Petter Mostad
Tillämpad Matematik och Statistik
Chalmers

Lösningsförslag för
MVE055 / MSG810 Matematisk statistik och diskret matematik
Omtenta 20 december 2016

1. Vi får

$$\begin{aligned} F_Z(x) &= P(Z \leq x) = P(\max(X_1, \dots, X_n) \leq x) \\ &= P((X_1 \leq x) \cap \dots \cap (X_n \leq x)) = P(X_1 \leq x) \cdot \dots \cdot P(X_n \leq x) = F_X(x)^n \end{aligned}$$

   och motsvarande

$$\begin{aligned} F_Y(x) &= P(Y \leq x) = P(\min(X_1, \dots, X_n) \leq x) \\ &= 1 - P(\min(X_1, \dots, X_n) > x) = 1 - P((X_1 > x) \cap \dots \cap (X_n > x)) \\ &= 1 - P(X_1 > x) \cdot \dots \cdot P(X_n > x) = 1 - (1 - F_X(x))^n \end{aligned}$$

2. De möjliga värdena till Y är $1/k$ för $k = 1, 2, \dots$. Vi har frekvensfunktionen $f_Y(1/k) = p(k) = (1/2)^k$, därmed har vi, för alla $y \in [1/k, 1/(k-1))$ att

$$F_Y(y) = P(Y \leq y) = \sum_{n=k}^{\infty} (1/2)^n = \frac{(1/2)^k}{1 - 1/2} = (1/2)^{k-1}$$

   och detta anger fördelningsfunktionen (cdf).

3. Eftersom $X_1, \dots, X_n \stackrel{iid}{\sim} N(\theta, \theta^2)$ så får vi för estimatorn $\hat{\theta} = \overline{X}$ att

$$\overline{X} \sim N(\theta, \theta^2/n).$$

   Därmed får vi även att

$$\frac{\overline{X} - \theta}{\sqrt{\theta^2/n}} \sim N(0, 1).$$

   Detta ger följande ekvivalenta ekvationer: (Märk att vi förutsätter att $n \geq 4$, så att $1 -$

$1.96/\sqrt{n} > 0$ i den sista linjen)

$$\begin{aligned}
Pr(-1.96 \leq \frac{\overline{X}-\theta}{\sqrt{\theta^2/n}} \leq 1.96) &= 0.95 \\
Pr(-1.96 \leq \sqrt{n}\frac{\overline{X}-\theta}{\theta} \leq 1.96) &= 0.95 \\
Pr(-1.96/\sqrt{n} \leq \frac{\overline{X}-\theta}{\theta} \leq 1.96/\sqrt{n}) &= 0.95 \\
Pr(-1.96/\sqrt{n} \leq \frac{\overline{X}}{\theta} - 1 \leq 1.96/\sqrt{n}) &= 0.95 \\
Pr(1 - 1.96/\sqrt{n} \leq \frac{\overline{X}}{\theta} \leq 1 + 1.96/\sqrt{n}) &= 0.95 \\
Pr\left(\frac{\overline{X}}{1+1.96/\sqrt{n}} \leq \theta \leq \frac{\overline{X}}{1-1.96/\sqrt{n}}\right) &= 0.95.
\end{aligned}$$

Detta betyder att

$$\left[\frac{\overline{X}}{1+1.96/\sqrt{n}}, \frac{\overline{X}}{1-1.96\sqrt{n}}\right]$$

är ett 95%-ig konfidensintervall för θ. Märk att man kan också få en enklare approximativ lösning genom att approximera

$$\frac{\overline{X}-\theta}{\sqrt{\overline{X}^2/n}} \sim N(0,1)$$

som ger ett approximativt 95%-ig konfidensintervall

$$\overline{X} \pm 1.96 \cdot \frac{\overline{X}}{\sqrt{n}}.$$

4. (a) Vi har

$$E[X] = \int_0^1 x \cdot 2x\,dx = \left[\frac{2}{3}x^3\right]_0^1 = \frac{2}{3}$$

och

$$E[X^2] = \int_0^1 x^2 \cdot 2x\,dx = \left[\frac{2}{4}x^4\right]_0^1 = \frac{1}{2}.$$

Därmed får vi

$$Var[X] = E[X^2] - E[X]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

.

