

Industriell matematik och statistik, LMA136 2013/14

21 Februari 2014

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

Disposition

Repetition

Punktskattning

Intervallskattning

- Konfidensintervall

- Väntevärden samt andelsparametrar

- Väntevärdesdifferens

- Varianser och standardavvikelser

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

Centrala gränsvärdessatsen

- ▶ $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- ▶ Om $X \sim N(\mu, \sigma)$ så gäller att $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$
- ▶ Om $X_1, X_2, \dots, X_n$ är n stycken oberoende och likafördelade stokastiska variabler med väntevärde μ och standardavvikelse σ . Då gäller att:

$$\sum_{i=1}^n X_i \text{ approx } N(n\mu, \sqrt{n}\sigma)\text{-fördelad}$$

$$\bar{X}_n \text{ approx } N(\mu, \sigma/\sqrt{n})\text{-fördelad}$$

- ▶ $X \sim \text{Bin}(n, p)$ så gäller för stora n att

$$X \text{ approx } N(np, \sqrt{np(1-p)})\text{-fördelad}$$

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

- Data, observationer; $\{x_1, x_2, \dots, x_n\}$
- Modellparametrar; $N(\mu, \sigma)$, $Bin(n, p)$, $Po(\lambda)$
- Vilken parameter beskriver datan bäst?

Skatta en parameter θ , så används en skattare $\hat{\theta}$. Det finns möjligtvis flera skattare till en och samma parameter.

Ex. Antag att data ser ut att vara oberoende och normalfördelat (μ, σ beskriver hela fördelning), dvs. $X_i \sim N(\mu, \sigma)$.

Väntevärdesskattare	Standardavvikelseskattare
$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$	$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$
$\hat{\mu} = \text{median}(\{X_1, X_2, \dots, X_n\})$	$\hat{\sigma} = Q/1.35$ (Q Kvartilavståndet)
$\hat{\mu} = \frac{1}{n+1} (2X_1 + \sum_{i=2}^n X_i)$	$\hat{\sigma} = \sqrt{(X_1 - \bar{X})^2}$

- ▶ Väntevärdesriktighet, $\mathbb{E}[\hat{\theta}] = \theta$
- ▶ Effektivitet, Variansen av $\hat{\theta}$ så liten som möjligt.

Ex. $X = \{X_1, X_2, X_3\}$ då $X_i \sim N(\mu, \sigma)$ och oberoende.

- ▶ $\hat{\mu}_1 = \bar{X}$
- ▶ $\hat{\mu}_2 = \frac{1}{6}(2X_1 + 3X_2 + X_3)$

Väntevärde	Varians
$\mathbb{E}[\hat{\mu}_1] = \mu$	$\text{Var}(\hat{\mu}_1) = \frac{1}{3}\sigma^2$
$\mathbb{E}[\hat{\mu}_2] = \mu$	$\text{Var}(\hat{\mu}_2) = \frac{1}{6^2}(4 + 9 + 1)\sigma^2 = \frac{7}{6}\frac{1}{3}\sigma^2$

Skattare av stickprovsvariansen

Data $x = \{x_1, x_2, \dots, x_n\}$

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

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

► Väntevärdesriktighet, $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$

$$\begin{aligned} \mathbb{E}[\hat{\sigma}^2] &= \frac{1}{n-1} \mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \dots \\ &= \frac{1}{n-1} \frac{n-1}{1} \sigma^2 = \sigma^2 \end{aligned}$$

Repetition

Punktskattning

Intervallskattning

Konfidsensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

X = antal lyckade utfall av n oberoende försök.

$X \sim \text{Bin}(n, p)$ men p är okänd.

Skatta sannolikheten att ett enskilt försök lyckas.

Låt $\hat{p} = X/n$ då gäller att

$$\mathbb{E}[\hat{p}] = \mathbb{E}[X/n] = \frac{1}{n}\mathbb{E}[X] = p$$

$$\text{Var}(\hat{p}) = \text{Var}(X/n) = \frac{1}{n^2}\text{Var}(X) = \frac{p(1-p)}{n}$$

Centrala gränsvärdessatsen ger att

X är approx $N(np, \sqrt{np(1-p)})$ -fördelad.

$\hat{p}$ är approx $N(p, \sqrt{\frac{p(1-p)}{n}})$ -fördelad.

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

- ▶ Väntevärdet för en s.v.
parameter μ ; punktskattning $\bar{X}$; standardavvikelse $\frac{\sigma}{\sqrt{n}}$;
medelfel $\sqrt{\frac{s^2}{n}}$
- ▶ Andelsparameter för en s.v.
parameter p ; punktskattning $\hat{p} = X/n$; standardavvikelse
 $\sqrt{\frac{p(1-p)}{n}}$; medelfel $\sqrt{\frac{\frac{x}{n}(1-\frac{x}{n})}{n}}$

Intervallskattning

(Använda centrala gränsvärdessatsen på skattarna.)

