Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

EKSAMEN I TMA4245 STATISTIKK

Løsningsforslag

XX. august 2010

Oppgave 1

a) Tettheten til Y blir

$$f_Y(y) = f_X(x(y)) \left| \frac{dx}{dy} \right| = \frac{1}{\sqrt{2\pi}\tau e^y} e^{-\frac{1}{2\tau^2}(\ln e^y - \nu)^2} e^y = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau^2}(y-\nu)^2}, \tag{1}$$

som er tettheten til en normalfordelt variabel med forventing ν og varians τ^2.

b) Fra definisjon av kumulativ tetthet og sammenhengen mellom X og Y får vi

$$F_X(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq \ln x) = P(Z \leq \frac{\ln x - \nu}{\tau}) = \Phi(\frac{\ln x - \nu}{\tau}), \tag{2}$$

hvor $Z = (Y - \nu)/\tau \sim N(0, 1)$.

c) For $\nu = 1.0$ og $\tau = 0.8$ blir

$$P(X_1 \leq 1.0) = \Phi(\frac{\ln 1.0 - 1.0}{0.8}) = 0.1057, \tag{3}$$

og

$$P(X_1 \cdot X_2 \leq 1) = P(\ln X_1 + \ln X_2 \leq \ln 1) = P(Y_1 + Y_2 \leq 0) = P(Z \leq \frac{-2\nu}{\sqrt{2\tau^2}}) = 0.0385, \tag{4}$$

siden $Y_1 + Y_2$ er normalfordelt med forventning 2ν og varians $2\tau^2$.

d) Sannsynligheten for at hver av jordprøvene har nikkelinnhold $X_i < 2.72$ mg blir

$$p = P(X_i < 2.72) = \Phi(\frac{\ln 2.72 - 1.0}{0.8}) = 0.5003 \tag{5}$$

Uavhengighet medfører at antallet jordprøver W hvor $X_i < 2.72$ blir binomisk fordelt med parametere p og $n = 5$ og

$$P(W \geq 4) = P(W = 4) + P(W = 5) = 5p^4(1 - p) + p^5 = 0.1879 \tag{6}$$

e) Fra definisjon av forventningsverdi er

$$\mu = E(X) = \int_0^\infty x \frac{1}{\sqrt{2\pi}\tau x} e^{-\frac{1}{2\tau^2}(\ln x - \nu)^2} dx. \tag{7}$$

Innfører vi $u = (\ln x - \nu)/\tau$ som ny integrasjonsvariabel slik at $du = \frac{1}{\tau x} dx$ og $x = e^{\nu + \tau u}$ får vi

$$\begin{aligned} EX &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{\nu + \tau u - \frac{1}{2}u^2} du \\ &= e^\nu \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u^2 - 2\tau u)} du \\ &= e^\nu \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u^2 - 2\tau u + \tau^2 - \tau^2)} du \\ &= e^{\nu + \tau^2/2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u - \tau)^2} du \\ &= e^{\nu + \tau^2/2}. \end{aligned} \tag{8}$$

På samme måte finner vi

$$\begin{aligned} E(X^2) &= \int_{-\infty}^\infty x^2 \frac{1}{\sqrt{2\pi}\tau x} e^{-\frac{1}{2\tau^2}(\ln x - \nu)^2} dx \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{2\nu + 2\tau u - \frac{1}{2}u^2} du \\ &= e^{2\nu} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u^2 - 4\tau u)} du \\ &= e^{2\nu} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u^2 - 4\tau u + 4\tau^2 - 4\tau^2)} du \\ &= e^{2\nu + 2\tau^2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u - 2\tau)^2} du \\ &= e^{2\nu + 2\tau^2}. \end{aligned} \tag{9}$$

Dermed er

$$\sigma^2 = \mathrm{Var} X = E(X^2) - (EX)^2 = e^{2v}(e^{2\tau^2} - e^{\tau^2}). \tag{10}$$

