Parametskattningar

Punktskattning

Låt $x_1, \dots, x_n$ vara observationer av oberoende s.v $X_1, \dots, X_n$ vars sannolikhetsfunktion $p(k, \theta)$ el. täthetsfunktion $f(x, \theta)$ innehåller en okänd parameter θ.

$\Rightarrow$ vi söker ett approximativt värde på θ, dvs en punktskattning.

- θ : verkligt värde
- $\hat{\theta}$: approximativt värde (skattning)
- $\hat{\Theta}$: stokastisk variabel (hur skattningen, $\hat{\theta}$, varierar)

Väntevärdesriktig skattning (vvr):

En skattning är vvr om:

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

$E(\hat{\sigma}^2) = \sigma^2$

Effektiva skattningar:

Om både $\hat{\theta}_1$ & $\hat{\theta}_2$ är vvr skattningar är $\hat{\theta}_1$ effektivare än $\hat{\theta}_2$ om:

$$\mathrm{var}(\hat{\theta}_1) < \mathrm{var}(\hat{\theta}_2)$$

Skattningar som ofta används:

$\hat{\mu} = \bar{x}$

$\hat{\sigma}^2 = s^2$ där σ^2 är varians & s^2 är stickprovsvarians

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

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right) = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right)$$

Repetition:

$V(X) = E(X^2) - E(X)^2$

$E(aX + b) = aE(X) + b$

$V(aX + b) = a^2 V(X)$

Exempel: vvr samt mest effektiv

Stickprov: $x_1 \ldots x_n$

$\bar{X} \sim N(\mu, \sigma)$ där μ o σ är okända

Det finns 2 olika skattningar på μ:

$\boxed{1}$: $\hat{\mu}_1 = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\boxed{2}$: $\hat{\mu}_2 = \frac{x_1 + x_n}{2}$

Är $\hat{\mu}_1$ o $\hat{\mu}_2$ vvr?

$\boxed{1}$: $E(\hat{\mu}_1) = E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) =$

$= \frac{1}{n} \cdot n \cdot \underbrace{E(X_i)}_{=\mu} = \frac{n}{n} \cdot \mu = \mu \Longrightarrow$ vvr!

$\boxed{2}$: $E(\hat{\mu}_2) = E\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{2} E(X_1 + X_n) = \frac{1}{2}\left(\underbrace{E(X_1)}_{=\mu} + \underbrace{E(X_n)}_{=\mu}\right) =$

$= \frac{1}{2}(\mu + \mu) = \frac{1}{2}(2\mu) = \mu \Longrightarrow$ vvr!

Vilken skattning är mest effektiv?

$\boxed{1}$: $V(\hat{\mu}_1) = V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{V(X_i)}_{=\sigma^2} =$

$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$

$\boxed{2}$: $V(\hat{\mu}_2) = V\left(\frac{X_1 + X_n}{2}\right) = \frac{1}{4} V(X_1 + X_n) = \frac{1}{4}\left(\underbrace{V(X_1)}_{=\sigma^2} + \underbrace{V(X_n)}_{=\sigma^2}\right) =$

$= \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{1}{4}(2\sigma^2) = \frac{\sigma^2}{2}$

då $n > 2$ fås:

$\frac{\sigma^2}{n} < \frac{\sigma^2}{2} \Longrightarrow \hat{\mu}_1$ är mer effektiv!

Konsistent skattning

Om $E(\hat{\theta}_n) = \theta$ och $var(\hat{\theta}_n) \to 0$ då $n \to \infty$

så är $\hat{\theta}_n$ en konsistent skattning av θ.

