Maximum likelihood-metoden Skattningarna $(\theta_1)^*_{\text{obs}}, \ldots, (\theta_r)^*_{\text{obs}}$ är de värden på $\theta_1, \ldots, \theta_r$ som maximerar *trolighetsfunktionen*

$$L(\theta_1, \ldots, \theta_r) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

vid kontinuerliga fördelningar, eller

$$L(\theta_1, \ldots, \theta_r) = \mathsf{P}(X_1 = x_1, \ldots, X_n = x_n)$$

vid diskreta.

Filosofin bakom Maximum likelihood-metoden är att vårt utfall (stickprovet) bestäms av parametrarna i de ingående fördelningarna. Skattningen av dessa parametrar är de värden som gör det observerade så troligt som möjligt.

Tips! Ofta är det enklare att maximera $\log L(\theta_1, \ldots, \theta_r)$.

Exempel: Binomialfördelningen. Låt x vara ett utfall av en binomialfördelad stokastisk variabel X.

$$L(p) = \mathsf{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Logaritmering ger

$$\ln L(p) = \ln\left(\binom{n}{x}\right) + x \ln p + (n-x)\ln(1-p).$$

Derivering bestämmer maximum:

$$\frac{d}{dp} \ln L(p) = 0 + x\frac{1}{p} - (n-x)\frac{1}{1-p} = 0$$

ger lösningen $p = x/n$. Alltså, ML-skattningen av p är $p^*_{\text{obs}} = x/n$.

Minsta kvadrat-metoden Låt $x_1, \ldots, x_n$ vara utfall av oberoende stokastiska variabler $X_1, \ldots, X_n$ (med samma varians). Skattningarna $(\theta_1)^*_{\text{obs}}, \ldots, (\theta_r)^*_{\text{obs}}$ de värden på $\theta_1, \ldots, \theta_r$ som minimerar *minsta kvadrat-avstånden* mellan observationer och väntevärden:

$$Q(\theta_1, \ldots, \theta_r) = \sum_{i=1}^{n} (x_i - \mathsf{E}[X_i])^2.$$

Även här minimeras Q oftast med hjälp av derivering.

Exempel: Låt x vara ett utfall av en binomialfördelad stokastisk variabel X. MK-metoden: finn p som minimerar

$$Q(p) = (x - \mathsf{E}[X])^2 = (x - np)^2.$$

Eftersom $Q(p) \geq 0$ med likhet om $p = x/n$ är MK-skattningen av p, $p^*_{\text{obs}} = x/n$.

Konfidensintervall

Definition: Ett konfidensintervall $I_\theta(x_1, \ldots, x_n)$ för θ är ett utfall av ett stokastiskt intervall $I_\theta(X_1, \ldots, X_n)$ sådant att

$$\mathsf{P}(\theta \in I_\theta(X_1, \ldots, X_n)) = 1 - \alpha = \text{konfidensgraden}.$$

Låt θ^*_{obs} vara en skattning av θ sådan att θ^* är $\mathrm{N}(\theta, D)$. Då kan vi ur $\mathrm{N}(0,1)$-tabeller bestämma kvantiler $\lambda_{\alpha/2}$ så att med sannolikhet $1 - \alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\theta^* - \theta}{D} \leq \lambda_{\alpha/2}$$

eller

$$-\lambda_{\alpha/2}D \le \theta^* - \theta \le \lambda_{\alpha/2}D$$

eller

$$\theta^* - \lambda_{\alpha/2}D \le \theta \le \theta^* + \lambda_{\alpha/2}D.$$

Det vill säga

$$\theta \in \theta^*_{\text{obs}} \pm \lambda_{\alpha/2}D$$

är ett *konfindensintervall* för θ med konfidensgrad $1-\alpha$.

