a \le x \le b$

$\mu = E[X] = \frac{a+b}{2},\ \sigma^2 = V(X) = \frac{(b-a)^2}{12}$

Normalfördelning $N(\mu, \sigma^2)$

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

$E[X] = \mu, V(X) = \sigma^2$

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

Exponentialfördelning $Exp(\lambda)$

$f(x) = \lambda e^{-\lambda x}, x \ge 0$

$\mu = E[X] = 1/\lambda, \sigma^2 = V(X) = 1/\lambda^2$

$P(X < t_1 + t_2 | X > t_2) = P(X < t_1)$

Poissonprocess

$N(t) =$ antal impulser i intervallet $[0, t]$

$N(t) \sim Poi(\lambda t)$, där $\lambda > 0$ är intensiteten.

$X =$ tiden mellan två impulser $\sim Exp(\lambda)$

Normalapproximation

Centrala gränsvärdessatsen

För oberoende, lika fördelade $X_1, \dots, X_n$ med $E[X_i] = \mu$ och $Var(X_i) = \sigma^2$ gäller att

$\sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2)$ och $\bar{X} \approx N(\mu, \sigma^2/n)$.

Av Binomialfördelningen:

$X \sim Bin(n, p)$ och $np > 5$ och $n(1-p) > 5$

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1)$$

Med kontinuitetskorrektion:

$$P(X \le x) = P(X \le x + 0.5) \approx \Phi\left(\frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$$

Av Poissonfördelningen:

$X \sim Poi(k)$ och > 20.

$$Z = \frac{X - k}{\sqrt{k}} \approx N(0,1)$$

Med kontinuitetskorrektion:

$$P(X \le x) = P(X \le x + 0.5) \approx \Phi\left(\frac{x + 0.5 - k}{\sqrt{k}}\right)$$

Gemensamma fördelningar för *X* och *Y*

Diskreta variabler – frekvensfunktion:

(1) $f_{XY}(x, y) \ge 0$

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$

Kontinuerliga variabler – täthetsfunktion

(1) $f_{XY}(x, y) \ge 0$

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$

(3) För en region $R \in \mathbb{R}^2$

$$P\big((X, Y) \in R\big) = \iint f_{XY}(x, y) dx dy$$

Marginalfördelning

$f_X(x) = P(X = x) = \sum_y f_{XY}(x, y)$ (diskreta)

$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$ (kontinuerliga)

Betingad fördelning

(1) $f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)} \ge 0$

(2) $f_{Y|x}(y) \ge 0$

(3) $\sum_y f_{Y|x(y)} = 1$ (diskret)

$\int_{-\infty}^{\infty} f_{Y|x}(y) dy = 1$ (kont.)

(4) $P(Y = y | X = x) = f_{Y|x}(y)$ (diskret)

$P(Y \in B | X = x) = \int_B f_{Y|x}(y) dy$ (kont.)

Kovarians

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

Korrelation

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Räkneregler för väntevärde och varians:

Oavsett om *X* och *Y* är oberoende eller inte, för konstanter *a* och *b*:

$E[aX + b] = aE[X] + b$

$V(aX + b) = a^2 V(X) + b$

$E[X + Y] = E[X] + E[Y]$

$E[XY] = E[X]E[Y] + Cov(X, Y)$

$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$

$V(X - Y) = V(X) + V(Y) - 2Cov(X, Y)$

För oberoende variabler gäller

$Cov(X, Y) = 0$

Parameterskattning

Maximum Likelihood

För stickprovet $X_1, \dots, X_n$, frekvens/täthetsfunktion $f(x)$ och parametrar $\theta_1, \dots, \theta_m$ bilda log-likelihoodfunktionen

$l(\theta_1, \dots, \theta_m) = \sum_{i=1}^{n} \log f(x_i)$

och lös ekvationssystemet $k = 1, \dots, m$

$$\left\{\frac{\partial l}{\partial \theta_k} = 0\right\}$$

Väntevärdesriktig (unbiased) skattning

$E[\hat{\theta}] = \theta$

Stickprovs-medelvärde och varians

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

Konfidensintervall – ett stickprov

För μ, när σ^2 är känd:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Två-sidigt intervall:

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

Ensidigt intervall:

$\mu > \bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$ eller $\mu < \bar{x} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$

För μ, när σ^2 är okänd:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\mu = \bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Ensidigt intervall:

$\mu > \bar{x} - t_{n-1,1-\alpha}\frac{\sigma}{\sqrt{n}}$ (undre) eller

$\mu < \bar{x} + t_{n-1,1-\alpha}\frac{\sigma}{\sqrt{n}}$ (övre)

För σ^2:

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}$$

Ensidigt intervall: byt $\alpha/2$ mot α i χ^2_{n-1}.

