Punktskattningar

1. Momentmetoden

(I) Beräkna $E(X)$:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$ (fått ett värde och satt hatt på den skattade parametern.)

(II) Momentmetoden ger...

$$E(X) = \overline{X} \Rightarrow \text{skattningen } \hat{\theta} = \ldots$$

Följdfrågor:

- Beräkna väntevärde och varians för skattningen:

$$E(\hat{\theta}) = \ldots$$

$$var(X) = E(X^2) - (E(X))^2, \quad E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\, dx$$

Regler:
- $E(aX + b) = aE(X) + b$
- $V(aX + b) = a^2 V(X)$

2. Maximum-Likelihood-metoden

(I) $L(\theta) = \prod_{i=1}^{n} f_X(x) =$

Allmän regel:

$$\prod_{i=1}^{n} (a \cdot x_i) = a^n \cdot \prod_{i=1}^{n} x_i$$

Ex

$$f_X(x) = \frac{x}{\theta^2} e^{-x/\theta} \Rightarrow L(\theta) = \frac{1}{\theta^{2n}} \cdot e^{\frac{1}{\theta}\sum_{i=1}^{n} x_i} \cdot \prod_{i=1}^{n} x_i$$

(II) $\ell(\theta) = \ln(L(\theta))$

$\ln \prod_{i=1}^{n} x_i \Rightarrow$

$\Rightarrow \sum_{i=1}^{n} \ln x_i$

Regler:
- $\ln(a \cdot b) = \ln a + \ln b$
- $\ln\left(\frac{a}{b}\right) = \ln a - \ln b$
- $\ln a^c = c \cdot \ln a$
- $\ln e^a = a$
- $e^{\ln b} = b$
- $\ln e = 1$

(III) $\frac{dl}{d\theta} = 0 \implies \hat{\theta} = \dots$

(IV) Kolla om det är en max-punkt:

$$\frac{d^2 l}{d\theta^2} < 0$$

Regel:

$$\sum_{i=1}^{n} c = n \cdot c$$

Följd frågor:

- Väntesvärdesriktig?

(I) Beräkna det allmäna väntevärdet för $\bar{X}$:

$$E(\bar{X}) = \dots$$

Regel:

- Kan behöva använda PI:

$$\int f(x) \cdot g(x)\, dx = F(x) \cdot g(x) - \int F(x) \cdot g'(x)\, dx$$

- Standard:

$$\int e^{ax}\, dx = \frac{1}{a} e^{ax}$$

(II) VVR om $E(\hat{\theta}) = \theta$

Varför man gör (I) är för att ofta så innehåller skattningen $\bar{X}$ och då kan man skriva om $E(\bar{X}) = ???$

- Effektivast?

$V(\hat{\theta}_1) < V(\hat{\theta}_2) \implies \hat{\theta}_1$ är mer effektiv än $\hat{\theta}_2$

- Konsistent?

Två egenskaper ska skattningen ha för att vara konsistent:

1) VVR: $E(\hat{\theta}) = \theta$

2) $Var(\hat{\theta}) \to 0$ då $n \to \infty \dots$

- $E(\bar{X}^2) = \int x^2 \cdot f_{\bar{X}}(x)\, dx$, $Var(X) = E(X^2) - (E(X))^2$

Ex) · $Var(\hat{\theta}) = \frac{3-\theta^2}{n} \to 0$ då $n \to \infty$

Konfidensintervall

Nedåt begr. intervall: $(\bar{x} - \lambda_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$
Uppåt begr. intervall: $(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma}{\sqrt{n}})$
uppåt begr. för σ eller p $(0, \bar{x} \text{----})$

1. Ett stickprov; $X \sim N(\mu, \sigma)$

Använd hjälpv. på s. 10 (1) eller (3) beroende på om σ är känd eller inte

Om s^2 ej är given kan man använda formeln på samma sida.