Alternativt kan vi bruke at

$$EX = E(e^Y) = M_Y(1) = e^{\nu + \tau^2/2}, \tag{11}$$

og

$$E(X^2) = E((e^Y)^2) = E(e^{2Y}) = M_Y(2) = e^{2\nu + 2\tau^2}. \tag{12}$$

f) Om vi bruker parameteriseringen (μ, σ^2), har hver måling forventning $E(X_i) = \mu$ og

$$E(\hat{\mu}) = E(\frac{1}{n}\sum X_i) = \mu \tag{13}$$

slik at $\hat{\mu}$ er forventningsrett for μ. Variansen blir

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{n}\mathrm{Var}X_i = \frac{\sigma^2}{n} \tag{14}$$

For datasettet i oppgaven får vi estimatet

$$\hat{\mu} = \frac{1}{n}\sum x_i = 62.5 \tag{15}$$

g) Siden $Y_1, Y_2, \ldots, Y_n$ er uavhengig normalfordelte med forventning og varians ν og τ^2 er det kjent at SMEene av ν og τ^2 er

$$\hat{\nu} = \frac{1}{n}\sum Y_i, \tag{16}$$

og

$$\hat{\tau}^2 = \frac{1}{n}\sum (Y_i - \bar{Y})^2. \tag{17}$$

Det følger at $\hat{\nu}$ er normalfordelt, forventningsrett for ν med varians $\mathrm{Var}\hat{\nu} = \tau^2/n$. Videre er $E(\hat{\tau}^2) = \frac{n-1}{n}\tau^2$. Siden $S^2(n-1)/\tau^2 = \hat{\tau}^2 n/\tau^2$ er kji-kvadrat med $n-1$ frihetsgrader er $\mathrm{Var}(\hat{\tau}^2) = 2(n-1)\tau^4/n^2$.

Parameteren μ er en funksjon av ν og τ. SMEen av μ, μ^*, er samme funskjon av SMEene av ν og τ,

$$\mu^* = e^{\hat{\nu} + \hat{\tau}^2/2}. \tag{18}$$

For det oppgitte datasettet får vi

$$\hat{\nu} = 3.958, \qquad \hat{\tau}^2 = 0.3260, \qquad \mu^* = 61.54. \tag{19}$$

Forvetningsverdien til μ^* kan finnes på følgende måte. Siden $\hat{\nu}$ og $\hat{\tau}^2$ er uavhengige er

$$E(\mu^*) = Ee^{\hat{\nu}} Ee^{\hat{\tau}^2/2} \tag{20}$$

Fordi $\hat{\nu}$ er normalfordelt er $e^{\hat{\nu}}$ lognormal og

$$Ee^{\hat{\nu}} = e^{\nu + \tau^2/(2n)}. \tag{21}$$

Vi ser videre at

$$Ee^{\hat{\tau}^2/2} = Ee^{\frac{\tau^2}{2n}\cdot\frac{\hat{\tau}^2 n}{\tau^2}} = M_{\chi^2_{n-1}}\left(\frac{\tau^2}{2n}\right) = \left(1 - \frac{\tau^2}{n}\right)^{-\frac{n-1}{2}} \tag{22}$$

Dermed blir forventingsverdien til μ^* etter litt omskriving

$$\begin{aligned} E(\mu^*) &= \mu e^{-\frac{\tau^2}{2}+\frac{\tau^2}{2n}} \left(1 - \frac{\tau^2}{n}\right)^{-\frac{n-1}{2}} \\ &= \mu e^{-\frac{\tau^2}{2}+\frac{\tau^2}{2n}-(\frac{n}{2}-\frac{1}{2})\ln(1-\tau^2/n)} \\ &\approx \mu e^{-\frac{\tau^2}{2}+\frac{\tau^2}{2n}-(\frac{n}{2}-\frac{1}{2})(-\frac{\tau^2}{n}-\frac{\tau^4}{2n^2})} \\ &= \mu e^{-(\frac{n}{2}-\frac{1}{2})(-\frac{\tau^4}{2n^2})} \\ &\approx \mu e^{\frac{\tau^4}{4n}} \end{aligned} \tag{23}$$

for store n. Av dette ser vi at μ^* overestimerer μ men at forventningsfeilen går mot null når n går mot uendelig.

