

FORMELSAMLING  
MATEMATISK STATISTIK, AK, FMS012 (9HP) OCH FMSF20 (7.5HP)  
[UPPDATERAD 2007-09-12]

## Sannolikhetsteori

### Sannolikhetsteorins grunder

- Följande gäller för sannolikheter:
  - \*  $0 \leq \mathbf{P}(A) \leq 1$
  - \*  $\mathbf{P}(\Omega) = 1$
  - \*  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ , om händelserna  $A$  och  $B$  är oförenliga (disjunkta).
- Additionssatsen för två händelser:  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ .
- Betingad sannolikhet:  $\mathbf{P}(B | A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$ .
- "Satsen om total sannolikhet":  $\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A | H_i) \mathbf{P}(H_i)$ ,  
 där händelserna  $H_1, \dots, H_n$  är parvis oförenliga (disjunkta) händelser och  $\bigcup_{i=1}^n H_i = \Omega$ .
- $A$  och  $B$  är oberoende  $\iff \mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$ .
- Antalet olika sätt,  $m$ , att dra  $k$  element ur  $n$  är:
  - \* Med återläggning, med hänsyn till ordning:  $m = n^k$
  - \* Med återläggning, utan hänsyn till ordning:  $m = \binom{n+k-1}{k}$
  - \* Utan återläggning, med hänsyn till ordning:  $m = n(n-1) \dots (n-k+1)$
  - \* Utan återläggning, utan hänsyn till ordning:  $m = \binom{n}{k}$

### Endimensionella stokastiska variabler

- Fördelningsfunktion för  $X$ :  $F_X(x) = \mathbf{P}(X \leq x)$ .
- Sannolikhetsfunktion för en diskret s.v.  $X$ :  $p_X(k) = \mathbf{P}(X = k)$ .
- Täthetsfunktionen för en kontinuerlig s.v.  $X$ :  $f_X(x) = \frac{dF_X(x)}{dx}$ .
- $\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a) = \begin{cases} \sum_{k=a+1}^b p_X(k) & \text{om } X \text{ är diskret och } a \text{ och } b \text{ är heltal,} \\ \int_a^b f_X(x) dx & \text{om } X \text{ är kontinuerlig.} \end{cases}$

## Flerdimensionella stokastiska variabler

- Simultan fördelningsfunktion:

$$F_{X,Y}(x,y) = \mathbf{P}(X \leq x, Y \leq y) = \begin{cases} \sum_{j \leq x, k \leq y} p_{X,Y}(j,k) & \text{om } (X, Y) \text{ är diskret,} \\ \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(t,u) dt du & \text{om } (X, Y) \text{ är kontinuerlig.} \end{cases}$$

- $\mathbf{P}(g(X, Y) \in A) = \iint_{g(x,y) \in A} f_{X,Y}(x,y) dx dy.$
- Marginell täthetsfunktion för  $X$ :  $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy.$
- Att  $X$  och  $Y$  är oberoende är ekvivalent med att  $F_{X,Y}(x,y) = F_X(x) F_Y(y)$  för alla  $x$  och  $y$ , samt
  - \*  $p_{X,Y}(j,k) = p_X(j) p_Y(k)$  för alla  $j$  och  $k$  om  $X$  och  $Y$  är diskreta s.v..
  - \*  $f_{X,Y}(x,y) = f_X(x) f_Y(y)$  för alla  $x$  och  $y$  om  $X$  och  $Y$  är kontinuerliga s.v..
- Betingad sannolikhetsfunktion för  $X$ , givet  $Y = k$ :  $p_{X|Y=k}(j) = \frac{p_{X,Y}(j,k)}{p_Y(k)}$
- Betingad täthetsfunktion för  $X$ , givet  $Y = y$ :  $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

## Summer av stokastiska variabler

- Om  $X$  och  $Y$  är oberoende, så gäller för  $Z = X + Y$ ,

$$\begin{aligned} * p_Z(k) &= \sum_{i=0}^k p_X(i) p_Y(k-i), \\ * f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx. \end{aligned}$$

## Väntevärden

- Väntevärdet av  $g(X, Y)$ :

$$\mathbf{E}(g(X, Y)) = \begin{cases} \sum_{j,k} g(j,k) p_{X,Y}(j,k) & \text{om } (X, Y) \text{ är diskret,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy & \text{om } (X, Y) \text{ är kontinuerlig.} \end{cases}$$

