1. Tyhjentävä tunnusluku (sufficient statistics)

Olkoon $(P(X = x|\theta) : \theta \in \Theta)$ todennäköisyysmalli havainnolle X. Datan funktio $T(X)$ on Tyhjentävä tunnusluku jos ehdollinen todennäköisyys (ehdollinen tiheysfunktio)

$$P(X = x|T(X) = t, \theta) = p(X = x|T(X) = t)$$

ei riipu tuntemattomasta θ:sta

Silloin uskottavuusfunktio faktorisoituu

$$\begin{aligned} &P(X = x|\theta) = \\ &P(X = x|T(X) = T(x))P(T(X) = T(x)|\theta) \end{aligned}$$

Esimerkki: Heitetaan kolikkoa n-kertaa, olkoon $Y_i = \mathbf{1}(\ i\text{-tulos on kruunu }\ )$. Oletamme että kolikon heitot ovar riippumattomia ja samoin jakautuneita (independent and uniformly distributed, i.i.d), tuntemattomalla kruunun todennäköisyydellä $\theta \in [0,1]$.

$$P(Y_1, \dots, Y_n | \theta) = \prod_{i=1}^{n} P(Y_i|\theta) =$$

$$\prod_{i=1}^{n} \theta^{Y_i}(1-\theta)^{(1-Y_i)} = \theta^{X_n}(1-\theta)^{n-X_n}$$

jossa $X_n = T(Y_1, \dots, Y_n) = (Y_1 + \dots + Y_n)$ on tyhjentävä tunnusluku.

$$P(X_n = x|\theta) = \sum_{y_i \in \{0,1\}: \sum_{i=1}^n y_i = x} P(Y_1 = y_1, \dots, Y_n = y_n|\theta)$$

$$= \theta^x (1-\theta)^{n-x} \# \{ y \in \{0,1\}^n : \sum_{i=1}^n y_i = x \}$$

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

Tyhjentävyydenperiaate Olkoon Y data ja $T(Y)$ tyhjentävä tunnusluku. Olkoon y ja y' havaintojen arvoja joilla $T(y) = T(y')$. Tilastollinen päättely θ:sta datan $\{Y = y\}$ ja $\{Y = y\}$ perustella johtaa samoihin tuloksiin.
Sekä freqventistisessa että Bayeslaisessa ajattelutavassa hyväksytään tyhjentävyyden periaatetta.

Uskottavuuden periaate (Likelihood principle)

Olkoon $X \sim p(x|\theta)$ ja $\widetilde{X} \sim \widetilde{p}(\widetilde{x}|\theta)$, kaksi kokeetta samoilla tuntematomalla parametrilla θ.
Jos joillakin kokeiden tulokseilla $\{ X = x\}$ ja $\{\widetilde{X} = \widetilde{x}\}$

$$p(X = x|\theta) = C\widetilde{p}(\widetilde{X} = \widetilde{x}|\theta) \quad \forall\theta$$

jossa $C = C(x, \widetilde{x})$ ei riipu θ:sta, silloin kokeiden tulokset $\{ X = x\}$ ja $\{\widetilde{X} = \widetilde{x}\}$ ovat yhtä informatiivisia θ:stä ja johtavat samaan päättelyyn.

Uskottavuusfunktio $\theta \mapsto p(X = x|\theta)$ modulo normalisointivakiota sisältää koko havainnon $\{X = x\}$:n informaatiota tuntemattomasta parametrista θ.
Esimerkiksi, suurimman uskottavuuden (maximum likelihood, ML) estimaattori noudattaa uskottavuuden periateetta:

$$\begin{aligned}
\widehat{\theta}_{ML} &= \arg\max_{\theta} p(X = x|\theta) p(\widetilde{X} = \widetilde{x}|\theta) \\
&= \arg\max_{\theta} C(x, \widetilde{x}) p(\widetilde{X} = \widetilde{x}|\theta) \\
&= \arg\max_{\theta} p(\widetilde{X} = \widetilde{x}|\theta)
\end{aligned}$$

Kuitenkin, kun suurimman uskottavuudeen estimattoriin liitetään frekventistiset luottamusvälit, rikotaan uskottavauuden periaatetta.