Hypotestest av $N(\mu, \sigma^2)$

Test av μ, σ^2 känd:

$$H_0\!: \mu = \mu_0,\ \ Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Alternativ:	Förkasta om:
$H_1\!: \mu \ne \mu_0$	$\lvert Z_0\rvert \ge z_{1-\alpha/2}$
$H_1\!: \mu < \mu_0$	$Z_0 \le z_\alpha$
$H_1\!: \mu > \mu_0$	$Z_0 \ge z_{1-\alpha}$

Test av μ, σ^2 okänd:

$$H_0\!: \mu = \mu_0,\ \ T_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

Alternativ:	Förkasta om:
$H_1\!: \mu \ne \mu_0$	$\lvert T_0\rvert \ge t_{n-1,1-\alpha/2}$
$H_1\!: \mu < \mu_0$	$T_0 \le t_{n-1,\alpha}$
$H_1\!: \mu > \mu_0$	$T_0 \ge t_{n-1,1-\alpha}$

P-värde:

Den "exakta" signifikansnivån.

$$P = \begin{cases} 2(1 - \Phi(|z_0|) \text{ om } H_1\!: \mu \ne \mu_0 \\ 1 - \Phi(z_0) \text{ om } H_1\!: \mu > \mu_0 \\ \Phi(z_0) \text{ om } H_1\!: \mu < \mu_0 \end{cases}$$

Test av σ^2:

$$H_0\!: \sigma^2 = \sigma_0^2,\ \ X_0^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Alternativ:	Förkasta om:
$H_1\!: \sigma^2 \ne \sigma_0^2$	$X_0^2 \le \chi^2_{n-1,1-\alpha/2}$ eller $X_0^2 \ge \chi^2_{n-1,\alpha/2}$
$H_1\!: \sigma^2 < \sigma_0^2$	$X_0^2 \le \chi^2_{1-\alpha,n-1}$
$H_1\!: \sigma^2 > \sigma_0^2$	$X_0^2 \ge \chi^2_{\alpha,n-1}$

Inferens på andelar

Skattning:

För $X \sim Bin(n,p)$ gäller att

$$\hat{p} = \frac{X}{n} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

Konfidensintervall:

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

Hypotestest:

$$H_0\!: p = p_0,\ \ Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$$

Alternativ:	Förkasta om:
$H_1\!: p \ne p_0$	$\lvert Z_0\rvert \ge z_{1-\alpha/2}$
$H_1\!: p < p_0$	$Z_0 \le z_\alpha$
$H_1\!: p > p_0$	$Z_0 \ge z_{1-\alpha}$

Inferens på skillnad i andelar

$X \sim Bin(n_Y, p_X)$ och $Y \sim Bin(n_Y, p_Y)$

Skattning:

$$\hat{p}_X - \hat{p}_Y = \frac{X}{n_X} - \frac{Y}{n_Y}$$

och

$$\hat{p}_X - \hat{p}_Y \approx N\left(p_X - p_Y, \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right)$$

Konfidensintervall:

$$p_X - p_Y = \hat{p}_X - \hat{p}_Y \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}$$

Hypotestest:

$H_0: p_X - p_Y = p_0$,

$$Z_0 = \frac{\hat{p}_X - \hat{p}_Y - p_0}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \approx N(0,1)$$

Alternativ:	Förkasta om:
$H_1: p_X - p_Y \neq p_0$	$\lvert Z_0 \rvert \geq z_{1-\alpha/2}$
$H_1: p_X - p_Y < p_0$	$Z_0 \leq z_\alpha$
$H_1: p_X - p_Y > p_0$	$Z_0 \geq z_{1-\alpha}$

Hypotestest– två oberoende stickprov

För $\mu_1 - \mu_2$, σ_1^2 och σ_2^2 kända:

$$H_0: \mu_1 - \mu_2 = d_0,\ \ Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Alternativ:	Förkasta om:
$H_1: \mu_1 - \mu_2 \neq d_0$	$\lvert Z_0 \rvert \geq z_{1-\alpha/2}$
$H_1: \mu_1 - \mu_2 < d_0$	$Z_0 \leq z_\alpha$
$H_1: \mu_1 - \mu_2 > d_0$	$Z_0 \geq z_{1-\alpha}$

För $\mu_1 - \mu_2$, σ_1^2 och σ_2^2 okända men lika:

$$H_0: \mu_1 - \mu_2 = d_0,\ \ T_0 = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$T_0 \sim t_{n_1+n_2-2}$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Alternativ:	Förkasta om:
$H_1: \mu_1 - \mu_2 \neq d_0$	$\lvert T_0 \rvert \geq t_{n_1+n_2-2,1-\alpha/2}$
$H_1: \mu_1 - \mu_2 < d_0$	$t \leq t_{n_1+n_2-2,\alpha}$
$H_1: \mu_1 - \mu_2 > d_0$	$t_0 \geq t_{n_1+n_2-2,1-\alpha}$

För $\mu_1 - \mu_2$, σ_1^2 och σ_2^2 okända och olika

$$H_0: \mu_1 - \mu_2 = d_0,\ \ T_0 = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\gamma$$

där γ skattas (avrundas nedåt)

$$\hat{\gamma} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}\right)}$$