Föreläsning 2: Punktskattningar

Johan Thim (johan.thim@liu.se)

27 augusti 2018

1 Repetition

Stickprov

Definition. Låt de stokastiska variablerna X_1, X_2,..., X_n vara oberoende och ha samma fördelningsfunktion F. Ett stickprov x_1, x_2,..., x_n består av observationer av variablerna X_1, X_2,..., X_n. Vi säger att *stickprovsstorleken* är n.

Punktskattning

Definition. En *punktskattning* $\widehat{\theta}$ av parametern θ är en funktion av de observerade värdena $x_1, x_2, \ldots, x_n$:

$$\widehat{\theta} = g(x_1, x_2, \ldots, x_n).$$

Vi definierar motsvarande *stickprovsvariabel* $\widehat{\Theta}$ enligt

$$\widehat{\Theta} = g(X_1, X_2, \ldots, X_n).$$

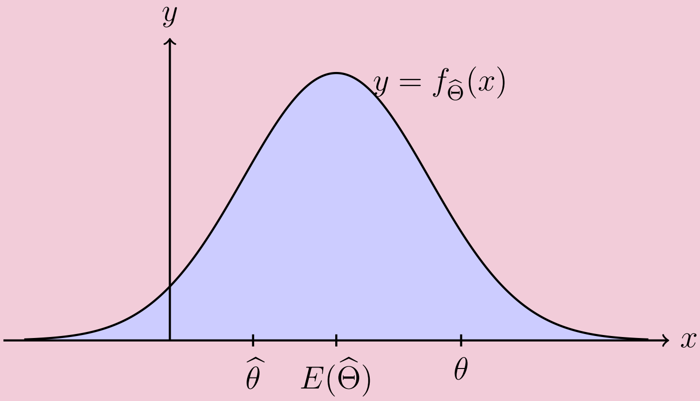
Stokastiskt eller ej?

Var noggran med att tydligt visa och göra skillnad på vad som är stokastiskt eller inte i din redovisning! Vi har tre storheter:

(i) θ – verkligt värde. Okänt. Deterministiskt.

(ii) $\widehat{\theta}$ – skattat värde. Känt (beräknat från stickprovet). Deterministiskt.

(iii) $\widehat{\Theta}$ – stickprovsvariabeln. Denna är stokastisk!



Sannolikheter som beräknas bör använda sig av $\widehat{\Theta}$ då $\widehat{\Theta}$ beskriver variationen hos $\widehat{\theta}$ för olika stickprov. Om bara $\widehat{\theta}$ och θ ingår är sannolikheten alltid noll eller ett (varför?).

Egenskaper för skattningar

Definition. Stickprovsvariabeln $\widehat{\Theta}$ kallas

(i) *väntevärdesriktig* (vvr) om $E(\widehat{\Theta}) = \theta$;

(ii) *konsistent* om det för varje $\epsilon > 0$ gäller att

$$\lim_{n\to\infty} P(|\widehat{\Theta}_n - \theta| > \epsilon) = 0,$$

där $\widehat{\Theta}_n$ är punktskattningen för varje stickprovsstorlek n;

(iii) *effektivare* än en skattning Θ^* om $V(\widehat{\Theta}) \leq V(\Theta^*)$.

Om en punktskattning inte är väntevärdesriktig pratar man ibland om ett systematiskt fel. Vi definierar detta som skillnaden $E(\widehat{\Theta}) - \theta$. En väntevärdesriktig skattning $\widehat{\Theta}$ har alltså inget systematiskt fel; i ”medel” kommer den att hamna rätt (tänk på de stora talens lag). Vi vill också gärna ha egenskapen att en punktskattning blir bättre ju större stickprov vi använder.

2 Vanliga punktskattningar

Vi stötte på medelvärdet och stickprovsvariansen på föregående föreläsning. Dessa skattningar är vettiga skattningar av väntevärdet och variansen i meningen att de är väntevärdesriktiga och konsistenta.

Medelvärde

Medelvärdet $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ är en väntevärdesriktig och konsistent skattning av väntevärdet.