Skattare $\hat{\theta}$ är kontinuerlig stokastisk variabel

► $P(\hat{\theta} = x) = 0$ för varje enskilt x

Vill ange ett intervall som θ ligger inom med viss sannolikhet.

Ex. antag att vi vet att en väntevärdesriktig skattare är normalfördelad med given standardavvikelse $S(\hat{\theta})$;

$$\hat{\theta} \sim N(\theta, S(\hat{\theta}))$$

dvs

$$\frac{\hat{\theta} - \theta}{S(\hat{\theta})} \sim N(0, 1)$$

($T = \frac{\hat{\theta} - \theta}{S(\hat{\theta})}$ kallas testvariabel eller statistika)

Repetition

Punktskattning

Intervallskattning

Konfidsensintervall
Väntevärden samt
andelsparametrar
Väntevärdesdifferen
Varianser och
standardavvikelser

Intervallskattning

$$Z \sim N(0, 1)$$

$$\blacktriangleright P(Z \leq q_{1-0.025}) = 1 - 0.025, q_{1-0.025} = 1.96$$

$$\blacktriangleright P(Z \leq q_{0.025}) = 0.025, q_{0.025} = -1.96$$

$$\begin{aligned} P(|Z| \leq 1.96) &= P(Z \leq q_{1-0.025}) - P(Z \leq q_{0.025}) \\ &= 1 - 0.025 - 0.025 \\ &= 0.95 \end{aligned}$$

$$\begin{aligned} 0.95 &= P\left(\left|\frac{\hat{\theta} - \theta}{S(\hat{\theta})}\right| \leq 1.96\right) \\ &= P(\hat{\theta} - 1.96S(\hat{\theta}) \leq \theta \leq \hat{\theta} + 1.96S(\hat{\theta})) \end{aligned}$$

Givet att $\hat{\theta} \sim N(\theta, S(\hat{\theta}))$ ges ett 95% konfidensintervall för parameter θ av intervallet $(\hat{\theta} - 1.96S(\hat{\theta}), \hat{\theta} + 1.96S(\hat{\theta}))$

Repetition

Punktskattning

Intervallskattning

KonfidensintervallVäntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

Konfidsensintervall normalfördelning

Repetition

Punktskattning

Intervallskattning

KonfidsensintervallVäntevärden samt
andelsparametrar

Väntevärdesdifferen

Varianser och
standardavvikelser

- ▶ $\hat{\theta} \sim N(\theta, S(\hat{\theta}))$
- ▶ $q_{\alpha/2}$ är $(\alpha/2)$ -quantilen för $Z \sim N(0, 1)$,
(dvs $P(Z < q_{\alpha/2}) = \alpha/2$)

Så är

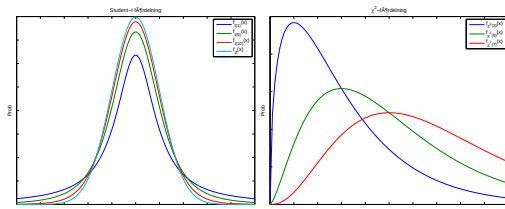
$$(\hat{\theta} + q_{\alpha/2} S(\hat{\theta}), \hat{\theta} - q_{\alpha/2} S(\hat{\theta}))$$

ett $(1 - \alpha)$ -konfidsensintervall för θ .

Andra fördelningar

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}, \text{ (en stokastisk variabel)}$$

- ▶ Student-t fördelning, fördelning för $\frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}}$ om $X = \{X_1, \dots, X_n\}$ oberoende normalfördelade med väntevärde μ ($(n-1)$ -frihetsgrader). Betecknas t (frihetsgrader)
- ▶ χ^2 -fördelning, fördelning för $\sum_{i=1}^n Z^2$, n -frihetsgrader. Betecknas χ^2 (frihetsgrader)



Figur: Frekvensfunktioner för Student-t- och χ^2 -fördelningar.

Repetition

Punktskattning

Intervallskattning

KonfidsensintervallVäntevärden samt
andelsparametrar

Väntevärdesdifferen

Varianser och
standardavvikelser

Ex. Skatta väntevärdet av likafördelade observationer

$x = \{x_1, x_2, \dots, x_n\}$ av s.v. X ,

- om vi känner $\sigma = \sqrt{\text{Var}(X)}$, $\hat{\theta} = \bar{X} \sim N(\mu, \sigma/\sqrt{n})$

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

Beräkna $\bar{x}$ och finn kvantiler till $N(0, 1)$ -fördelningen

- om σ okänd

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim t(n - 1)$$

Beräkna $\bar{x}$, $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ och finn kvantiler till $t(n - 1)$ -fördelningen.