Miniräknare

Ska kunna hitta s, kolla upp hur

Minsta-kvadrat-metoden

Låt $x_1, \dots, x_n$ vara observationer av oberoende s.v $X_1 \dots X_n$

med $E(X_i) = \mu_i(\theta)$ och $var(X_i) = \sigma^2$

Det värde $\hat{\theta}$ som minimerar:

$$Q(\theta) = \sum_{i=1}^{n} (x_i - \mu_i(\theta))^2$$

kallas för minsta-kvadrat-skattning

Exempel - Normalfördelning

$x_1 \dots x_n$ observationer av oberoende s.v. $X_1 \dots X_n$

$X_i \sim N(\mu, \sigma)$, σ är känt

skatta μ med minsta-kvadrat-metoden:

$$Q(\mu) = \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{\partial Q}{\partial \mu} = -2 \sum_{i=1}^{n} (x_i - \mu) = 0 \iff 0 = \sum_{i=1}^{n} (x_i - \mu) = \sum_{i=1}^{n} x_i - n\mu$$

vilket ger $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$

Maximum - Likelihood - Metoden

Maximum - likelihood - funktionen

Diskret: $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i \, ; \, \theta)$

Kontinuerlig: $L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i \, ; \, \theta)$

} sannolikheten att alla händelser händer

Det värde $\hat{\theta}$ som maximerar $L(\theta)$ kallas för maximum-likelihood-skattningen (ML-skattningen) av θ.

Anmärkning: Vid maximering av $L(\theta)$ ska θ betraktas som en variabel o x_i som konstant.

Anmärkning: Det är oftast enklare att maximera

$$\Rightarrow \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i \, ; \, \theta)$$

Repetition Logaritmlagar:

$$\ln(xy) = \ln(x) + \ln(y) \Rightarrow \ln\left(\prod_{i=1}^{n} x_i\right) = \ln(x_1 \cdot \ldots \cdot x_n) = \sum_{i=1}^{n} \ln x_i$$

$$\ln x^y = y \ln x$$

$$\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$$

$$\ln e = 1$$

Exempel:

$$f_X(x) = \frac{x}{a} e^{-\frac{x^2}{2a}} \quad , x \geq 0$$

Ange ML-skattningen av a:

(I) Ställ upp maximum-likelihood-funktionen

$$L(a) = \prod_{i=1}^{n} f_X(x_i) = \frac{x_1}{a} \cdot e^{-\frac{x_1^2}{2a}} \cdot \ldots \cdot \frac{x_n}{a} e^{-\frac{x_n^2}{2a}} = \left(\frac{1}{a}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot e^{-\frac{1}{2a}\sum_{i=1}^{n} x_i}$$

(II) Logaritmera funktionen

$$\ell(a) = \ln(L(a)) = \ln\left(\left(\frac{1}{a}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot e^{-\frac{1}{2a}\sum_{i=1}^{n} x_i^2}\right) =$$

$$= \ln\left(\frac{1}{a^n}\right) + \ln\left(\prod_{i=1}^{n} x_i\right) + \ln\left(e^{-\frac{1}{2a}\sum_{i=1}^{n} x_i^2}\right) =$$

$$= -n\ln a + \sum_{i=1}^{n} \ln x_i + -\frac{1}{2a}\sum_{i=1}^{n} x_i^2 \cdot \underbrace{\ln e}_{=1} =$$

(III) Derivera med avseende på a:

$$\frac{d\ell}{da} = -\frac{n}{a} + \frac{1}{2a^2} \cdot \sum_{i=1}^{n} x_i^2 = 0$$

$$\Leftrightarrow \frac{1}{2a^2}\sum_{i=1}^{n} x_i^2 = \frac{n}{a} \Leftrightarrow \boxed{a = \frac{1}{2n}\sum_{i=1}^{n} x_i^2}$$

Den optimala skattningen blir då $\hat{a} = \dfrac{\sum_{i=1}^{n} x_i^2}{2n}$

(IV) Visa att det är en maximipunkt:

- Teckentabell
- Studera andraderivatan, ska vara < 0

Repetition summor o produkter:

Summor:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

$$\sum_{i=1}^{n} a x_i = a x_1 + a x_2 + \ldots + a x_n = a(x_1 + x_2 \ldots + x_n) = a\sum_{i=1}^{n} x_i$$

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

Produkter:

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot \ldots \cdot x_n$$

$$\prod_{i=1}^{n} a x_i = (a x_1)\cdot(a x_2) \cdot \ldots \cdot (a x_n) = a^n \cdot (x_1 \cdot x_2 \cdot \ldots x_n) = a^n \cdot \prod_{i=1}^{n} x_i$$

Konfidensintervall

Problem: Hur nära det sanna värdet hamnar $\hat{\theta}$?