(b) $V = \frac{4}{3}\pi X^3$, sï¿½

$$E[V] = \frac{4}{3}\pi E[X^3] = \frac{4}{3}\pi \int_0^1 x^3 \cdot 2x\,dx = \frac{4}{3}\pi \left[\frac{2}{5}x^5\right]_0^1 = \frac{4}{3}\pi\frac{2}{5} = \frac{8\pi}{15}.$$

5. (a) Definitionen är $m_X(t) = E[e^{tX}]$.

   (b) Vi har $f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ och därmed

$$m_X(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \cdot e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^t\lambda)^x}{x!} = e^{-\lambda}\exp(e^t\lambda) = \exp(\lambda(e^t - 1))$$

   (c) Vi får

$$m_{\overline{X}}(t) = \prod_{i=1}^{n} m_{X_i}(t/n) = \prod_{i=1}^{n}\exp(\lambda_i(e^{t/n} - 1)) = \exp\left(\left(\sum_{i=1}^{n}\lambda_i\right)(e^{t/n} - 1)\right).$$

6. För Louisiana-farmen beräknar vi ett medelvärde på $\overline{x} = 14.57143$ och en stickprovsvarians på $s_x^2 = 2.612381$, medan för Arizona-farmen beräknar vi medelvärdet $\overline{y} = 12.94$ och stickprovsvariansen $s_y^2 = 2.171556$. Vi antar att fångsterna är fördelade enligt normalfördelningarna $N(\mu_x, \sigma^2)$ respektive $N(\mu_y, \sigma^2)$. Vi använder $H_0 : \mu_x = \mu_y$ och $H_1 : \mu_x \neq \mu_y$. Den poolade variansen blir

$$s_p^2 = \frac{(7-1)s_x^2 + (10-1)s_y^2}{7+10-2} = \frac{6 \cdot 2.612381 + 9 \cdot 2.171556}{15} = 2.347886.$$

   Teststatistikan blir

$$\frac{\overline{x} - \overline{y}}{\sqrt{s_p^2(1/7 + 1/10)}} = 2.169501$$

   Förkastningsområdet är alla värden utanför intervallet $[T_{0.025}, T_{0.975}]$, där T har en T-fördelning med $7+10-2 = 15$ frihetsgrader: Enligt tabellen blir intervallet $[-2.131, 2.131]$. Eftersom test-statistikan inte ligger i detta intervallet förkastas nollhypotesen.

7. (a) Definera två stokastiska variabler A och B med $A \sim \exp(1)$ och $B \sim \exp(2)$, och kom i håg att för dessa är fördelningsfunktionerna då $F_A(t) = 1 - e^{-t}$ och $F_B(t) = 1 - e^{-t/2}$. För fördelningsfunktionen till X får vi

$$F_X(t) = Pr(X \leq t) = \frac{1}{2}Pr(A \leq t) + \frac{1}{2}Pr(B \leq t) = \frac{1}{2}(1-e^{-t}) + \frac{1}{2}(1-e^{-t/2}) = 1 - \frac{e^{-t} + e^{-t/2}}{2}$$

   (b) Här använder vi Bayes formel:

$$Pr(\text{typ A} \mid X > t) = \frac{Pr(X > t \mid \text{typ A})Pr(\text{typ A})}{Pr(X > t)} = \frac{e^{-t} \cdot \frac{1}{2}}{(e^{-t} + e^{-t/2})/2} = \frac{1}{1 + e^{t/2}}$$

8. (a) Vi får

$$\begin{aligned} 1 &= \int_0^1 \int_0^x f(x,y)\,dy\,dx = \int_0^1 \int_0^x c(x-y)\,dy\,dx \\ &= \int_0^1 c\left[xy - \frac{y^2}{2}\right]_0^x = c\int_0^1 (x^2 - \frac{1}{2}x^2)\,dx \\ &= \frac{c}{2}\int_0^1 x^2\,dx = \frac{c}{2}\left[\frac{1}{3}x^3\right]_0^1 = \frac{c}{6} \end{aligned}$$

och därmed $c = 6$.