- Väntevärden är linjära, dvs  $\mathbf{E}(ag(X) + bh(Y)) = a\mathbf{E}(g(X)) + b\mathbf{E}(h(Y)).$
- Varians:  $\mathbf{V}(X) = \mathbf{E}[(X - \mathbf{E}(X))^2] = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2.$
- Standardavvikelse:  $\mathbf{D}(X) = \sqrt{\mathbf{V}(X)}.$
- Kovarians:  $\mathbf{C}(X, Y) = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))] = \mathbf{E}(XY) - \mathbf{E}(X) \mathbf{E}(Y).$
- Korrelationskoefficient:  $\rho(X, Y) = \frac{\mathbf{C}(X, Y)}{\mathbf{D}(X) \mathbf{D}(Y)}.$

- Kovariansen är bilinjär, dvs  $\mathbf{C}(\sum_j a_j X_j, \sum_k b_k Y_k) = \sum_j \sum_k a_j b_k \mathbf{C}(X_j, Y_k)$ .
- Väntevärde av linjärkombination  $\mathbf{E}(\sum_i a_i X_i + b) = \sum_i a_i \mathbf{E}(X_i) + b$ .
- Varians av linjärkombination  $\mathbf{V}(\sum_i a_i X_i + b) = \sum_i a_i^2 \mathbf{V}(X_i) + 2 \sum_{i < j} a_i a_j \mathbf{C}(X_i, X_j)$ .
- $X, Y$  oberoende  $\Rightarrow X, Y$  okorrelerade, dvs  $\mathbf{C}(X, Y) = 0$ .
- Betingat väntevärde för  $X$ , givet  $Y = k$ :  $\mathbf{E}(X | Y = k) = \sum_j j p_{X|Y=k}(j)$ .
- Betingat väntevärde för  $X$ , givet  $Y = y$ :  $\mathbf{E}(X | Y = y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x) dx$ .
- För betingade väntevärden gäller

$$\mathbf{E}(X) = \begin{cases} \sum_k \mathbf{E}(X | Y = k) p_Y(k), \\ \int_{-\infty}^{\infty} \mathbf{E}(X | Y = y) f_Y(y) dy. \end{cases}$$

- Gauss approximationsformler:

$$\mathbf{E}(g(X_1, \dots, X_n)) \approx g(\mathbf{E}(X_1), \dots, \mathbf{E}(X_n)).$$

$$\mathbf{V}(g(X_1, \dots, X_n)) \approx \sum_{i=1}^n c_i^2 \mathbf{V}(X_i) + 2 \sum_{i < j} c_i c_j \mathbf{C}(X_i, X_j),$$

där  $c_i = \left. \frac{\partial g}{\partial x_i}(x_1, \dots, x_n) \right|_{x_k = \mathbf{E}(X_k), \forall k}$ .

### Normalfördelning och centrala gränsvärdessatsen

- Om  $X_1, \dots, X_n$  är oberoende och  $N(\mu_1, \sigma_1), \dots, N(\mu_n, \sigma_n)$  och  $c_1, \dots, c_n \in \mathbb{R}$ , så gäller att

$$\sum_{i=1}^n c_i X_i \in N\left(\sum_{i=1}^n c_i \mu_i, \sqrt{\sum_{i=1}^n c_i^2 \sigma_i^2}\right).$$

- Centrala gränsvärdessatsen (CGS):

Om  $X_1, X_2, \dots, X_n$  är oberoende och likafördelade med  $\mathbf{E}(X_i) = \mu$  och  $\mathbf{D}(X_i) = \sigma$ , så gäller att

$$Y_n = X_1 + \dots + X_n \in \text{AsN}(\mathbf{E}(Y_n), \mathbf{D}(Y_n)),$$

då  $n \rightarrow \infty$ .

- Med utnyttjande av, bland annat, CGS gäller följande approximationer

Hypergeometrisk	$\rightarrow$	Binomial	om $n/N \leq 0.1$ .
Hypergeometrisk	$\rightarrow$	Poisson	om $p + n/N \leq 0.1$ och $n \geq 10$ .
Hypergeometrisk	$\rightarrow$	Normal	om $\frac{N-n}{N-1} npq \geq 10$ .
Binomial	$\rightarrow$	Poisson	om $p \leq 0.1$ och $n \geq 10$ .
Binomial	$\rightarrow$	Normal	om $npq \geq 10$ .
Poisson	$\rightarrow$	Normal	om $\mu \geq 15$ .