Lösning: Vi kan beräkna en intervallskattning

Konfidensgrad: $1-\alpha$ där α är risk för fel.
andelen $(1-\alpha)$ av intervallen kommer innehålla θ

Konstruera ett konfidensintervall

(I) Bestäm $\hat{\theta}$

Använd dig av lämplig punktskattning.

(II) Ta fram fördelningen för motsvarande s.v. $\hat{\Theta}$:
Konstruera m.h.a $\hat{\Theta}$ en hjälpvariabel som innehåller θ men inga andra okända parametrar och som har en "känd" fördelning.

(III) Stäng in hjälpvariablerna i ett intervall med sannolikhetsmassa $1-\alpha$.

(IV) Skriv om intervallet i (III) till ett villkor på θ:

$$\Rightarrow P(a_1(\bar{X}_1, \dots, \bar{X}_n) < \theta < a_2(\bar{X}_1, \dots, \bar{X}_n)) = 1-\alpha$$

(V) Sätt in observationerna, beräkna:

$$I_\theta^{1-\alpha} = (a_1(x_1, \dots, x_n), a_2(x_1, \dots, x_n))$$

Exempel: Normalfördelning

x_i : 151320 149832 150458 150020 149263
149908 149395 150956 150003

$\sigma = 442$, konfidensgrad $= 0{,}95$

$I_\mu^{0{,}95}$ sökt , $X_i \sim N(\mu, \sigma)$

(I) Bestäm $\hat{\mu}$:

$\hat{\mu} = \bar{x} = 150\,128{,}33$

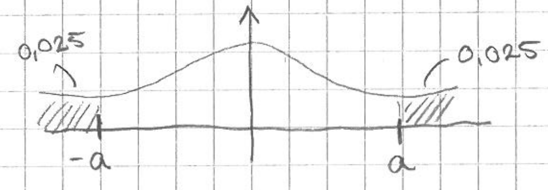
(II) $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$, där $n=9$ (Fördelning)

ger hjälpvariabeln:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

För konstruktion av I_μ då $\underline{\sigma \text{ är känt}}$.

(III) Instängning



ty $\alpha = 0,05$

$\Phi(a) = \underbrace{0,975}_{=\lambda_\alpha}$, tabell ger $a = 1.96$

Vi får då:

$$P\left(-a < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < a\right) = P\left(-1,96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1,96\right) = 0,95$$

konfidensgraden

(IV) Skriv om intervallet:

$$P\left(-1,96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1,96\right) = P\left(-1,96\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < 1,96\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(\bar{X} - 1,96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(a_1(X_1, \ldots X_n) < \mu < a_2(X_1, \ldots X_n)\right) = 0,95$$

(V) Beräkna I_μ:

$$I_\mu = \left(\bar{x} - 1,96\frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + 1,96\frac{\sigma}{\sqrt{n}}\right) =$$

$$= \left(150128.33 - 1,96 \cdot \frac{442}{\sqrt{9}}\ ;\ 150128.33 + 1,96 \cdot \frac{442}{\sqrt{9}}\right) =$$

$$= (149839.56\ ;\ 150417,1066) \approx (149840\ ;\ 150420)$$

Svar: $I_\mu^{0,95} = (149840\ ;\ 150420)$

Två "nya" fördelningar

Dessa två fördelningar används ofta vid konstruktion av konfidensintervall:

- χ^2-fördelning

$\Rightarrow$ används vid konstruktion av konfidensintervall för σ el. σ^2.

- t-fördelning

$\Rightarrow$ används vid konstruktion av konfidensintervall för μ när σ^2 är okänd.

χ-fördelning:

En s.v X med täthetsfunktion av typen

$$f(x) = k \cdot x^{f/2-1} \cdot e^{-x/2} \quad , \; x > 0$$

kallas χ^2-fördelad med frihetsgraden f.