Bevis: Variablerna X_k är oberoende och likafördelade. Låt $E(X_i) = \mu$ och $V(X_i) = \sigma^2$ för alla i. Eftersom väntevärdesoperatorn är linjär så gäller att

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu.$$

Alltså är $\overline{X}$ en väntevärdesriktig skattning av μ.
Då variablerna är oberoende kan vi göra en liknande kalkyl för variansen:

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Här ser vi att $V(\overline{X}) \to 0$ då $n \to \infty$, så enligt satsen ovan är skattningen konsistent.

Stickprovsvarians

Stickprovsvariansen $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ är en väntevärdesriktig skattning av variansen.

Bevis: Detta bevis är lite bökigare, men följer samma princip.

$$
\begin{aligned}
E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) &= \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)\\
&= \frac{1}{n-1}\sum_{i=1}^{n}\big(E(X_i^2) - 2E(X_i\overline{X}) + E(\overline{X}^2)\big).
\end{aligned}
$$

Vi vet att $E(\overline{X}) = \mu$ och att $V(\overline{X}) = \sigma^2/n$. Steiners formel säger att $E(Y^2) = V(Y) + E(Y)^2$ för en stokastisk variabel Y, vilket vi kan utnyttja för att skriva

$$
E(X_i^2) = V(X_i) + E(X_i)^2 = \sigma^2 + \mu^2 \quad \text{samt} \quad E(\overline{X}^2) = \sigma^2/n + \mu^2.
$$

Vidare så ser vi att

$$
E(X_i\overline{X}) = E\left(X_i\frac{1}{n}\sum_{k=1}^{n}X_k\right) = \frac{1}{n}\sum_{k=1}^{n}E(X_iX_k)
$$

och eftersom $E(X_iX_k) = E(X_i)E(X_k) = \mu^2$ om $i \neq k$ (eftersom dessa variabler är oberoende) och $E(X_i^2) = \sigma^2 + \mu^2$ (då $i = k$) kan vi skriva

$$
E(X_i\overline{X}) = ((n-1)\mu^2 + \sigma^2 + \mu^2)/n = \mu^2 + \sigma^2/n.
$$

Vi återgår till det sökta väntevärdet:

$$
E(S^2) = \frac{1}{n-1}\sum_{i=1}^{n}\big(\sigma^2 + \mu^2 - 2(\mu^2 + \sigma^2/n) + \sigma^2/n + \mu^2\big) = \frac{n\sigma^2 - n\sigma^2/n}{n-1} = \sigma^2.
$$

Alltså är S^2 en väntevärdesriktig skattning (av σ^2). Värt att notera är att $S = \sqrt{S^2}$ **inte** är en väntevärdesriktig skattning av σ (men den används oftast ändå!).

3 Metoder för att hitta punktskattningar

Vi har slarvat lite i definitionen av punktskattningar när det gäller vilka värden på den okända parametern θ som är tillåtna. Vi inför begreppet **parameterrum**.

Parameterrum

Definition. Vi låter Ω_θ beteckna **parameterrummet** av alla tillåtna värden på parametern θ.

Parameterrummet är alltså en delmängd av $\mathbf{R}^p$ där p är antalet parametrar (tänk på att $\boldsymbol{\theta}$ kan vara en vektor $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$).

Exempel

(i) Om $X \sim N(\mu, \sigma^2)$ kan vi tänka oss $\boldsymbol{\theta} = (\mu, \sigma^2)$, i vilket fall parameterrummet kan representeras som $\mathbf{R} \times (0, \infty)$.

(ii) Om $X \sim \text{Bin}(n, p)$ där n är fixerad är parameterrummet $\Omega_p = [0, 1]$.

Skulle vi med någon metod hitta en skattning som faller utanför parameterrummet måste den förkastas. Så åter till frågan hur vi hittar skattningar mer systematiskt.

3.1 Momentmetoden

Vi såg momentmetoden i förra föreläsningen. Låt oss endast repetera vad den gick ut på.

Momentskattning med flera parametrar

Definition. Låt $X \sim F(x\,; \theta_1, \theta_2, \ldots, \theta_j)$ bero på j okända parametrar $\theta_1, \theta_2, \ldots, \theta_j$ och definiera $m_i(\theta_1, \theta_2, \ldots, \theta_j) := E(X^i)$, $i = 1, 2, \ldots$ Momentskattningarna för θ_k, $k = 1, 2, \ldots, j$, ges av lösningen till ekvationssystemet

$$m_i(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_j) = \frac{1}{n}\sum_{k=1}^{n} x_k^i, \quad i = 1, 2, \ldots, j.$$

3.2 MK-skattning

Minsta kvadrat-metoden har vi egentligen stött på i tidigare kurser, mer specifikt när vi hittade approximativa lösningar till överbestämda ekvationssystem. Faktum är att vi kommer att upprepa den proceduren senare i denna kurs i sammband med linjär regression.