Behöver ej räkna ut; slå in på miniräknaren $(x\sigma_{n-1})$

2. Parvisa mätningar $(I_{\Delta i})$

(I)

- Döp första mätningen till x_i och andra till y_i
- Antag att $X_i \sim N(\mu, \sigma_1^2)$ o $Y_i \sim N(\mu + \Delta, \sigma_2^2)$
- Bilda $z_i = y_i - x_i$ som är en obs. från $Z_i = Y_i - X_i \sim N(\Delta, \sigma^2)$
- $\hat{\Delta} = \bar{z} =$
- $\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2$ (slå på miniräknaren)

(II)

Hjälpvariabel: $\frac{\bar{Z} - \Delta}{S/\sqrt{n}} \sim t(n-1)$

(III) stänser in och får intervallet

$$I_\Delta = \left(\bar{z} \mp t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}\right)$$

Om man vill visa med instängning:

$$-t(n-1) \leq \frac{\bar{z} - \Delta}{s/\sqrt{n}} < t(n-1)$$

IV.

- Om intervall positivt $\Rightarrow \bar{y} > \bar{x}$ (verkar det som

3. $\sigma_1 = \ldots \sigma_3 = \sigma \ldots$

(I) Bilda $I_{\mu_i - \mu_j}$

där $\mu_i - \mu_j$ skattas med $\hat{\mu}_i - \hat{\mu}_j = \bar{x}_i - \bar{x}_j$ som är obs.

från $\bar{X}_i - \bar{X}_j \sim N\left(\mu_i - \mu_j \, , \, \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)\right.$

(II)

σ^2 skattas med $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{(n_1-1)+(n_2-1)+(n_3-1)}$

(III) Hjälpvariabeln: $\dfrac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t(n_1 + n_2 + n_3 - 3)$

(IV) Stäng in:

$$I_{\mu_i - \mu_j} = \left(\bar{x}_i - \bar{x}_j \mp t_{\alpha/2}(n_1+n_2+n_3-3) \cdot s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right)$$

(V)

Om 0 ingår i intervallet så är det ingen direkt skillnad

4. Konfidensintervall för variansen σ^2

(I) Använd hjälpv. ② på s.10 $\Rightarrow \dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{0,05}(17) = 8{,}67$

s^2 pss. som ovan

(II) $I_{\sigma^2} = \left(0; \dfrac{(n-1)s'^2}{\chi^2_{0,05}(17)}\right)$

Alt: Hypotesprövning

Om $\dfrac{(n-1)s'^2}{\sigma_0^2} > \chi^2(n-1) \Rightarrow$ förkasta H_0

5. Bestäm konfidensgraden

Givet: $f_X(x) = \theta x^{\theta-1} \quad 0 \le x \le 1$, $y = -\frac{1}{\ln x}$, $I_\theta = \left(\frac{y}{2}, y\right)$

Lösn: $P\left(\frac{Y}{2} \le \theta \le Y\right) = P(\theta \le Y \le 2\theta) = P\left(\theta \le -\frac{1}{\ln x} \le 2\theta\right) =$

$= P(e^{-1/\theta} \le X \le e^{-1/(2\theta)}) = \int_{e^{-1/\theta}}^{e^{-1/2\theta}} \theta x^{\theta-1}\,dx = e^{-1/2} - e^{-1} \approx 0.24$

6. Approximativ konfidensgrad – enl. CGS är $\mu^* = \bar{X} \approx N$ om antalet obs. är tillräckligt stort.

(I) Poisson

(I) Låt $x_1, \dots, x_n$ vara observationer av $X_1, \dots, X_n$, där $X_i \sim Po(\mu)$

(II) $\hat{\mu} = \bar{x}$ som är en obs. av $\hat{M} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

(III.) Den s.v $\sum_{i=1}^{n} X_i \sim Po(n\mu)$ som är appr. $N(n\mu, \sqrt{n\mu})$ om $n\hat{\mu} > 15$

(IV.) Då följer att $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \approx N(\mu, \sqrt{\frac{\mu}{n}})$