## Stokastiska processer med diskret tid

- Övergångssannolikhet:

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$

$\mathbf{P} = \{p_{ij}\}$  är övergångsmatrisen.

- Övergångssannolikhet av ordning  $m$ :

$$p_{ij}^{(m)} = P(X_{n+m} = j \mid X_n = i)$$

$\mathbf{P}^{(m)} = \{p_{ij}^{(m)}\}$  är övergångsmatrisen av ordning  $m$ .

- Chapman-Kolmogorovs sats:

$$\mathbf{P}^{(m)} = \mathbf{P}^m$$

- Absoluta sannolikheter:  $p_i(n) = P(X_n = i)$ .  $\mathbf{p}(n)$  är radvektorn  $\{p_i(n)\}$ .  $\mathbf{p}(0)$  kallas initialfördelningen.

$$\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{P} = \mathbf{p}(0)\mathbf{P}^n$$

- Stationär fördelning:  $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$
- Beständighet: tillstånd  $i$  är beständigt om

$$P(X_n = i \text{ för något } n > 0 \mid X_0 = i) = 1$$

annars transient.

- Kommunikation: tillstånd  $i$  kommunicerar ensidigt med  $j$  om  $p_{ij}^{(m)} > 0$  för något  $m > 0$ . Om  $i$  kommunicerar ensidigt med  $j$  och vice versa kommunicerar  $i$  och  $j$  tvåsidigt.
- Irreducibilitet: alla tillstånd kommunicerar tvåsidigt. Alla tillstånd är då antingen transienta, positivt beständiga eller nollbeständiga, och de har samma period.
- Existens av stationär fördelning: om  $\{X_n\}$  är irreducibel så existerar en (unik) stationär fördelning om och endast om tillstånden är positivt beständiga.

## Poissonprocessen

- En homogen Poissonprocess  $\{X(t), t \geq 0\}$  har oberoende och stationära ökningar och

$$X(t+s) - X(s) \in Po(\lambda t).$$

- Avstånden mellan konsekutiva händelser är oberoende och  $Exp(\lambda)$ -fördelade.
- En inhomogen Poissonprocess  $\{X(t), t \geq 0\}$  har oberoende ökningar och

$$X(t+s) - X(s) \in Po\left(\int_s^{s+t} \lambda(u) du\right).$$

Fördelning	Tabell 1: Vanliga fördelningar		Väntevärde	Varians
Binomialfördelning, $Bin(n, p)$	$p(k) = \binom{n}{k} p^k q^{n-k}$	$k = 0, 1, \dots, n$	$np$	$npq$
Hypergeometrisk fördelning	$p(k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}$	$k \leq Np,$ $n - k \leq Nq$	$np$	$\frac{N-n}{N-1} npq$
Poissonfördelning, $Po(\mu)$	$p(k) = e^{-\mu} \frac{\mu^k}{k!}$	$k = 0, 1, 2, \dots$	$\mu$	$\mu$
Geometrisk fördelning	$p(k) = pq^k$	$k = 0, 1, 2, \dots$	$q/p$	$q/p^2$
ffg-fördelning	$p(k) = pq^{k-1}$	$k = 1, 2, \dots$	$1/p$	$q/p^2$
Normalfördelning, $N(\mu, \sigma)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x \in \mathbb{R}$	$\mu$	$\sigma^2$
Gammafördelning, $\Gamma(p, \lambda)$	$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$ (*)	$x \geq 0$	$p/\lambda$	$p/\lambda^2$
Exponential- fördelning, $Exp(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$x \geq 0$	$1/\lambda$	$1/\lambda^2$
Rektangel- fördelning, $R(a, b)$	$f(x) = \frac{1}{b-a}$	$a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Dubbel exponential- fördelning	$F(x) = e^{-e^{-(x-b)/a}}$ (OBS! fördelningsfunktion)	$x \in \mathbb{R},$ $a > 0$	$b + \gamma a$ (**)	$\frac{a^2 \pi^2}{6}$
Weibull- fördelning	$F(x) = 1 - e^{-\left(\frac{x-b}{a}\right)^c}$ (OBS! fördelningsfunktion)	$x \geq b,$ $a, c > 0$	$b + a\Gamma(1 + 1/c)$ $a^2 \left[ \Gamma(1 + \frac{2}{c}) - \Gamma^2(1 + \frac{1}{c}) \right]$	
Lognormal- fördelning $\ln X \in N(b, a)$	$f(x) = \frac{1}{\sqrt{x^2 a^2 2\pi}} e^{-\frac{(\ln x - b)^2}{2a^2}}$	$x \geq 0$	$e^{b+a^2/2}$	$e^{2b+2a^2} - e^{2b+a^2}$