Esimerkki

Koe I:
Heitetaan kolikkoa n kertaa, jossa kruunun todennäköisyys on tuntematon parametri $\theta \in [0,1]$. Olkon satunnaismuuttuja $X_n =$ kruunujen määrä Uskottavuusfunktio on

$$p(X_n = x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Esimerkki

Koe I:

Heitetaan kolikkoa n kertaa, jossa kruunun todennäköisyys on tuntematon parametri $\theta \in [0,1]$. Olkon satunnaismuuttuja $X_n =$ kruunujen määrä Uskottavuusfunktio on

$$p(X_n = x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Koe II:

heitetaan samaa kolikkoa jolla on edelleen tuntematon todennäköissys parametri $\theta \in [0,1]$, niin kauan kun saadan x-kruunuja. Olkoon satunnaismuuttuja $N_x =$ heittojen määrä . N_x on x-geometristen satunnaismuuttujen muuttujen konvoluutio, negatiivinen binomialinen jakauma. x on otoskoko ja uskottavuusfunktio on

$$\widetilde{P}(N_x = n|\theta) = \binom{n-1}{x-1} \theta^x (1-\theta)^{n-x}$$

Tämä seuraa koska kun $x = 1$, N_1 on geometrisesti jakautunut parametrilla θ,

$$\widetilde{P}(N_1 = n_1|\theta) = (1-\theta)^{(n_1-1)}\theta$$

ja kun $N_0 = 0$

$$\widetilde{P}(N_1 = n_1, N_2 - N_1 = n_2, \dots N_x - N_{x-1}|\theta) =$$
$$\prod_{i=1}^{x} \widetilde{P}(N_i - N_{i-1} = n_i|\theta) =$$
$$(1-\theta)^{(n_1-1)}\theta \times (1-\theta)^{(n_2-1)}\theta \times (1-\theta)^{(n_x-1)}\theta$$
$$= (1-\theta)^{\sum_{i=1}^{x}(n_i-1)}\theta^{\sum_{i=1}^{x} n_i}$$
$$(1-\theta)^{(n-x)}\theta^{x}$$

jossa $n = n_1 + n_2 + \cdots + n_x$.

Koska $N_x = \sum_{i=1}^{x}(N_i - N_{i-1})$,

$$\widetilde{P}(N_x = n|\theta) =$$

$$\sum_{n_i:\sum_{i=1}^{x} n_i = n} P(N_1 = n_1, N_2 - N_1 = n_2, \dots N_x - N_{x-1}|\theta)$$

$$(1-\theta)^{(n-x)}\theta^x \#\{(n_1,\dots,n_x) : \sum_{i=1}^{x} n_i = n\}$$

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

Oletetaan että kokeessa I) olemme saaneet $X_n = x$, jossa n on otoskoko, ja kokeessa II) olemme saaneet $N_x = n$ jossa x. Havainnot $\{X_n = x\}$ kokeessa (I) ja $\{N_x = n\}$ kokeessa (II) sisältävät samaa informaatiota θ:sta ja kun noudataan uskottavuus-periaatetta joudutaan samaan päättelyyn.

Tässä tapauksessa Suurimman uskottavuuden estimaattorit ovat $(I): \ \widehat{\theta}_n = \frac{X_n}{n} = \frac{x}{n}$, ja $(II): \ \widehat{\theta}_x = \frac{x}{N_x} = \frac{x}{n}$. molemmissa tapauksissa.

Laskemme *ML*-estimattorien varianssieja:

$$I): \ E_{\theta_0}(X_n) = \theta_0 n, \quad \mathrm{Var}_{\theta_0}(X_n) = \frac{\theta_0(1-\theta_0)}{n}$$

Asymptottisesti, suurten lukujen laista

$$\widehat{\theta}_n = \frac{X_n}{n} \to \theta_0 \quad \text{kun } n \to \infty$$

P_{ϑ_0}-todennäköisyydellä 1, ja
ja keskeisesta raja-arvolauseesta

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \theta_0(1-\theta_0))$$

jakauman konvergenssin mielessä.

$$II)\ E_{\theta_0}(N_x) = \frac{x}{\theta_0}, \quad \mathrm{Var}_{\theta_0}(N_x) = x\frac{(1-\theta)}{\theta^2}$$

Suurten lukujen laista seuraa

$$\frac{N_x}{x} \longrightarrow \frac{1}{\theta_0} \quad \text{kun } x \to \infty$$

$\widetilde{P}_{\vartheta_0}$-todennäköisyydellä 1, ja keskeisesta raja-arvolauseesta

$$\sqrt{x}\left(\frac{N_x}{x} - \frac{1}{\theta_0}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1-\theta}{\theta^2}\right) \quad \text{kun } x \to \infty$$

jakauman mielessä.