(V). Hjälpv. $\frac{\bar{X} - \mu}{\sqrt{\bar{x}/n}} \approx N(0,1)$

(VI) $I_\mu = (\hat{\mu} \mp \lambda_{\alpha/2} \cdot d(\hat{\mu})) = (\bar{x} \mp \lambda_{\alpha/2} \cdot \sqrt{\frac{\bar{x}}{n}})$

(Ex) $X \sim Bin(n, p)$

(I) $p^* = \frac{x}{n}$ obs. från från $p^* = \frac{X}{n} \approx N$ om $np^*(1-p^*) > 10$

(II) $p^* \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

(III) $I_p = \left(p^* \mp \lambda_{\alpha/2} \sqrt{\frac{p^*(1-p^*)}{n}}\right)$

Hypotesprövning

1. Rimligt att anta lika $\sigma_1 = \dots = \sigma_4 = \sigma$?

(I) $H_0 : \sigma_1 = \sigma_4$ mot $H_1 : \sigma_1 \neq \sigma_4$

om H_1 är $\sigma_1 > \sigma_2$ så stryk: $v < \frac{1}{a}$ då väcker $v > a$ och $F_{\alpha}(r_1 - 1, r_2 - 1)$

divideras ej på två

(II) Teststorhet: $v = \frac{s_1^2}{s_4^2}$

(III) Förkasta H_0 om $v < \frac{1}{a}$ eller om $v > a$ där

$$a = F_{\alpha/2}(n_1 - 1, n_2 - 1).$$

2. Vanlig Hypotesprövning

* Binomial

(I) $H_0 : p = 0.5$ mot $H_1 : p < 0.5$

(II) Förkasta H_0 om $x \leq c$

där c fås av $0.1 \geq P(X \leq c \mid X \sim Bin(n, p)) =$

I texten står det "på nivån högst.."

$$= \sum_{i=0}^{c} \binom{10}{i} p^i (1-p)^{10-i}$$ vilket kan kollas upp i tabell:

1. Välj rätt n och p
2. Se när det tar stopp

Ex om $n = 10$ och $p = 0.5$ och $\alpha = 0.1$ som ovan så är $k = 2$ det sista som ger chansen mer när 0.1 (0.0547)

$\Rightarrow c = 2$

(III) Styrkefunktionen ges av

$h(p) = P(H_0 \text{ förkastas om } H_0 \text{ är sann}) = P(X \leq 2 \mid X \sim Bin(n, p)) =$ Tabell.

* Normal ($\bar{X} \sim N(\mu, \sigma^2)$)

om $\mu > 1 \Rightarrow Z > Z_{0.95}$

(I) $H_0 : \mu = 1$ mot $H_1 : \mu \neq 1$ på nivån $\alpha = 0.05$

(II) Testvariabel: $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$ under H_0

(III) Förkasta H_0 om $|Z| > Z_{0.975}$

Idealt så skulle man aldrig förkasta H_0 om den är sann, och därför vill man att $h(\theta)$ ska vara så liten som möjligt.

(IV) Styrkefunktion:

$h(\mu) = P(\text{förkasta } H_0 \text{ då } \mu \text{ är det sanna värdet}) =$

$= P\left(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \lambda_\alpha\right) = P\left(\bar{X} > \dfrac{\sigma}{\sqrt{n}} + 0.2\right) =$

$= P\left(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \lambda_\alpha + \dfrac{0.2 - \mu}{\sigma/\sqrt{n}}\right) =$

$= 1 - \Phi(\lambda_\alpha + 0.2$

Skriv om

* Approximativ nivå

Ex. $H_0 : \lambda = 3$ mot $H_1 : \lambda > 3$, Poisson fördelat.