Låt $x_1, x_2, \ldots, x_n$ vara observationer av oberoende stokastiska variabler $X_1, X_2, \ldots, X_n$ sådana att $E(X_k) = \mu_k(\boldsymbol{\theta})$ och $V(X_k) = \sigma^2$ för $k = 1, 2, \ldots, n$ (alltså samma varians men potentiellt olika väntevärden).

Minsta kvadrat-skattning

Definition. Minsta kvadrat-skattningen för $\boldsymbol{\theta}$ ges av den vektor $\widehat{\boldsymbol{\theta}}$ som minimerar

$$Q(\widehat{\boldsymbol{\theta}}) = \sum_{k=1}^{n} \Big(x_k - \mu_k(\widehat{\boldsymbol{\theta}})\Big)^2.$$

Exempel

Låt $X_1, \ldots, X_n$ vara ett slumpmässigt stickprov från en fördelning F. Hitta MK-skattningen för väntevärdet μ.

Lösning. Vi ställer upp funktionen

$$Q(\mu) = \sum_{k=1}^{n} (x_k - \mu)^2, \mu \in \mathbf{R}.$$

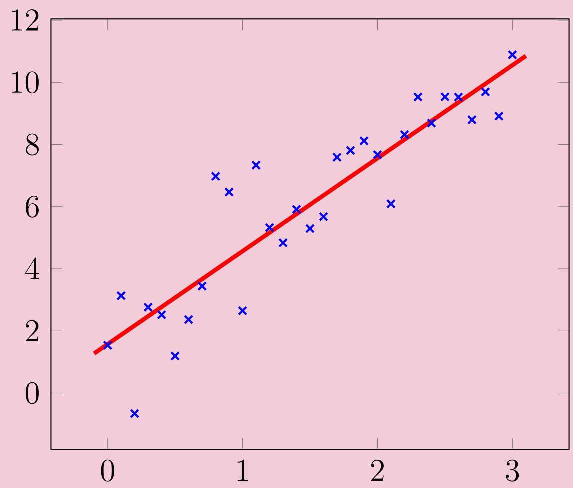
Vi söker nu det värde $\widehat{\mu}$ som minimerar Q. Enklast är att ta till envariabelanalysen och derivera och söka efter stationära punkter:

$$0 = Q'(\mu) = -2\sum_{k=1}^{n}(x_k - \mu) \quad \Leftrightarrow \quad n\mu = \sum_{k=1}^{n} x_k \quad \Leftrightarrow \quad \mu = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}.$$

Är detta ett minimum? Eftersom $Q''(\overline{x}) = 2n > 0$ är det mycket riktigt ett minimum. Den eftersökta MK-skattningen av väntevärdet är alltså $\widehat{\mu} = \overline{x}$.

Enkel linjär regression

Antag att vi gjort mätningar y_k på något vid vissa värden x_k, $k = 1, 2, \ldots, n$ och att ett spridningsdiagram visar något i stil med figuren till höger. Det förefaller rimligt att det föreligger ett approximativt linjärt samband. Kan vi hitta en linje som passar in i mätserien? vi söker alltså en linje $y = \beta_0 + \beta_1 x$ som i någon mening approximerar mätresultaten. I vilken mening? Där finns flera sätt, men det vanligaste är nog att minimera kvadraten i felen.



Lösning. Vi betraktar varje punkt (x_k, y_k) som att x_k är fixerad och y_k är en observation av en stokastisk variabel $Y = \beta_0 + \beta_1 x_k + \epsilon_k$ där ϵ_k är oberoende stokastiska variabler med $E(\epsilon_k) = 0$ och $V(\epsilon_k) = \sigma^2$. Detta är den typiska modellen vid linjär regression. Konstanterna β_0 och β_1 är okända och det är dessa vi vill bestämma. Eftersom

$$E(Y_k) = \beta_0 + \beta_1 x_k \qquad \text{och} \qquad V(Y_k) = \sigma^2$$

så blir

$$Q(\beta_0, \beta_1) = \sum_{k=1}^{n}\big(y_k - E(Y_k)\big)^2 = \sum_{k=1}^{n}\big(y_k - \beta_0 - \beta_1 x_k\big)^2.$$