Om θ^* är approximativt $\mathrm{N}(\theta, D)$ där $D = \mathrm{D}(\theta^*)$ inte är känd får den skattas och

$$\theta \in \theta^*_{\text{obs}} \pm \lambda_{\alpha/2}d(\theta^*)$$

är ett (approximativt) konfindensintervall för θ med konfidensgrad $\approx 1-\alpha$.

Exempel: Låt $x_1, \ldots, x_n$ vara utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$.

Väntevärdet μ skattas med $\overline{x}$ där

$$\overline{X} \text{ är } \mathrm{N}(\mu, \sigma/\sqrt{n})$$

Ett konfidensintervall för μ är

$$\mu \in \overline{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (1-\alpha)$$

Exempel: Opinionsundersökning. Låt x vara antalet sympatisörer av n intervjuade. Vi modellerar x som ett utfall av en binomialfördelad stokastisk variabel X, X är $\mathrm{Bin}(n, p)$.

Då är $p^*_{\text{obs}} = x/n$ och $\mathrm{D}(p^*) = \sqrt{p(1-p)/n}$. Om p^* är approximativt $\mathrm{N}(p, \mathrm{D}(p^*))$ så fås ($\approx 95\%$)-konfidensintervall

Parti	Antal, x	Andel [%], p^*_{obs}	Medelfel [%], $d(p^*)$	$p^*_{\text{obs}} \pm 1.96d(p^*)$
(S)	668	35.0	1.1	(32.9, 37.2)
(V)	139	7.3	0.6	(6.1, 8.5)
(Mp)	109	5.7	0.5	(4.7, 6.8)
(M)	463	24.3	1.0	(22.4, 26.2)
(C)	126	6.6	0.6	(5.5, 7.7)
(Fp)	244	12.8	0.8	(11.3, 14.3)
(Kd)	101	5.3	0.5	(4.3, 6.3)
Övr	57	3.0	0.4	(2.2, 3.8)
	1907			

Normalapproximationen motiveras av att skattningen av

$$\mathrm{V}(X) = np(1-p)$$

alla är mycket större än 10.

Exakta konfidensintervall i normalfördelningen

Låt $x_1, \ldots, x_n$ vara utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$.

Fördelningen för

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} Z_i^2$$

är fördelningen för en summa av kvadrater av oberoende $\mathrm{N}(0,1)$-variabler, och beror inte på μ eller σ. Fördelningen kallas χ^2-fördelningen och parametern, antalet termer i summan, kallas dess frihetsgrader. Fördelningen finns tabulerad i formelsamlingen.

Sats. Man kan visa att

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

är $\chi^2(n-1)$-fördelad (och oberoende av $\overline{X}$).

Vi kan således för varje sannolikhet α bestämma kvantilerna $\chi^2_{1-\alpha/2}$ och $\chi^2_{\alpha/2}$ så att

$$\mathsf{P}\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1-\alpha.$$

Så med sannolikhet $1-\alpha$ är

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

eller

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}.$$

Detta är ger att

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

är ett [symmetriskt] konfidensintervall för σ^2 med konfidensgrad $1-\alpha$. Ett intervall för σ är

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

Vidare så är

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \qquad \mathrm{N}(0,1)$$

och

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \text{ är } t(n-1)\text{-fördelad,}$$

alltså, storheten har en fördelning som inte beror på μ eller σ utan bara på antalet observationer n. Utnyttjar vi tabellerna för t-fördelningen kan vi för varje sannolikhet α bestämma kvantiler $t_{\alpha/2}$ så att med sannolikhet $1-\alpha$ är

$$-t_{\alpha/2} \le \frac{\overline{X}-\mu}{S/\sqrt{n}} \le t_{\alpha/2}$$

eller

$$-t_{\alpha/2}\frac{S}{\sqrt{n}} \le \overline{X}-\mu \le t_{\alpha/2}\frac{S}{\sqrt{n}}$$

eller

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}.$$

Alltså, ett [symmetriskt] konfidensintervall för μ med konfidensgrad $1-\alpha$ är

$$\mu \in \overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}.$$