Palautan mieleen delta-menetelmaa: jos

$$\sqrt{x}(\widehat{\psi}_x - \psi_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

ja $\psi \mapsto \theta = f(\psi)$ jossa f on sileä, seuraa estimattorille $\widehat{\theta}_x := f(\widehat{\psi}_x)$

$$\sqrt{x}(\widehat{\theta}_x - \theta_0) = \sqrt{x}(f(\widehat{\psi}_x) - f(\psi_0)) \xrightarrow{d} \mathcal{N}(0, f'(\psi_0)^2\sigma^2)$$

jossa $\theta_0 := f(\psi_0)$

Kun $f(x) = 1/x$, Delta menetelmällä seuraa

$$\sqrt{x}(f(N_x/x) - f(1/\theta_0)) = \sqrt{x}(x/N_x - \theta_0) =$$
$$\sqrt{x}(\widehat{\theta}_x - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, f'(\psi_0)^2 \frac{(1-\theta_0)}{\theta_0^2}\right)$$
$$= \mathcal{N}\left(0, (1-\theta_0)\theta_0^2\right)$$

Kun otoskoot n ja x ovat tarpeeksi suureet, saadan freventistiset α-luottamustason luottamusvälit
Koe I): Kun $X_n = x$, $\widehat{\theta}_n = x/n$

$$\theta_0 \sim \hat{\theta}_n \pm q_\alpha \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}$$

Koe II): Kun $N_x = n$ $\widehat{\theta}_x = x/n$

$$\theta_0 \sim \hat{\theta}_n \pm q_\alpha \sqrt{\frac{(1-\widehat{\theta}_x)\widehat{\theta}_x^2}{x}}$$

Vaikka suurimman uskottavuuden estimointi on sopusoinnissa uskottavuus periaatteen kanssa, freqventistiset luottamusvälit eivät ole !

Bayesin ratkaisu

Oletamme tasainen priori $p(\theta) = 1$ kun $\theta \in [0,1]$, $p(\theta) = 0$ muuten

Posteriori jakauma on

Koe I:

$$p(\theta|X_n = x) = \frac{p(X_n = x|\theta)p(\theta)}{\int_0^1 p(X_n = x|\vartheta)p(\vartheta)d\vartheta}$$
$$= \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\binom{n}{x}\int_0^1 \vartheta^x(1-\vartheta)^{n-x}d\vartheta}$$
$$= \frac{1}{\mathrm{Beta}(x+1, n-x+1)}\theta^x(1-\theta)^{n-x}$$

jossa

$$\mathrm{Beta}(a,b) := \int_0^1 \vartheta^{a-1}(1-\vartheta)^{b-1}d\vartheta$$

on Beta funktio. Siis a posteriori ehdolla $X_n = x$ θ on

Koe II:

$$
\begin{aligned}
p(\theta| \ N_x = n) &= p(\theta|N_x = n) \\
&= \frac{p(N_x = n|\theta)p(\theta)}{\int_0^1 p(N_x = n|\vartheta)p(\vartheta)d\vartheta} \\
&= \frac{\binom{n-1}{x-1}\theta^x(1-\theta)^{n-x}}{\binom{n-1}{x-1}\int_0^1 \vartheta^x(1-\vartheta)^{n-x}d\vartheta} \\
&= \frac{1}{\mathrm{Beta}(x+1, n-x+1)}\theta^x(1-\theta)^{n-x}
\end{aligned}
$$

on θ:n ehdollinen jakauma ehdolla $N_x = n$.
Koska Bayesin kaavassa käytetään vain uskottavuusfunktion arvo data pisteissä, seuraa että Bayeslainen tilastotiede on sopusoinnissa uskottavuusperiaatteen kanssa.