Minimering av denna funktion med avseende på β_0 och β_1 ger skattningarna $\widehat{\beta}_0$ och $\widehat{\beta}_1$. Jakten på minimum sker nog enklast med lite flervariabelanalys:

$$\mathbf{0} = \nabla Q = (Q'_{\beta_0}, Q'_{\beta_1}) = -2\sum_{j=1}^{n}(y_j - \beta_0 - \beta_1 x_j, x_j(y_j - \beta_0 - \beta_1 x_j))$$

så

$$n\beta_0 + \beta_1 \sum_{j=1}^{n} x_j = \sum_{j=0}^{n} y_j \quad \Leftrightarrow \quad \beta_0 + \beta_1 \overline{x} = \overline{y}$$

och

$$\beta_0 \sum_{j=0}^{n} x_j + \beta_1 \sum_{j=0}^{n} x_j^2 = \sum_{j=1}^{n} x_j y_j \quad \Leftrightarrow \quad n\beta_0\overline{x} + \beta_1 \sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} x_j y_j.$$

Första ekvationen ger att $\beta_0 = \overline{y} - \beta_1\overline{x}$, så

$$n\overline{x}\,\overline{y} - \beta_1 n\overline{x}^2 + \beta_1 \sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} x_j y_j$$

vilket om vi löser ut β_1 leder till

$$\beta_1 = \frac{\sum_{j=1}^{n} x_j y_j - n\overline{x}\,\overline{y}}{\sum_{j=1}^{n} x_j^2 - n\overline{x}^2} = \frac{\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y})}{\sum_{j=1}^{n}(x_j - \overline{x})^2}.$$

3.3 ML-skattning

Låt $X_1, X_2, \ldots, X_n$ vara oberoende stokastiska variabler med täthets- eller sannolikhetsfunktioner $f_i(x; \boldsymbol{\theta})$ respektive $p_i(k; \boldsymbol{\theta})$. Vi antar att samtliga endera är kontinuerliga eller diskreta. Det typiska är att alla variablerna har samma fördelning, men det är inget nödvändigt krav för metoden (däremot förenklar det så klart). Samtliga fördelningar beror dock på en och samma parameter $\boldsymbol{\theta}$ som kan vara vektorvärd.

ML-skattning

Definition. ML-skattningen för $\boldsymbol{\theta}$ är det värde som gör att **likelihood-funktionen** $L(\boldsymbol{\theta})$ maximeras, där

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{n} f_k(x_k; \boldsymbol{\theta}) = f_1(x_1; \boldsymbol{\theta}) \cdot f_2(x_2; \boldsymbol{\theta}) \cdots f_n(x_n; \boldsymbol{\theta})$$

i det kontinuerliga fallet och

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{n} p_k(x_k; \boldsymbol{\theta}) = p_1(x_1; \boldsymbol{\theta}) \cdot p_2(x_2; \boldsymbol{\theta}) \cdots p_n(x_n; \boldsymbol{\theta})$$

i det diskreta fallet.

Så vad är då ML-skattningen? Ganska enkelt är det den skattning som gör att det stickprov vi observerat är det mest troliga. Eftersom vi antar att variablerna som stickprovet är observationer av är oberoende ges den simultana täthets- eller sannolikhetsfunktionen av produkten av de marginella, så vi väljer helt enkelt den skattning som maximerar den simultana tätheten/sannolikheten.
Ofta när man arbetar med ML-skattningar nyttjar man den så kallade log-likelihood-funktionen:

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}).$$

Denna funktion bevarar de flesta av de egenskaper vi är intresserade av eftersom ln är strängt växande och $L(\boldsymbol{\theta}) \in [0, 1]$. Specifikt så har $L(\boldsymbol{\theta})$ och $l(\boldsymbol{\theta})$ samma extrempunkter.

Exempel

Låt $x_1, x_2, \ldots, x_n$ vara ett stickprov av en exponentialfördelning med okänd intensitet θ. Hitta ML-skattningen för θ.

