TMA4245 Statistikk
Eksamen juni 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsskisse

Oppgave 1

a) På figuren er det vanskelig åse noen trend for samsvarende verdier for de to variablene X og Y. Variablene kan se ut som uavhengige. Derfor vil korrelasjon være (tilnærmet) 0.

$$EX \approx 2,\ \sqrt{\text{Var}(X)} \approx 1,\ EY \approx 0,\ \sqrt{\text{Var}(Y)} \approx 1.$$

Oppgave 2

a) Ja, de kan være avhengige. Spesielt når

$$P(A \cap B) > 0, P(A \cap C) > 0, P(B \cap C) > 0$$

mens

$$P(A \cap B \cap C) = 0.$$

I dette tilfelle

$$P((A \cap B) \cap C) = 0 \neq P(A \cap B)P(C);$$

$$P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) = P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) =$$

$$= P(A \cap C) + P(B \cap C) = P(A)P(C) + P(B)P(C) = P(C)(P(A) + P(B)) =$$

$$= P(C)(P(A \cup B) + P(A \cap B)) \neq P(C)P(A \cup B).$$

Alternativt kan vi konstruere et eksempel. Kan bruke terningkast. Betrakt hendelsene

$$A = \{1, 2\},\ B = \{2, 3\},\ C = \{1, 3, 4\}.$$

Da er

$$P(A) = 1/3,\ P(B) = 1/3,\ P(C) = 1/2.$$

$$P(A \cap C) = 1/6 = P(A)P(C)$$

$$P(B \cap C) = 1/6 = P(B)P(C)$$

$$P((A \cap B) \cap C) = 0 \neq 1/12 = P(A \cap B)P(C)$$

$$P((A \cup B) \cap C) = 1/3 \neq 1/4 = P(A \cup B)P(C)$$

Hvis A og B er disjunkte, kan ikke $A \cup B$ og C være avhengige:

$$P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C) = P(A \cap C) + P(B \cap C) =$$

$$= P(A)P(C) + P(B)P(C) = P(C)(P(A) + P(B)) = P(A \cup B)P(C)$$

dvs $A \cup B$ og C er uavhengige.

Oppgave 3

a) $\sigma = 0.01$.

$\mu = 0.1$.

$$P(B) = P(|X - 0.1| > 2\sigma) = P\left(\left|\frac{X - 0.1}{\sigma}\right| > 2\right) =$$

$$= P(|Z| > 2) = 2P(Z < -2) = 2 \cdot 0.0228 = \underline{\underline{0.0456}}$$

$$P(A) = P(|X - 0.1| > \sigma) = P\left(\left|\frac{X - 0.1}{\sigma}\right| > 1\right) =$$

$$= P(|Z| > 1) = 2P(Z < -1) = 2 \cdot 0.1587 = 0.3174$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)}{P(A)} = \frac{0.0456}{0.3174} = \underline{\underline{0.1437}}$$

$\mu = 0.11$.

$$P(|X - 0.1| > 2\sigma) = P(X - 0.1 < -2\sigma)P(X - 0.1 > 2\sigma) =$$

$$= P(X - 0.11 < -2\sigma - 0.01) + P(X - 0.11 > 2\sigma - 0.01) =$$

$$= P\left(Z < -2 - \frac{0.01}{\sigma}\right) + P\left(Z > 2 - \frac{0.01}{\sigma}\right) = P(Z < -3) + P(Z > 1) =$$

$$= P(Z < -3) + P(Z < -1) = 0.0013 + 0.1587 = \underline{\underline{0.16}}$$

b) Rimelighetsfunksjonen blir

$$L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right) = (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)\right).$$

Logaritme

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Deriverer (mhp σ^2)

$$\frac{\partial \ln L}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Løser ligningen

$$\frac{\partial \ln L}{\partial(\sigma^2)} = 0$$

. Løsningen

$$\underline{\underline{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}}$$

er SME (en positiv løsning og $L(\sigma^2) \to 0$ når $\sigma^2 \to 0$ og når $\sigma^2 \to \infty$, derfor er det maksimum og ikke minimum; alternativt kan man derivere en gang til og vise at den andre deriverte er negativ).

c) Forventningsverdien til en χ^2-fordeling med n frihetsgrader er lik n (se "Tabeller og formler i statistikk"). Så,

$$E\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\right] = n,\ E\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\right] = n-1$$

eller

$$E\hat{\sigma}^2 = \sigma^2,\ ES^2 = \sigma^2$$

dvs estimatorene er forventningsrette. Igjen, ved bruk av "Tabeller og formler i statistikkfår vi at

$$\text{Var}\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) = 2n,\ \text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

som impliserer

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n},\ \text{Var}(S^2) = \frac{2\sigma^4}{n-1}.$$

$\hat{\sigma}^2$ er mer effektiv enn S^2 (har mindre varians). Det er fornuftig å bruke den.

d) I generelt tilfelle får man $(1-\alpha)$konfidensintervall på følgende måte

$$1-\alpha = P\left(\chi^2_{1-\alpha/2,n} \le \frac{n\hat{\sigma}^2}{\sigma^2} \le \chi^2_{\alpha/2,n}\right) = P\left(\frac{n\hat{\sigma}^2}{\chi^2_{\alpha/2,n}} \le \sigma^2 \le \frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n}}\right)$$

dvs $(1-\alpha)$konfidensintervall er

$$\left[\frac{n\hat{\sigma}^2}{\chi^2_{\alpha/2,n}}, \frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n}}\right].$$

For tallene som er gitt:

$$n\hat{\sigma}^2 = 0.0018,\ \chi^2_{0.05,20} = 31.410,\ \chi^2_{0.95,20} = 10.851.$$

Intervallet blir $\underline{\underline{[0.000057, 0.000166]}}$.

