TMA4245 Statistikk
Vår 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 11, blokk II
Løsningsskisse

Oppgave 1

(Merk: I følgje oppgåveteksten skal konfidensintervallet *utleias*, ikkje berre setjas opp!)

Vi har at $X_1, \ldots, X_n$ er u.i.f. $\mathrm{N}(\mu_1, \sigma_0^2)$ og at $Y_1, \ldots, Y_m$ er u.i.f. $\mathrm{N}(\mu_2, \sigma_0^2)$, og også at alle X_i-ane er uavhengige av alle Y_j-ane, $i = 1, \ldots, n,\ j = 1, \ldots, m$. Forventningsverdiane μ_1 og μ_2 er ukjende, medan variansen σ_0^2 er felles og kjend.

Naturleg estimator: $\hat{\mu_1} - \hat{\mu_2} = \bar{X} - \bar{Y}$

Då estimatoren er ein lineærkombinasjon av uavhengige normalfordelte variable er han sjølv normalfordelt med:

$$\begin{aligned}
\mathrm{E}(\hat{\mu_1} - \hat{\mu_2}) &= \mathrm{E}(\bar{X}) - \mathrm{E}(\bar{Y}) = \mu_1 - \mu_2 \\
\mathrm{Var}(\hat{\mu_1} - \hat{\mu_2}) &= \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma_0^2}{n} + \frac{\sigma_0^2}{m}.
\end{aligned}$$

D.v.s:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathrm{N}(0, 1)$$

som gjev:

$$\begin{aligned}
P\left(-z_{\frac{\alpha}{2}} \le \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}} \le z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\
P\left(\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}} \le \mu_1 - \mu_2 \le \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}\right) &= 1 - \alpha
\end{aligned}$$

D.v.s. at vi får $(1 - \alpha)100\%$ konfidensintervall ved:

$$\left[\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sigma_0 \sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

For å få numerisk svar set vi inn talverdiane: $\bar{x} = 28.80$, $\bar{y} = 26.07$, $\sigma_0 = 2$, $m = n = 10$ og $z_{0.025} = 1.96$. Får då eit 95%-konfidensintervall på:

$\underline{\underline{[0.977, 4.483]}}$

Oppgave 2 $E(\hat{\mu}) = \mu$

$Var(\hat{\mu}) = \frac{\tau_0^4 Var(X) + \sigma_0^4 Var(Y)}{(\tau_0^2+\sigma_0^2)^2} = \frac{\tau_0^4\sigma_0^2 + \sigma_0^4\tau_0^2}{(\tau_0^2+\sigma_0^2)^2} = \frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}$

$\hat{\mu}$ er en lineærkombinasjon av normalfordelte variable, og er dermed normalfordelt.

$Z = \frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}} \sim N(0,1)$

$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$

$P(\hat{\mu} - z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}} < \mu < \hat{\mu} + z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}) = 1-\alpha$

Ett $1-\alpha$ konfidensintervall for μ er da :

$(\hat{\mu} - z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}, \hat{\mu} + z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}})$

Oppgave 3

a) $P(B) = 2P(X < 0.1 - 3\sigma) = 2P(\frac{X-0.1}{\sigma} < -3) = 2 \cdot 0.0013 = 0.0026$

$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{0.0026}{2P(X<0.1-2\sigma)} = \frac{0.0026}{2\cdot 0.0228} = 0.057$

Med $\mu = 0.11$ blir $P(B)$ lik:

$P(B) = P(X > 0.1 + 3\sigma) + P(X < 0.1 - 3\sigma) = P(\frac{X-0.11}{\sigma} > \frac{-0.01}{\sigma} + 3) + P(\frac{X-0.11}{\sigma} < \frac{-0.01}{\sigma} - 3) = P(\frac{X-0.11}{\sigma} > 2) + P(\frac{X-0.11}{\sigma} < -4) = 0.0228 + 0 = 0.0228$

b) En god estimator er forventningsrett og har liten varians.

I beregningene benytter vi at $\frac{\sum(X_i-\mu)^2}{\sigma^2}$ er χ^2 fordelt med n frihetsgrader. Dvs at $E(\frac{\sum(X_i-\mu)^2}{\sigma^2}) = n$ og at $Var(\frac{\sum(X_i-\mu)^2}{\sigma^2}) = 2n$.