$$(*) \quad \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \quad p > 0$$

$$\Gamma(p) = (p-1) \cdot \Gamma(p-1)$$

$$\Gamma(p) = (p-1)! \quad \text{om } p \text{ heltal}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$(**) \quad \gamma \approx 0.57722.$$

**Tvådimensionell normalfördelning**

$(X, Y)$  är tvådimensionellt normalfördelat med väntevärdesvektor  $\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$  och kovariansmatris

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \text{ om } f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} e^{-Q/2}, \quad (x,y) \in \mathbb{R}^2,$$

$$\text{där } \det(\Sigma) = \sigma_X^2\sigma_Y^2(1 - \rho^2) \text{ och } Q = \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}.$$

Den betingade fördelningen för  $X$  givet att  $Y = y$  är en endimensionell normalfördelning med

$$\mathbf{E}(X | Y = y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y),$$

$$\mathbf{V}(X | Y = y) = \sigma_X^2(1 - \rho^2).$$

**Fördelningar besläktade med normalfördelningar**

$$\chi^2\text{-fördelning} \quad X_1, \dots, X_n \in N(0, 1), \quad \text{oberoende} \Rightarrow \sum_{i=1}^n X_i^2 \in \chi^2(n).$$

$$\chi^2(f) = \Gamma(f/2, 1/2)$$

$$t\text{-fördelning, } t(f) \quad X \in N(0, 1), Y \in \chi^2(f), \quad \text{oberoende} \Rightarrow \frac{X}{\sqrt{Y/f}} \in t(f).$$

$$F\text{-fördelning, } F(f_1, f_2) \quad X \in \chi^2(f_1), Y \in \chi^2(f_2), \quad \text{oberoende} \Rightarrow \frac{X/f_1}{Y/f_2} \in F(f_1, f_2).$$

**Additionsformler**

Om  $X_1$  och  $X_2$  oberoende så gäller:

$$X_1 \in \text{Bin}(n_1, p), X_2 \in \text{Bin}(n_2, p) \Rightarrow X_1 + X_2 \in \text{Bin}(n_1 + n_2, p).$$

$$X_1 \in \text{Po}(\mu_1), X_2 \in \text{Po}(\mu_2) \Rightarrow X_1 + X_2 \in \text{Po}(\mu_1 + \mu_2).$$

$$X_1 \in \Gamma(p_1, a), X_2 \in \Gamma(p_2, a) \Rightarrow X_1 + X_2 \in \Gamma(p_1 + p_2, a).$$

$$X_1 \in \chi^2(f_1), X_2 \in \chi^2(f_2) \Rightarrow X_1 + X_2 \in \chi^2(f_1 + f_2).$$

## Statistikteori

### Punktskattningar vid normalfördelning och helt okänd fördelning

#### Ett stickprov

Låt  $x_1, \dots, x_n$  vara observationer av oberoende och likafördelade s.v.  $X_1, \dots, X_n$  med väntevärde  $\mu$  och standardavvikelse  $\sigma$ . Väntevärdesriktiga skattningar av  $\mu$  och  $\sigma^2$  är då

- $\mu^* = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$  Om  $X_i \in N(\mu, \sigma)$  så  $\mu^* \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- $(\sigma^2)^* = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$  då  $\mu$  känd. Om  $X_i \in N(\mu, \sigma)$  så  $\frac{n(\sigma^2)^*}{\sigma^2} \in \chi^2(n)$
- $(\sigma^2)^* = S^2 = \frac{Q}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  då  $\mu$  okänd. Om  $X_i \in N(\mu, \sigma)$  så  $\frac{Q}{\sigma^2} \in \chi^2(n-1)$