(b) Vi får

$$f_X(x) = \int_0^x f(x,y)\,dy = \int_0^x 6(x-y)\,dy = 6\left[xy - \frac{1}{2}y^2\right]_0^x = 6\left(x^2 - \frac{1}{2}x^2\right) = 3x^2$$

och

$$f_Y(y) = \int_y^1 f(x,y)\,dx = \int_y^1 6(x-y)\,dx = 6\left[\frac{1}{2}x^2 - xy\right]_y^1 = 6\left(\frac{1}{2} - y - \frac{1}{2}y^2 + y^2\right) = 3(1-y)^2$$

(c)

$$\begin{aligned} P(X \le 2Y) &= \int_0^1 \int_0^x I(x \le 2y) f(x,y)\,dy\,dx \\ &= \int_0^1 \int_{x/2}^x 6(x-y)\,dy\,dx \\ &= 6\int_0^1 \left[xy - \frac{1}{2}y^2\right]_{x/2}^x\,dx \\ &= 6\int_0^1 x^2 - \frac{1}{2}x^2 - \frac{1}{2}x^2 + \frac{1}{8}x^2\,dx \\ &= 6\int_0^1 \frac{1}{8}x^2\,dx = \frac{6}{8}\left[\frac{1}{3}x^3\right]_0^1 = \frac{2}{8} = \frac{1}{4} \end{aligned}$$

9. Definera $Y = X_1 + X_3 - 3X_2$. Då får vi

$$E(Y) = E(X_1) + E(X_3) - 3E(X_2) = 1 + 1 - 3 = -1$$

och

$$Var(Y) = Var(X_1) + Var(X_3) + 9Var(X_2) = 1 + 1 + 9 = 11$$

Eftersom Y är en linjärkombination av normalfördelade variabler får vi då även att

$$Y \sim N(-1, 11)$$

Nu kan vi beräkna, om vi låter Z vara en variabel med standard normalfördelning,

$$\begin{aligned}
P(X_1 + X_3 \geq 3X_2) &= P(X_1 + X_3 - 3X_2 \geq 0)) = P(Y \geq 0) \\
&= P\left(\frac{Y+1}{\sqrt{11}} \geq \frac{1}{\sqrt{11}}\right) = P\left(Z \geq \frac{1}{\sqrt{11}}\right) = P(Z \geq 0.30) \\
&= 1 - P(Z \leq 0.30) = 1 - 0.6179 = 0.3821
\end{aligned}$$

10. Genom att använda formlerna för parametrarna i regressionslinjen får vi

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{24 \cdot 12518.69 - 21.873 \cdot 8965}{24 \cdot 29.51779 - 21.873^2} = 453.73$$

och

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \frac{8965}{24} - 453.73\frac{21.873}{24} = -39.98$$

Vi vill sen testa $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. Vi beräknar

$$\begin{aligned}
S_{xx} &= \frac{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}{n} = \frac{24 \cdot 29.51779 - 21.873^2}{24} = 9.58329 \\
S_{yy} &= \frac{n\sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}{n} = \frac{24 \cdot 6516925 - 8965^2}{24} = 3168124 \\
S_{xy} &= \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n} = \frac{24 \cdot 12518.69 - 21.873 \cdot 8965}{24} = 4348.218 \\
\text{SSE} &= S_{yy} - \hat{\beta}_1 S_{xy} = 3168124 - 453.73 \cdot 4348.218 = 1195207 \\
S^2 &= \text{SSE}/(n-2) = 54327.59
\end{aligned}$$

För teststatistikan får vi

$$T = \frac{\hat{\beta}_1}{S/\sqrt{S_{xx}}} = \frac{453.73}{\sqrt{54327.59/9.58329}} = 6.03$$

som skall jämföras med en T-fördelning med $24 - 2 = 22$ frihetsgrader. Enligt tabellen förkastar vi utanför intervallet [-2.074, 2.074], så H_0 blir förkastad. Från tabellen får vi även att p-värdet blir mycket litet, mindre än 0.001.