$\Rightarrow X \sim \chi^2(f)$

Om X o Y är oberoende, $X \sim \chi^2(f_1)$ o $Y \sim \chi^2(f_2)$ gäller:

$X + Y \sim \chi^2(f_1 + f_2)$

Sats:

Om $X_1, \ldots, X_n$ är oberoende o $X_i \sim N(\mu, \sigma)$ gäller:

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

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

- $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

- $\bar{X}$ o S^2 är oberoende s.v.

Exempel χ^2-fördelad

vi söker I_σ ; vill veta om $\sigma = 442$ är rimligt.

σ är alltså okänt.

(I) Bestäm s

$$s = \sqrt{\frac{1}{9-1}\sum_{i=1}^{9}(x_i - \bar{x})^2} = 676.76$$

(II) Sätt upp fördelning samt skapa hjälpvariabel

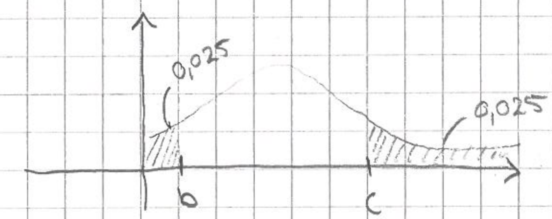
$$\frac{8s^2}{\sigma^2} \sim \chi^2(8)$$

Hjälpvariabeln blir:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

(III) Instängning

$F(b) = 0,025$ ger $b = 2.18$ $\quad$ $F(c) = 0,975$ ger $c = 17,55$



$$P\left(2.18 < \frac{8s^2}{\sigma^2} < 17.55\right) = 0,95$$

(IV) Skriv om intervall till villkor för σ^2

$$P\left(2.18 < \frac{8s^2}{\sigma^2} < 17.55\right) = 0,95 \iff$$

$$\iff P\left(\frac{8s^2}{17.55} < \sigma^2 < \frac{8s^2}{2.18}\right) = 0,95$$

(V) Sätt in observationer och beräkna:

$$I_\sigma = \left(s\sqrt{\frac{8}{17.55}} \; ; \; s\sqrt{\frac{8}{2.18}}\right) \approx (457 \; ; \; 1296)$$

$\Rightarrow$ med 95% sannolikhet ligger $457 < \sigma < 1296$

då $442 \notin I_\sigma$ är det inte speciellt troligt att $\sigma = 442$.

t-fördelning:

En s.v X kallas t-fördelad med frihetsgraden f om täthetsfunktionen är av typen:

$$f(x) = c \cdot \frac{1}{((1+x^2)/f)^{(f+1)/2}}$$

En $t(f)$-fördelning konvergerar mot $N(0,1)$-fördelning då $f \to \infty$

Sats:

Om X o Y är oberoende $X \sim N(0,1)$, $Y \sim \chi^2(f)$ så gäller att:

$$\frac{X}{\sqrt{Y/f}} \sim t(f)$$

Denna sats används när vi tar fram hjälpvariabeln.

Exempel: t-fördelning

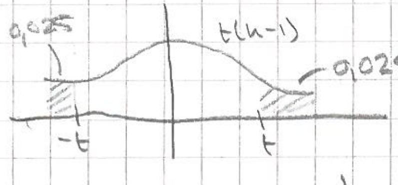
konfidensintervall för μ när σ är okänt.

(I) Gjort tidigare i exemplet.

(II) Konstruera hjälpvariabel:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

(III) $F(t) = 0{,}975$, $f = n-1 = 8$ tabell ger $t = 2.31$



$$P\left(-2{,}31 < \frac{\bar{X} - \mu}{s/\sqrt{n}} < 2.31\right) = 0{,}95$$

(IV) lös ut för μ som i tidigare exempel

(V) $I_\mu = \left(\bar{x} \pm 2.31 \cdot \frac{s}{\sqrt{9}}\right) \approx (149610, 150650)$