Oppgave 4

a) La $F(x)$ og $F_T(t)$ være kumulative fordelingsfunksjoner for X_i og T, henholdvis. Da er

$$F(x) = \int_0^x \frac{1}{\mu}e^{-u/\mu}d\mu = 1 - e^{-x/\mu},\ x > 0$$

og

$$F_T(t) = P(\min(X_1, ..., X_n) \leq t) = 1 - P(\min(X_1, ..., X_n) > t) =$$
$$= 1 - P(X_1 > t, ..., X_n > t) = 1 - P(X_1 > t) \cdot ... \cdot P(X_n > t) =$$
$$= 1 - (1 - P(X_1 \leq t)) \cdot ... \cdot (1 - P(X_n \leq t)) = 1 - (1 - F(t))^n = 1 - e^{-tn/\mu}.$$

Tilsvarende sannsynlighetstetthet er den deriverte av $F_T(t)$:

$$f_T(t) = F'_T(t) = \underline{\underline{\frac{n}{\mu} e^{-tn/\mu}}}.$$

En ser at T er eksponesialfordelt med parameter μ/n, derfor er $E(T) = \underline{\underline{\mu/n}}$ og $\text{Var}(T) = \underline{\underline{\mu^2/n^2}}$.

b)

$$\alpha = P_{\mu=1}(T \leq c_1) = 1 - e^{-nc_1},$$

derfor er

$$\underline{\underline{c_1 = \frac{1}{n} \ln \frac{1}{1 - \alpha}}}.$$

c) Jo mindre μ er, desto mindre er $\bar{X}$ i gjennomsnitt. Derfor er det fornuftig å forkaste H_0 for små verdier av $\bar{X}$ (fordi nullhypotesa $H_0 : \mu = 1$ testes mot alternativ $H_1 : \mu < 1$). Således har forkastningsområdet formen $\underline{\underline{\bar{X} \leq c_2}}$. For å finne c_2 (tilnærmet) bruker vi sentralgrenseteoremet:

$$\frac{\bar{X} - E\bar{X}}{\sqrt{\text{Var}\bar{X}}} \sim N(0, 1).$$

Siden $E\bar{X} = \mu$ og $\text{Var}\bar{X} = \mu^2/n$, har vi at under H_0

$$\sqrt{n}(\bar{X} - 1) \sim N(0, 1).$$

Da

$$\alpha = P(\bar{X} \leq c_2) = P(\sqrt{n}(\bar{X} - 1) \leq \sqrt{n}(c_2 - 1)) = P(Z \leq \sqrt{n}(c_2 - 1))$$

og derfor

$$\sqrt{n}(c_2 - 1) = -z_\alpha$$

eller

$$\underline{\underline{c_2 = 1 - \frac{z_\alpha}{\sqrt{n}}}}.$$

d) Teststyrken til Test 1 er

$$1 - \beta_1(\mu) = P_\mu \left(T \leq \frac{1}{n} \ln \frac{1}{1 - \alpha} \right) =$$
$$= 1 - e^{[\ln(1-\alpha)]/\mu} = 1 - (1 - \alpha)^{1/\mu}.$$

Teststyrken til Test 2 er

$$1 - \beta_2(\mu) = P_\mu \left(\bar{X} \leq 1 - \frac{z_\alpha}{\sqrt{n}} \right) = P_\mu \left(\sqrt{n} \frac{\bar{X} - \mu}{\mu} \leq \frac{\sqrt{n}}{\mu} \left(1 - \mu - \frac{z_\alpha}{\sqrt{n}} \right) \right) =$$

$$= P_\mu\left(Z \leq \frac{\sqrt{n}}{\mu}\left(1 - \mu - \frac{z_\alpha}{\sqrt{n}}\right)\right) = \Phi\left(\frac{\sqrt{n}}{\mu} - \sqrt{n} - \frac{z_\alpha}{\mu}\right).$$

Test 2 er best fordi den har størst styrke dvs minst sannsynlighet av Type 2 feil mens signifikansnivå (sannsynlighet av Type 1 feil) er det samme for de to testene. For eksempel, for $\alpha = 0.05$, $n = 30$ og $\mu = 0.8$, er styrkene $1 - \beta_1 = \underline{\underline{0.06}}$, $1 - \beta_2 = \underline{\underline{0.25}}$.

Test 1 er en dårlig test fordi styrken er uavhengig av n dvs sannsynligheten for Type 2 feil avtar ikke når utvalgsstørrelsen vokser. En ser også at teststyrken for Test1 vil være svært liten for alle relevante verdier på α og μ.