#### Flera stickprov

Låt  $x_{i1}, \dots, x_{in_i}$  vara ober. obs. från  $N(\mu_i, \sigma)$  då  $i = 1, \dots, k$ . Då är

- $(\sigma^2)^* = S^2 = \frac{Q}{f} = \frac{(n_1 - 1)S_1^2 + \dots + (n_k - 1)S_k^2}{(n_1 - 1) + \dots + (n_k - 1)}$  Eftersom  $X_{ij} \in N(\mu_i, \sigma)$  så  $\frac{Q}{\sigma^2} \in \chi^2(f)$

#### Vanliga skattningsmetoder

- ML-skattning:

Låt  $x_1, \dots, x_n$  vara observationer av  $X_1, \dots, X_n$ , som är oberoende s.v. med täthets- (sannolikhets-) funktion  $f(x_i; \theta)$ ,  $i = 1, \dots, n$  ( $p(x_i; \theta)$ ,  $i = 1, \dots, n$ ).

ML-skattningen av parametern  $\theta$  är det  $\theta_{ML}^*$  som maximerar likelihood-funktionen

$$L(\theta; x_1, \dots, x_n) = \begin{cases} p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta), \\ f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta). \end{cases}$$

- MK-skattning:

Låt  $x_1, \dots, x_n$  vara oberoende observationer av stokastiska variabler med  $\mathbf{E}(X_i) = \mu_i(\theta)$ , där funktionerna  $\mu_i$  är kända och parametern  $\theta$  okänd.

MK-skattningen av parametern  $\theta$  är det  $\theta_{MK}^*$  som minimerar förlustfunktionen

$$Q(\theta; x_1, \dots, x_n) = \sum_{i=1}^n (x_i - \mu_i(\theta))^2.$$

- Viktad MK-skattning:

(Förutsättningar enligt MK-skattning ovan.) Den viktade MK-skattningen av  $\theta$  är det  $\theta_{MK}^*$  som minimerar förlustfunktionen

$$Q(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \lambda_i (x_i - \mu_i(\theta))^2,$$

där  $\lambda_i$  är en följd av vikter, t.ex.  $\lambda_i = 1/\sigma_i^2$ , där  $\sigma_i^2 = \mathbf{V}(X_i)$ .

## Konfidensintervall

- Konfidensintervall med konfidensgrad  $1 - \alpha$  för väntevärdet av en normalfördelad skattning:

Om  $\theta^*(X_1, \dots, X_n) \in N(\theta, \mathbf{D}(\theta^*))$  så

- ♦  $I_\theta = (\theta^* \pm \lambda_{\alpha/2} \cdot \mathbf{D}(\theta^*))$  om  $\mathbf{D}(\theta^*)$  är känd
- ♦  $I_\theta = (\theta^* \pm t_{\alpha/2}(f) \cdot \mathbf{d}(\theta^*))$  om  $\mathbf{D}(\theta^*) = c \cdot \sigma$  där  $\sigma = \mathbf{D}(X_i)$ ,  $c$  är en konstant och  $(\sigma^2)^* = S^2 = \frac{Q}{f}$  med  $\frac{Q}{\sigma^2} \in \chi^2(f)$

- Konfidensintervall med konfidensgrad  $1 - \alpha$  för väntevärdet i en normalapproximation:

Om  $\theta^*(X_1, \dots, X_n) \in N(\theta, \mathbf{D}(\theta^*))$  enligt CGS (el. dyl.) så

- ♦  $I_\theta \approx (\theta^* \pm \lambda_{\alpha/2} \cdot \mathbf{D}(\theta^*))$  om  $\mathbf{D}(\theta^*)$  är känd
- ♦  $I_\theta \approx (\theta^* \pm \lambda_{\alpha/2} \cdot \mathbf{d}(\theta^*))$  om  $\mathbf{D}(\theta^*)$  skattas med  $\mathbf{d}(\theta^*)$
- ♦  $I_\theta \approx (\theta^* \pm t_{\alpha/2}(f) \cdot \mathbf{d}(\theta^*))$  om  $\mathbf{D}(\theta^*)$  skattas med  $\mathbf{d}(\theta^*)$  där  $\mathbf{D}(\theta^*) = c \cdot \sigma$  med  $\sigma = \mathbf{D}(X_i)$ ,  $c$  är en konstant och  $(\sigma^2)^* = S^2 = \frac{Q}{f}$