Lösning. Täthetsfunktionen ges av $f(x) = \theta e^{-\theta x}$, $x \geq 0$, så

$$L(\theta) = \prod_{k=1}^{n} \theta e^{-\theta x_k} = \theta^n \exp\left(-\theta \sum_{k=1}^{n} x_k\right) \quad \Rightarrow \quad l(\theta) = \ln L(\theta) = n \ln \theta - \theta \sum_{k=1}^{n} x_k.$$

Vi undersöker vart det finns extrempunkter och finner att

$$0 = l'(\theta) = \frac{n}{\theta} - \sum_{k=1}^{n} x_k \quad \Leftrightarrow \quad \theta = \frac{k}{\sum_{k=1}^{n} x_k} = \frac{1}{\overline{x}},$$

under förutsättning att $\overline{x} \neq 0$. Är detta ett maximum? Använd det ni lärt er i envariabelanalysen! Till exempel ser vi att

$$l''(\theta) = -\frac{n}{\theta^2},$$

så $l''(\theta) < 0$ för alla $\theta > 0$. Således är det ett maximum vi funnit.

Exempel

Låt $X \sim \text{Bin}(n, p)$ med p okänd och låt x vara en observation av X. Hitta ML-skattningen för p.

Lösning. Sannolikhetsfunktionen ges av $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, så

$$\begin{aligned} L(p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &\Rightarrow \quad l(p) = C(n,x) + x \ln p + (n-x) \ln(1-p), \end{aligned}$$

där $C(n,x)$ är en konstant (med avseende på p). Parameterrummet ges av $\Omega_p = (0,1)$. Vi deriverar och erhåller att

$$0 = l'(p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{(1-p)x - (n-x)p}{p(1-p)} = \frac{x - np}{p(1-p)} \quad \Leftrightarrow \quad x = np \quad \Leftrightarrow \quad p = \frac{x}{n}.$$

ML-skattningen är således $\widehat{p} = \dfrac{x}{n}$ om detta är ett maximum. Vi kontrollerar:

		$\widehat{p}$	
$l'(p)$	$+$	0	$-$
$l(p)$	$\nearrow$	max	$\searrow$

Vad skulle hända om observationen blev $x = 0$ (eller $x = n$)?

Exempel

Låt $x_1, x_2, \ldots, x_n$ vara ett stickprov från $N(\mu, \sigma^2)$ där både μ och σ^2 är okända. Hitta ML-skattningarna för μ och σ^2.

Lösning. Vi har nu två okända parametrar och likelihoodfunktionen ges av

$$L(\mu, v) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x_k-\mu)^2}{2v}\right) = \frac{1}{(2\pi v)^{n/2}} \exp\left(-\frac{1}{2v}\sum_{k=1}^{n}(x_k-\mu)^2\right),$$

där $v = \sigma^2$, så

$$l(\mu, v) = \text{konstant} - \frac{n}{2}\ln v - \frac{1}{2v}\sum_{k=1}^{n}(x_k-\mu)^2.$$

Parameterrummet ges av $\Omega_{\mu,v} = \mathbf{R} \times (0, \infty)$ och vi vill maximera $l(\mu, v)$. Stationära punkter finner vi där $\nabla l(\mu, v) = (0, 0)$, så vi beräknar de partiella derivatorerna:

$$l'_\mu(\mu, v) = \frac{1}{v}\sum_{k=1}^{n}(x_k-\mu) = \frac{n}{v}(\overline{x}-\mu)$$

och

$$l'_v(\mu, v) = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{k=1}^{n}(x_k-\mu)^2.$$

Det är tydligt att $\mu = \overline{x}$ och

$$\frac{n}{2v} = \frac{1}{2v^2}\sum_{k=1}^{n}(x_k-\mu)^2 \quad \Leftrightarrow \quad v = \frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)^2,$$

så $\nabla l = 0$ precis då

$$\mu = \overline{x} \quad \text{och} \quad v = \frac{1}{n}\sum_{k=1}^{n}(x_k-\overline{x})^2.$$