$E(\hat{\sigma}^2) = E(\frac{\sigma^2}{n}\frac{\sum(X_i-\mu)^2}{\sigma^2}) = \frac{\sigma^2}{n}E(\frac{\sum(X_i-\mu)^2}{\sigma^2}) = \sigma^2$

$Var(\hat{\sigma}^2) = \frac{\sigma^4}{n^2}Var(\frac{\sum(X_i-\mu)^2}{\sigma^2}) = 2\sigma^4/n$

I beregningene benytter vi at $\frac{\sum(X_i-\bar{X})^2}{\sigma^2}$ er χ^2 fordelt med $n-1$ frihetsgrader. Dvs at $E(\frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = n-1$ og at $Var(\frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = 2(n-1)$.

$E(S^2) = E(\frac{\sigma^2}{n-1}\frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = \frac{\sigma^2}{n-1}E(\frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = \sigma^2$

$Var(S^2) = \frac{\sigma^4}{(n-1)^2}Var(\frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = 2\sigma^4/(n-1)$

Både $\hat{\sigma}^2$ og S^2 er forventningsrette, men $\hat{\sigma}^2$ har mindre varians, og er derfor å foretrekke.

c) $P(\chi^2_{0.95,20} < \frac{\sum(X_i-\mu)^2}{\sigma^2} < \chi^2_{0.05,20}) = 0.9$

Vi flytter om innenfor P tegnet til vi får σ^2 for seg selv.

$P(\frac{\sum(X_i-\mu)^2}{\chi^2_{0.05,20}} < \sigma^2 < \frac{\sum(X_i-\mu)^2}{\chi^2_{0.95,20}}) = 0.9$

Ved å innsette tall: $\sum(X_i-\mu)^2 = 0.0018$, samt $\chi^2_{0.95,20} = 10,85$ og $\chi^2_{0.05,20} = 31,41$, blir et 0.9 konfidensintervall for σ^2 gitt ved $(5.7 \cdot 10^{-5}, 17 \cdot 10^{-5})$.

d)

Vårt ønske på konfidensintervallet er at: $\frac{\sum(X_i-\mu)^2}{\chi^2_{0.05,20}} - \frac{\sum(X_i-\mu)^2}{\chi^2_{0.95,20}} = \frac{n\hat{\sigma}^2}{\chi^2_{0.05,20}} - \frac{n\hat{\sigma}^2}{\chi^2_{0.95,20}} \leq \frac{\hat{\sigma}^2}{2}$

$\hat{\sigma}^2$ forkortes slik at: $\frac{n}{\chi^2_{0.05,20}} - \frac{n}{\chi^2_{0.95,20}} \leq \frac{1}{2}$

Ved å prøve seg fram i tabellen får man at dette er oppfyllt for $n > 100$.

Oppgave 4

a) $X \sim bin(n,p)$ fordi:

- Undersøker n uavhengige delar av DNA-strukturen.
- Finn for kvar del ut om denne delen av DNA-stukturen er samanfallande eller ikkje.
- Sannsynet for samanfallande er det same for alle delane ($P(\text{samanfell}) = p = 0.15$).

$$\begin{aligned}
P(X=2) &= \binom{5}{2} 0.15^2 (1-0.15)^{5-2} = \underline{\underline{0.138}} \\
P(X \geq 2) &= 1 - P(X<2) = 1 - P(X \leq 1) = 1 - 0.835 = \underline{\underline{0.165}} \\
P(X=2|X\geq 2) &= \frac{P(X=2 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(X=2)}{P(X \geq 2)} \\
&= \frac{0.138}{0.165} = \underline{\underline{0.836}}
\end{aligned}$$

b)

$$\begin{aligned}
P(\text{Type-I-feil}) &= P(\text{forkaste } H_0|H_0) \\
&= P(X=5|p=0.15) = 0.15^5 = \underline{\underline{0.000076}} \\
P(\text{Type-II-feil}) &= P(\text{ikkje forkaste } H_0|H_1) \\
&= P(X<5|p=1) = \underline{\underline{0}}
\end{aligned}$$

Ser så på det generelle uttrykket for sannsynet for type-I-feil når vi har n forsøk. Finn derfrå kor stor n må vere for å oppnå ønska sannsyn for type-I-feil.

$$\begin{aligned}
P(\text{Type-I-feil}) = P(X=5|p=0.15) = 0.15^n &< 0.000001 \\
n \ln(0.15) &> \ln(0.000001) \\
n &> \frac{\ln(0.000001)}{\ln(0.15)} = 7.28
\end{aligned}$$

Minst $\underline{\underline{\text{8 delar}}}$ frå DNA-strukturen må undersøkast dersom sannsynet for type-I-feil skal vere mindre enn 0.000001.