Repetition

Punktskattning

Intervallskattning

KonfidensintervallVäntevärden samt
andelsparametrar

Väntevärdesdifferens

Varianser och
standardavvikelser

- ▶ Statistika T har fördelning F
(T är en funktion av $\hat{\theta}$ och θ)
- ▶ $q_{\alpha/2}$ och $q_{1-\alpha/2}$ är $(\alpha/2)$ -kvantil respektive $(1 - \alpha/2)$ -kvantil för fördelningen F .
- ▶ Så gäller

$$P(q_{\alpha/2} < T < q_{1-\alpha/2}) = 1 - \alpha$$

- ▶ Lös ut den sökta parametern ur olikheterna.
- ▶ $(1-\alpha)$ -konfidsensintervallet för θ definieras av det resulterande intervallet.

- ▶ μ = väntevärdet
- ▶ σ = standardavvikelsen
- ▶ p = sannolikhet för lyckat resultat i enskilt försök
- ▶ $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
- ▶ $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Storhet	Statistika	Fördelning	Tumregler
Väntevärde	$\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$	$N(0, 1)$	σ känd
Väntevärde	$\frac{\bar{X}_n - \mu}{\sqrt{\hat{\sigma}^2/n}}$	$t(n-1)$	σ okänd, skatta s
Andel	$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}}$	$N(0, 1)$	

Beräkna $\bar{x}$ och $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ (tal ej s.v.!) och beräkna konfidensintervall för μ

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt andelsparametrar

Väntevärdesdifferens

Varianser och standardavvikelser

Statistikor för väntevärdesdifferenser

- ▶ Data $x = \{x_1, x_2, \dots, n_X\}, y = \{y_1, y_2, \dots, n_Y\}$.
- ▶ Observationer av respektive s.v. X och Y med väntevärden μ_X, μ_Y och standardavvikelser σ_X, σ_Y .
- ▶ $\Delta(X, Y) = \bar{X} - \bar{Y}$ (en s.v. med väntevärde $\mathbb{E}[\Delta(X, Y)] = \mu_X - \mu_Y$)
- ▶ $S(\sigma_X, \sigma_Y) = \sqrt{\text{Var}(\Delta(X, Y))}$

Statistika	Fördelning
σ_X, σ_Y kända	$\frac{\bar{\Delta} - (\mu_X - \mu_Y)}{S(\sigma_X, \sigma_Y)} \quad N(0, 1)$

- ▶ Skatta s_X, s_Y om σ_X, σ_Y okända

Statistika	Fördelning	Tumregler
$\frac{\bar{\Delta} - (\mu_X - \mu_Y)}{S(s_X, s_Y)}$	$N(0, 1)$	n_X, n_Y stora
$\frac{\bar{\Delta} - (\mu_X - \mu_Y)}{S(s_X, s_Y)}$	$t(n_X + n_Y - 2)$	$\sigma_X = \sigma_Y$
$\frac{\bar{\Delta} - (\mu_X - \mu_Y)}{S(s_X, s_Y)}$	$t(df)$	$df \approx \frac{(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y})^2}{(\frac{s_X^2}{n_X})^2 + (\frac{s_Y^2}{n_Y})^2}$

- ▶ Beräkna $\bar{\Delta} = \bar{x} - \bar{y}$, s_X, s_Y (tal!) och beräkna konfidsintervall för $\mu_X - \mu_Y$

Repetition

Punktskattning

Intervallskattning

Konfidsintervall

Väntevärden samt andelsparametrar

Väntevärdesdifferenser

Varianser och standardavvikelser

Statistikor och konfidensintervall för varianser och standardavvikelser

$$\blacktriangleright \bar{X}_n = \frac{1}{n} \sum_{i=1}^n x_i, \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$$

Statistika	Fördelning
$\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$	$\chi^2(n-1)$

q kvantiler till $\chi^2(n-1)$ -fördelningen

$$1 - \alpha = P(q_{\alpha/2} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < q_{1-\alpha/2})$$

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

undre gräns		övre gräns
$\frac{(n-1)s^2}{q_{1-\alpha/2}}$	$< \sigma^2 <$	$\frac{(n-1)s^2}{q_{\alpha/2}}$
$\sqrt{\frac{(n-1)s^2}{q_{1-\alpha/2}}}$	$< \sigma <$	$\sqrt{\frac{(n-1)s^2}{q_{\alpha/2}}}$

Repetition

Punktskattning

Intervallskattning

Konfidensintervall

Väntevärden samt
andelsparametrar

Väntevärdesdifferens

**Varianser och
standardavvikelser**