Är detta ett maximum? Vi undersöker närmare:

$$H(\mu, v) = \begin{pmatrix} l''_{\mu\mu} & l''_{\mu v} \\ l''_{v\mu} & l''_{vv} \end{pmatrix} = \begin{pmatrix} -\frac{n}{v} & -\frac{n}{v^2}(\overline{x}-\mu) \\ -\frac{n}{v^2}(\overline{x}-\mu) & \frac{n}{2v^2} - \frac{1}{v^3}\sum_{k=1}^{n}(x_k-\mu)^2 \end{pmatrix},$$

där vi låter $\text{SS} = \sum_{k=1}^{n}(x_k-\mu)^2$ och i punkten $(\mu, v) = \left(\overline{x}, \frac{1}{n}\text{SS}\right)$ blir

$$H\left(\overline{x}, \frac{1}{n}\text{SS}\right) = \begin{pmatrix} -\frac{n^2}{SS} & 0 \\ 0 & \frac{n^3}{2SS^2} - \frac{n^3}{SS^3}SS \end{pmatrix} = \begin{pmatrix} -\frac{n^2}{SS} & 0 \\ 0 & -\frac{n^3}{2SS^2} \end{pmatrix},$$

vilket är en negativt definit matris, så detta är ett maximum.
Vi vet sedan tidigare att skattningen för v behöver ha faktorn $1/(n-1)$ för att vara väntevärdesriktig, så ML-skattningen av σ^2 är således inte väntevärdesriktig.

4 Flera stickprov; sammanvägd variansskattning

Antag att vi har två stickprov $x_1, x_2, \ldots, x_m$ och $y_1, y_2, \ldots, y_n$ från normalfördelningar med olika väntevärde men samma varians. ML-skattningarna för respektive väntevärde blir $\widehat{\mu}_1 = \overline{x}$ respektive $\widehat{\mu}_2 = \overline{y}$. För standardavvikelsen kan man visa att den **sammanvägda variansskattningen** (*pooled variance*) blir

$$s^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{n+m-2},$$

där s_1^2 och s_2^2 är stickprovsvarianserna för respektive stickprov. Formeln generaliserar naturligt till fler stickprov. Vi kan även direkt se att

$$E(S^2) = \frac{1}{m+n-2}\big((m-1)E(S_1^2) + (n-1)E(S_2^2)\big) = \frac{1}{m+n-2}\big((m+n-2)\sigma^2\big) = \sigma^2,$$

så skattningen är väntevärdesriktig.

5 Medelfel

Vi har använt variansen $V(\widehat{\Theta})$ (eller standardavvikelsen $D(\widehat{\Theta})$) för att jämföra olika skattningar (effektivitet och konsistens). Mindre varians betyder helt enkelt att skattningen i någon mening är bättre. Detta är ett problem då dessa storheter i allmänhet inte är kända. Vad vi gör är att vi helt enkelt skattar de okända storheterna i $D(\widehat{\Theta})$ och kallar resultatet för medelfelet.

Medelfel

Definition. En skattning $d = d(\widehat{\Theta})$ av standardavvikelsen $D(\widehat{\Theta})$ kallas för skattningens **medelfel**.

Vi ersätter alltså helt enkelt okända storheter i $V(\widehat{\Theta})$ med skattningar. Givetvis påverkar detta precisionen och sättet vi väljer att ersätt de okända storheterna har inverkan på resultatet.

Exempel

Om $X_1, \ldots, X_n$ är ett slumpmässigt stickprov av en $N(\mu, \sigma^2)$-fördelning där både μ och σ^2 är okända kan vi uppskatta μ med medelvärdet $\widehat{M} = \overline{X}$. Således är $D(M) = \dfrac{\sigma}{\sqrt{n}}$, men då σ är okänd behöver vi skatta σ med något. Förslagsvis med stickprovsstandardavvikelsen s, vilket ger medelfelet

$$d(\widehat{M}) = \frac{s}{\sqrt{n}}.$$

Detta är inte på något sätt unikt. En annan skattning av σ ger ett annat medelfel. Med det sagt är detta ett ganska naturligt val för medelfelet.

Ett annat vanligt exempel är vid skattningar av andel. Ofta gör vi som i följande exempel.

Exempel

Ett annat vanligt exempel är när p ska skattas i binomialfördelning. Låt $X \sim \mathrm{Bin}(n, p)$. Vi vet att $V(X) = np(1-p)$ så om vi skattar p med $\widehat{P} = \dfrac{X}{n}$ erhåller vi att $D(\widehat{P}) = \sqrt{\dfrac{p(1-p)}{n}}$. Eftersom p är okänd känner vi inte denna storhet exakt, men medelfelet skulle bli

$$d(\widehat{P}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$