Svar: $149610 < \mu < 150650$

Prediktionsintervall

$$X_0 - \bar{X} \sim N\left(0, \sigma\sqrt{1+\frac{1}{n}}\right)$$

Hjälpvariabel:

$$\frac{X_0 - \bar{X}}{S\sqrt{1+\frac{1}{n}}} \sim t(n-1)$$

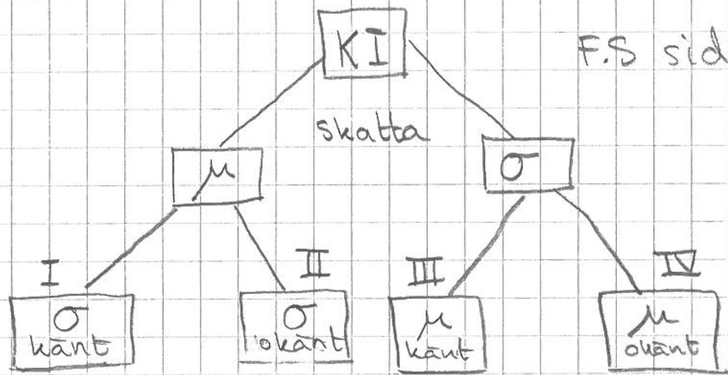
Instängning ger:

$$P\left(\bar{X} - ts\sqrt{1+\frac{1}{n}} < X_0 < \bar{X} + ts\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

Vi får prediktionsintervallet:

$$I_{X_0} = \left(\bar{x} - ts\sqrt{1+\frac{1}{n}} \; ; \; \bar{x} + ts\sqrt{1+\frac{1}{n}}\right)$$

Konfidensintervall bild:



F.S sida 10.

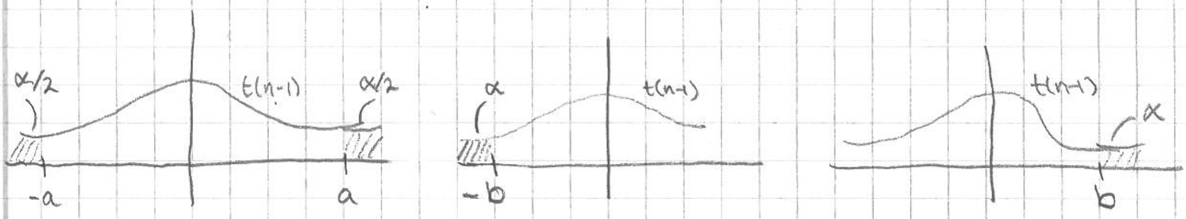
I: hjälpvariabel $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $I_\mu = (\bar{x} \pm \lambda_{\alpha/2} \cdot \sigma/\sqrt{n})$

II: hjälpvariabel $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, $I_\mu = (\bar{x} \pm t_{\alpha/2}(t)\, S \cdot \sqrt{n})$

III: hjälpvariabel $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$, $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$

IV: hjälpvariabel $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

Tre fall för instängning av hjälpvariabel:



Tvåsidigt — Ensidigt uppåt-begränsat — Ensidigt nedåt-begränsat

Exempel:

- Uppåt begränsat

$$P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} > -b\right) = 1-\alpha \quad \text{ger intervallet} \Rightarrow I_\mu = \left(-\infty, \bar{x}+b\frac{s}{\sqrt{n}}\right)$$

- Nedåt begränsat

$$P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} < b\right) = 1-\alpha \quad \text{ger intervallet} \Rightarrow I_\mu = \left(\bar{x} - b\cdot\frac{s}{\sqrt{n}} \; ; \; \infty\right)$$

- tvåsidigt

$$P\left(-a < \frac{\bar{X}-\mu}{S/\sqrt{n}} < a\right) = 1-\alpha \quad \Rightarrow I_\mu = \left(\bar{x} \pm b\cdot\frac{s}{\sqrt{n}}\right)$$

Konfidensintervall för parvisa mätningar

(I) $I_{\mu_1-\mu_2}$ då σ_1 o σ_2 är kända:

Här är μ_1 o μ_2 olika o likaså σ_1 o σ_2. Vi söker $I_{\mu_1-\mu_2}$:

- Punktskattning: $\hat{\mu}_1-\hat{\mu}_2 = \bar{x}-\bar{y}$
- Fördelning: Normalfördelning ty två oberoende s.v