(I) $\bar{X} \sim Po(30\mu) \approx N(30\lambda, \sqrt{30\lambda})$

(II) $K \cdot \hat{\mu} = \bar{X} \Rightarrow \hat{\mu} = \dfrac{\bar{X}}{K} \approx N\left(\hat{\mu}, \sqrt{\dfrac{\hat{\mu}}{K}}\right)$

(III) När H_0 är sann gäller att $Z = \dfrac{\hat{\mu} - 3}{\sqrt{\mu / K}} \approx N(0,1)$

(IV) Förkasta H_0 om $Z > c = \lambda_{0.95}$

Varians analys

1. R^2

$$R^2 = \frac{SS_R}{SS_{TOT}}, \quad R^2 \rightarrow 1$$

2. Testa om en/flera variabler påverkar...

* En-variabel:

Alt 1:

(I) $H_0: \beta_3 = 0$ mot $H_1: \beta_3 \neq 0$ på nivån...

(II) Teststorhet: $t = \dfrac{\hat{\beta_3}}{d(\hat{\beta_3})} \sim t(n-k-1)$

(III) Förkasta H_0 om $|t| > c = t_{\alpha/2}(n-k-1)$

Alt 2: Konstruera lämpligt KI - 95%:

(I) $I_{\beta_3} = (\hat{\beta_3} \mp t_{\alpha/2}(n-k-1)\, d(\hat{\beta_3}))$

om noll ej ingår intervallet så verkar det som β_3 gör skillnad.

* Flera variabler:

OBS! Om de frågar efter: "gör båda/alla förkl. variabler nytta" så gör testet för en var. ovan och gör för båda

(I) $H_0: \beta_1 = \beta_2 = 0$

(II) Teststorhet: $V = \dfrac{SS_R / k}{SS_E / (n-k-1)} \sim F(k, n-k-1)$ $k = df_{REGR}$, $(n-k-1) = df_{RES}$

(III) Förkasta H_0 om $V > c = F_\alpha(k, n-k-1)$

3. Jämför två modeller

(I) $H_0: \beta_1 = \beta_3 = 0$ mot H_1: minst en av β_1 och $\beta_3 \neq 0$ på nivå 0.1

(I detta fall är de variabler som är extra i modell 2, β_1 och β_3 och alltså därför man kollar om de ger någon betydelse i modellen.)

(II) $\omega = \dfrac{SS_E^{(1)} - SS_E^{(2)} / p}{SS_E^{(2)} / \underbrace{(n-p-k-1)}_{df\, SS_E^{(2)}}}$ (antalet variabler som skiljer i modellerna)

Om det endast skiljer en variabel så testa om den gör nytta i modell 2 => Då är den modellen bättre.
$|t| > c$ ...

(III) Förkasta H_0 om $\omega > c = F_\alpha(p, n-k-p-1)$

3. Konfidens intervall / Prediktions intervall

(I) Bilda ett prediktionsintervall för $u' = (1 \quad x_1 \quad x_2)'$

(II) Hjälpvariabel:

$$\text{Pred}: \quad \frac{u'\hat{\beta} - u'\beta}{s\sqrt{1 + u'(X'X)^{-1}u}} \sim t(n-k-1)$$

$$\text{konf}: \quad \frac{u'\hat{\beta} - u'\beta}{s\sqrt{u'(X'X)^{-1}u}} \sim t(n-k-1)$$

(III)

$$I_{u'\beta} = \left(u'\hat{\beta} \mp t_{\alpha/2}(n-k-1) \cdot s\sqrt{1 + u'(X'X)^{-1}u}\right)$$

Matrisräkning:

$u' = (1 \times 3)$ - matris

$(X'X)^{-1} = (3 \times 3)$ - matris

$u = (3 \times 1)$ - matris

$\hat{\beta} = (3 \times 1)$ - matris - (fås från uppgift)

$$s = \frac{SSE}{(n-k-1)}$$

4. Ange fördelningen för $\hat{\beta}$

$\underline{\hat{\beta}} = (\hat{\beta}_0 \; \hat{\beta}_1)' \sim N(\underline{\beta}, \sigma^2(X'X)^{-1})$ där $(X'X)^{-1} = \ldots$