- Konfidensintervall med konfidensgrad  $1 - \alpha$  för variansen i en normalfördelning:

Om  $X_1, \dots, X_n \in N(\mu, \sigma)$  med  $(\sigma^2)^* = S^2 = \frac{Q}{f}$  och  $\frac{Q}{\sigma^2} \in \chi^2(f)$  så

$$I_{\sigma^2} = \left( \frac{f \cdot s^2}{\chi_{\alpha/2}^2(f)}, \frac{f \cdot s^2}{\chi_{1-\alpha/2}^2(f)} \right)$$

## Hypotestest

Styrkefunktion:  $h(\theta) = \mathbf{P}(H_0 \text{ förkastas} \mid \theta \text{ är det rätta parametervärdet})$

Speciellt: Signifikansnivån,  $\alpha = \mathbf{P}(H_0 \text{ förkastas} \mid H_0 \text{ sann})$

## Regression

### Enkel linjär regression

- Modell:  $y_i = \alpha + \beta x_i + \varepsilon_i$ ,  $\varepsilon_i \in N(0, \sigma)$ .

- Parameterskattningar:

$$\alpha^* = \bar{y} - \beta^* \bar{x}, \quad \beta^* = \frac{S_{xy}}{S_{xx}}, \quad (\sigma^2)^* = s^2 = \frac{Q_0}{n-2}, \quad Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

$$\alpha^* \in N\left(\alpha, \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right), \quad \beta^* \in N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right), \quad \frac{Q_0}{\sigma^2} \in \chi^2(n-2),$$

$$\mathbf{C}(\alpha^*, \beta^*) = -\bar{x} \cdot \frac{\sigma^2}{S_{xx}}, \quad \mu_0^* = \alpha^* + \beta^* x_0 \in N\left(\alpha + \beta x_0, \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right).$$

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



- Ett prediktionsintervall med konfidensgrad  $1 - p$  för  $y(x_0) = \alpha + \beta x_0 + \varepsilon_0$  ges av

$$I_{y(x_0)} = \left( \alpha^* + \beta^* x_0 \pm t_{p/2}(n-2) \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

- Ett kalibreringsintervall med konfidensgrad  $1 - p$  för  $x_0 = \frac{y_0 - \alpha}{\beta}$  ges av

$$I_{x_0} = \left( x_0^* \pm t_{p/2}(n-2) \cdot \frac{s}{\beta^*} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0^* - \bar{x})^2}{S_{xx}}} \right) \quad \text{där} \quad x_0^* = \frac{y_0 - \alpha^*}{\beta^*}$$

### Multipel linjär regression

- Modell:  $y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_p x_i^{(p)} + \varepsilon_i$ ,  $\varepsilon_i \in N(0, \sigma)$ .
- Med matrisrepresentation kan modellen skrivas  $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}$  med

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1^{(1)} & \dots & x_1^{(p)} \\ 1 & x_2^{(1)} & \dots & x_2^{(p)} \\ \vdots & \vdots & & \vdots \\ 1 & x_n^{(1)} & \dots & x_n^{(p)} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Parameterskattningar:

$$\boldsymbol{\beta}^* = (X^T X)^{-1} X^T \mathbf{y}, \quad \mathbf{V}(\boldsymbol{\beta}^*) = \sigma^2 (X^T X)^{-1},$$

$$(\sigma^2)^* = s^2 = \frac{Q_0}{n - (p + 1)}, \quad Q_0 = \mathbf{y}^T \mathbf{y} - \boldsymbol{\beta}^{*T} X^T \mathbf{y}, \quad \frac{Q_0}{\sigma^2} \in \chi^2(n - (p + 1)),$$

$$\beta_i^* \in N\left(\beta_i, \sigma \sqrt{\text{element}(ii) \text{ i } (X^T X)^{-1}}\right),$$

$$\mu_0^* = \mathbf{x}_0 \boldsymbol{\beta}^* \in N\left(\mu_0, \sigma \sqrt{\mathbf{x}_0 (X^T X)^{-1} \mathbf{x}_0^T}\right) \quad \text{där} \quad \mathbf{x}_0 = \left( 1 \quad x_0^{(1)} \quad \dots \quad x_0^{(p)} \right)$$