  $\bar{X}-\bar{Y}$ är en kombination av oberoende normalvariabler.

$$\Rightarrow \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

- hjälpvariabel: $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

(II) $I_{\mu_1-\mu_2}$ då $\sigma_1=\sigma_2=\sigma$ (okänd men samma)

Vi söker:

1. I_σ el. I_{σ^2}
2. $I_{\mu_1-\mu_2}$
3. $I_{c_1\mu_1+c_2\mu_2}$

Täthetsfunktionerna är förskjutna i förhållande till varandra <u>men</u> de har samma form.

- Om vi söker I_σ el. I_{σ^2}:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$ F.S sid. 11

$$\Rightarrow \frac{(n_1+n_2-2)s^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$$ F.S. sid 11

- Om vi söker $I_{\mu_1-\mu_2}$:

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

- Om vi söker $I_{c_1\mu_1+c_2\mu_2}$:

$\boxed{\sigma_1 = \sigma_2 = \text{okända}}$

$$\frac{c_1\bar{X}+c_2\bar{Y}-(c_1\mu_1+c_2\mu_2)}{s\sqrt{\frac{c_1^2}{n_1}+\frac{c_2^2}{n_2}}} \sim t(n_1+n_2-2)$$

$\boxed{\sigma_1 \neq \sigma_2 \text{, okända}}$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t(v)$$

där $$v = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

<u>F-fördelning</u>:

Vid jämförelse av varianser behöver vi F-fördelning.
Kommer användas i samband med regressionsanalys. (variansanalys)
Om Y_1 o Y_2 är oberoende $Y_1 \sim \chi^2(r_1)$ o $Y_2 \sim \chi^2(r_2)$
så gäller:

$$V = \frac{Y_1/r_1}{Y_2/r_2} \sim F(r_1, r_2)$$

Anm: $\frac{1}{V} \sim F(r_2, r_1)$

Centrala gränsvärdessatsen

$x_1, \ldots, x_n$ är observationer av oberoende likafördelade s.v $X_1, \ldots, X_n$ som inte är normalfördelade.

Om:

- $X \sim hyp(N, n, p)$ o $\frac{N-n}{N-1} np(1-p) \geq 10$

  gäller $\Rightarrow X \approx N\left(np, \sqrt{\frac{N-n}{N-1} np(1-p)}\right)$

- $X \sim Bin(n, p)$ o $np(1-p) \geq 10$

  gäller $\Rightarrow X \approx N(np, \sqrt{np(1-p)})$

centrala gränsvärdessatsen säger:

- $\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$

- $\bar{X}_n \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

- $\sum_{j=1}^{n} X_j \approx N(n\mu, \sqrt{n}\,\sigma)$

Hypotesprövning

När man bildar konfidensintervall vill man dra slutsatser om intressanta parametrar. Kallas signifikanstest eller hypotesprövning.

Vi vill pröva en nollhypotes mot en mothypotes:

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0 \quad <, >$

Det finns två typer av fel:

- Fel av typ I:

  H_0 förkastas fast än H_0 är sann

  α = risk för fel I

  Vi vill att risken att vi slänger en hypotes som är sann ska vara så liten som möjligt

- Fel av typ II:

  H_0 förkastas inte fast än H_0 är falsk

  β = risk för fel II.

  Vi vill att risken att vi sparar en hypotes som är falsk är så liten som möjligt.

Signifikansnivå:

$\alpha = P(H_0 \text{ förkastas trots att } H_0 \text{ är sann})$

α ska vara så lågt som möjligt

Styrka:

För θ_1:

$h(\theta_1) = P(H_0 \text{ förkastas då } \theta = \theta_1)$

hög styrka $\Longrightarrow$ bra

$\beta = 1 - h(\theta_1)$

Styrkefunktionen:

Beroende på vad θ är så kommer testet vara olika "starkt"

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

P-värde:

P = sannolikheten, givet H_0, att få ett visst värde/utfall

$P < \alpha \Rightarrow$ Förkasta

$P > \alpha \Rightarrow$ Förkasta ej

Exempel:

$\bar{X} \sim N(\mu; 0.2)$

$H_0 : \mu = 4.0$

$H_1 : \mu > 4.0$

$\alpha = 0,05$

10 mätningar gav $\bar{x} = 4.1$

0,05 = P(H_0 förkastas trots H_0 är sann)

$\bar{X} \sim N(\mu, 0.2/\sqrt{N})$

$$Z = \frac{\bar{X} - \mu}{0.2/\sqrt{n}} \sim N(0,1)$$

$P(Z > a \mid H_0 \text{ sann}) = 0,05$

$P(Z < a) = 0,95 = \Phi(a)$

$\Rightarrow a = 1,645$ från tabell

$\therefore$ Sannolikheten att $z > 1,645$ är 0,05 %, givet H_0 = sann
samt $\mu = 4.0$

Studera nu utfallet:

$\bar{x} = 4.1$, $n = 10$, $\mu = 4$

$$z = \frac{\bar{x} - \mu}{0.2/\sqrt{n}} = \frac{4.1 - 4}{0.2/\sqrt{10}} \approx 1,5811 < 1.645$$

$\rightarrow$ H_0 kan inte förkastas.

Multivariat normalfördelning

Beroendemått

Som beroendemått används ofta kovarians o korrelation.

Kovarians:

$$\mathrm{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$\boxed{\text{F.S sid 7}}$

Korrelation:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

Stokastiska vektorer:

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A\underline{X} + b$$

$$\begin{cases} E(\underline{Y}) = A\,E(\underline{X}) + b \\ C_y = A\,C_x\,A^T \end{cases}$$

$$\underline{Y} \sim N(E(\underline{Y}), C_y)$$

Linjärkombination:

$$Z = a_1 X_1 + \dots + a_n X_n$$

$$\mathrm{var}(Z) = C_z = (a_1, \dots, a_n)\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Korrelation ρ

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

$\rho\sigma_1\sigma_2$ = kovariansen

$\sigma_1^2 = \mathrm{var}(X)$ i C

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \text{kovariansmatris}$$

om X o Y är oberoende är $\rho = 0$

$\Rightarrow$ C blir en diagonalmatris

om X o Y är oberoende o likafördelade

$$\Rightarrow C = \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} = \sigma^2 I$$

Flera dimensioner N, täthetsfunktionen:

$$f(\underline{Y}) = \frac{1}{\sqrt{2\pi}^{\,n}\sqrt{\det C}} \cdot e^{-\frac{1}{2}\left((Y-\mu)^T C^{-1} (Y-\mu)\right)}$$

Exempel:

$X_1, X_2 \sim N(0,1) \Rightarrow$ oberoende o likafördelade

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 - 2X_2 \\ X_1 + cX_2 \end{pmatrix}$$

a) Ange $\underline{Y}$:s fördelning

b) Bestäm c så Y_1 o Y_2 är oberoende.

a) (I) Bestäm A

$$\underline{Y} = \underbrace{\begin{pmatrix} 1 & -2 \\ 1 & c \end{pmatrix}}_{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

(II) Bestäm $E(Y)$

$E(Y) = A\,E(X)$

då $E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ fås

$$E(Y) = \begin{pmatrix} 1 & -2 \\ 1 & c \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(III) Bestäm C_Y

$C_Y = A\,C_X\,A^T$

$C_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$ ty $X_1, X_2 \sim N(0,1)$ oberoende o likafördelade

$$C_Y = \begin{pmatrix} 1 & -2 \\ 1 & c \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -2 & c \end{pmatrix} = \begin{pmatrix} 5 & 1-2c \\ 1-2c & 1+c^2 \end{pmatrix}$$

svar: $\underline{Y} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 1-2c \\ 1-2c & 1+c^2 \end{pmatrix}\right)$

b) C_Y ska då vara en diagonalmatris, $1-2c = 0$:

$1 - 2c = 0 \Rightarrow c = \frac{1}{2}$

svar: $c = 1/2$

Linjär algebra repetition.

- Inversen

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \Rightarrow \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Regressionsanalys