h) Gitt $\tau = \tau_0$ og sammenhengen $\mu = e^{\nu+\tau_0^2/2}$ er H_0, d.v.s., utsagnet

$$\mu \leq \mu_0 \tag{24}$$

ekvivalent med

$$e^{\nu+\tau_0^2/2} \leq \mu_0 \tag{25}$$

og

$$\nu \leq \ln \mu_0 - \tau_0^2/2, \tag{26}$$

d.v.s. H_0' om vi lar $\nu_0 = \ln \mu_0 - \tau_0^2/2$. På tilsvarende måte er H_1 ekvivalent med H_1'. Å teste H_0 mot H_1 er dermed ekvivalent med å teste H_0' mot H_1'.

For signifikansnivå valgt lik α er en rimelig test av H_0' mot H_1' å forkaste H_0' hvis testobservatoren

$$Z = \frac{\bar{Y} - \nu_0}{\tau_0/\sqrt{n}} > z_\alpha, \tag{27}$$

hvor z_α er α-kvantilen i standardnormalfordelingen.

i) Teststyrken blir

$$\begin{aligned}
P(Z > z_\alpha) &= P(\frac{\bar{Y} - \nu_0}{\tau_0 \sqrt{n}} > z_\alpha) \\
&= P(\bar{Y} > \nu_0 + z_\alpha \tau_0/\sqrt{n}) \\
&= P(\frac{\bar{Y} - \nu}{\tau_0/\sqrt{n}} > \frac{\nu_0 - \nu}{\tau_0/\sqrt{n}} + z_\alpha) \\
&= 1 - \Phi\left(\frac{\ln(\mu_0/\mu)}{\tau_0/\sqrt{n}} + z_\alpha\right) \\
&= \Phi\left(\frac{\sqrt{n}\ln\gamma}{\tau_0} - z_\alpha\right) \\
&= \Phi(0.878\ln(\gamma) - 1.64),
\end{aligned} \tag{28}$$

for parameterverdiene oppgitt i oppgaven.

Dersom vi ønsker at teststyrken skal være større enn 0.90 når $\mu = 1.5\mu_0$ ($\gamma = 1.5$) må

$$\Phi\left(\frac{\sqrt{n}\ln\gamma}{\tau_0} - z_\alpha\right) > 0.9 \tag{29}$$

og dermed

$$\frac{\sqrt{n}\ln\gamma}{\tau_0} - z_\alpha > z_{0.1} \tag{30}$$

slik at

$$n > \tau^2\left(\frac{z_\alpha + z_{0.1}}{\ln\gamma}\right)^2 = 0.36\left(\frac{1.64 + 1.28}{\ln 1.5}\right)^2 = 18.75 \tag{31}$$

som betyr at vi må ha $n \geq 19$.

j) Det følger at

$$P(-z_{\alpha/2} < \frac{\bar{Y} - \nu}{\tau_0/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha \tag{32}$$

slik at

$$P(\bar{Y} - z_{\alpha/2}\tau_0/\sqrt{n} < \nu < \bar{Y} + z_{\alpha/2}\tau_0/\sqrt{n}) = 1 - \alpha \tag{33}$$

og $\bar{Y} \pm z_{\alpha/2}\tau_0/\sqrt{n}$ et $(1-\alpha)$-konfidensintervall for ν.

Fra (33) følger at

$$P(e^{\bar{Y} - z_{\alpha/2}\tau_0/\sqrt{n} + \tau_0^2/2} < e^{\nu + \tau_0^2/2} < e^{\bar{Y} + z_{\alpha/2}\tau_0/\sqrt{n} + \tau_0^2/2}) = 1 - \alpha \tag{34}$$

slik at $e^{\bar{Y} \pm z_{\alpha/2}\tau_0/\sqrt{n} + \tau_0^2/2}$ er et $(1-\alpha)$-konfidensintervall for μ. For observasjonene gitt i oppgaven blir intervallet $(43.24, 90.96)$.