Yleisemmin, oletetaan että heitetaan kolikkoita tietyn satunnaisen pysähdysstrategia mukaan:
Kokeensunnitelija määrää jonoa pysähdystodennäköisyyksiä $q_k(Y_1, \ldots, Y_k) : \{0,1\}^k \to [0,1]$, jossa $k \in \mathbb{N}$. Koska θ on tuntematon koesuunnitelijalle, pysähdystodennäköisyydet eivät riipu parametrista θ.
Jos koe on jatkanut k-heiton asti, todennäköisyydellä $q_k(Y_1, \ldots, Y_k)$ datan kerääminen loppuu siihen. Myös alussa todennäköisyydella q_0 päätetään kerätäänkö ollenkaan dataa. Jos data ei kerätä, $N = 0$
Koska θ on tuntematon koesuunnitelijalle määrää pysähdys-todennäköisyydet $q_k(Y_1, \ldots, Y_k)$.

Silloin data $= (N, Y_1, \ldots, Y_N)$, jossa N on satunnainen, uskottavuusfunktiolla

$$P(N = n, Y_1 = y_1, \ldots, Y_n = y_n|\theta) = q_0 \quad \text{jos } n = 0, \text{ muuten}$$
$$= \left\{(1 - q_0)\left(\prod_{k=1}^{n-1}(1 - q_k(y_1, \ldots, y_k))\right) \times q_n(y_1, \ldots, y_n)\right\} \times$$
$$\times \left\{\prod_{i=1}^{n} \theta^{y_i}(1 - \theta)^{1-y_i}\right\}$$
$$= C(n, y_1, \ldots, y_n)\theta^{x_n}(1 - \theta)^{n-x_n}$$

jossa $x_n = y_1 + \cdots + y_n$. Huomataan satunnaispari (N, X_N) on tyhjentävä tunnusluku. Kun noudataan uskottavuuden periaatteetta tilastollinen päättely riippuu vain kerätystä datasta, eikä datan-keräämisen strategiasta

Edellisessä esimerkissa käytettiin pysähdysstrategiat $q_k(y_1,\dots,y_k) = \mathbf{1}(k = n)$ ja vastaavasti $q_k(y_1,\dots,y_k) = \mathbf{1}(y_1 + \dots + y_k = x)$ jossa n, ja vastaavasti x kiinnitettiin. Frekventtistiset luottamusvälit ja p-arvot riippuvat ei pelkään kerätystä datasta mutta myös etukäteen sovitusta pysähdysstrategiasta.
Miten jos jostakin syystä joudutaan poikkeamaan etukäteen sovitusta pysähdysstrategiasta ? Silloin uusi pysähdystodennäköisyys on tuntematon, ja ei pysty enää laskemaan ML-estimattorin varianssi, freqventistiset luottamusvälit ja p-arvot.

Lääketieteellisessä tutkimuksessa tämä on aiheuttanut ja aiheuttaa (aivan turhaan, jos hyväksytään uskottavuuden periaattetta) eettisia kysymyksiä: kun testataan uusia lääkkeitä, ja joudutaan poikeaamaan protokollasta joidenkin potilaiden suhteen, koko koetutkimuksen aineisto sattaa tulla frekventisen tilastotieteen näkökulmasta kelvottamaksi, koska p-arvot ei pysty enää laskemaan. Toisaalta, lääkärit ovat velvollisia antamaan yksittäisille potilaille jokaisessa vaiheessa parhasta mahdollista hoitoa protokollan huolimatta.

Delta-menetelmästä

Delta menetelmä perustuu Taylorin kaavaan:

Kun $\widehat{\theta}_n \stackrel{P_{\theta_0}}{\to} \theta_0$ kun $n \uparrow \infty$ $\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \stackrel{d}{\to} \mathcal{N}\big(0, \sigma^2\big)$, jos funktio f on derivoituva $\widehat{f}(\theta_n) \stackrel{P_{\theta_0}}{\to} f(\theta_0)$

$$
\begin{aligned}
&\sqrt{n}(f(\hat{\theta}_n) - f(\theta_0)) = \\
&f'(\theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0) \\
&+n^{-1/2}\frac{1}{2}f''(\theta_n^*)\big(\sqrt{n}(\widehat{\theta}_n - \theta_0)\big)^2 \\
&\stackrel{d}{\to} \mathcal{N}\big(0, f'(\theta_0)^2\sigma^2\big)
\end{aligned}
$$

Kun $n \uparrow \infty$, $\theta_n^* \stackrel{P_{\theta_0}}{\to} \theta_0$ ja toisen asteen jäännöstermi suppenee kohti nollaan P_{θ_0}:n mielessä.