Lösningar till Matematisk statistik för Z, del A den 27/10-01

1a) $1-(0.028+0.072-0.002) = 0.902$, b) Ja, ty med 3 decimalers noggranhet gäller $0.028 \cdot 0.072 = 0.002$. c) $0.002/0.072 = 0.028$ med 3 decimalers noggranhet.

2a) Utfallsrummet måste vara $S = \{1, 2, \ldots\}$ och det gäller att konstatera att $f(k) \geq 0$ för $k \in S$ och att $1 = \sum_{k \in S} f(k)$. Det första faktumet fås av att både $p > 0$ och $1 - p > 0$. Det andra gäller eftersom

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = \frac{p}{1-(1-p)} = 1$$

b) Observera först att $P(X > 0) = 1$. För $k \geq 1$ gäller

$$P(X > k) = 1 - P(X \leq k) = 1 - \sum_{l=1}^{k} p(1-p)^{l-1} = 1 - \frac{p(1-(1-p)^k)}{1-(1-p)} = (1-p)^k$$

I både a och b har jag använt mig av fakta från stycket "Geometrisk serie" i formelsamlingen.

3) Om T är processorns livslängd, så är $T \sim \mathrm{Exp}(10\,000)$ (se stycket "Exponentialfördelningen" i formelsamlingen) och a)

$$p = P(T > 20000) = \int_{20000}^{\infty} \frac{1}{10000} e^{-x/10000}\, dx = \int_{2}^{\infty} e^{-t}\, dt = e^{-2} \approx 0.135$$

(gör substitutionen $t = x/10000$). Den i b) sökta sannolikheten är $p + p - p^2 = 2e^{-2} - e^{-4} \approx 0.252$.

4) Tätheten $f_{X,Y}(x,y) > 0$ på triangeln $S = \{(x,y) : 27 \leq y \leq x \leq 33\}$ och a) kravet att volymen under $f_{X,Y}(x,y)$ ska vara 1 leder därför till

$$1 = \int_{27}^{33} \int_{27}^{x} (c/x)\, dy\, dx = c \int_{27}^{33} (x-27)/x\, dx = c \int_{27}^{33} (1 - 27/x)\, dx = c(6 - 27(\ln 33 - \ln 27))$$

(ty $27 \leq y \leq x$ då x är fixt) så $c = 1/(6 - 27 \ln(33/27)) \approx 1.72$ vilket skulle visas. b) X:s sannolikhetstäthet fås genom att integrera ut y för varje fixt x:

$$f_X(x) = \int_{27}^{x} (c/x)\, dy = c(x-27)/x = c(1 - 27/x), \quad 27 \leq x \leq 33$$

ty $27 \leq y \leq x$. Analogt fås Y:s sannolikhetstäthet så här:

$$f_Y(y) = \int_{y}^{33} (c/x)\, dx = c(\ln 33 - \ln y) = c \ln(33/y), \quad 27 \leq y \leq 33$$

ty då y är fixt är $y \leq x \leq 33$. c) Se Milton & Arnolds uträkning på sida 162. Men man kan också räkna ut den sökta sannolikheten så här:

$$P(X \leq 30, Y \leq 28) = P(X \leq 28) + P(28 < X \leq 30, Y \leq 28)$$

$$= \int_{27}^{28} f_X(x)\, dx + \int_{28}^{30} \int_{27}^{28} f_{X,Y}(x,y)\, dy\, dx = \ldots$$

$$= c(1 - 27 \ln 28 + 27 \ln 27) + c(\ln 30 - \ln 28) = c(1 + \ln 30 + 27 \ln 27 - 28 \ln 28) \approx 0.15$$

5) Trolighetsfunktionen är

$$L(\lambda) = \lambda e^{-\lambda x_1} \cdot \ldots \cdot \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_i x_i}$$

Dess logaritm är

$$\mathcal{L}(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum_i x_i$$

och genom att lösa ut λ ur $\mathcal{L}'(\lambda) = 0$ ser vi att $\mathcal{L}(\lambda)$ (och $L(\lambda)$) antar sitt maximum då $\lambda = n/\sum_i x_i = 1/\bar{x}$. Trolighetsskattningen av λ är därför $\widehat{\lambda} = 1/\bar{X}$. Den är inte väntevärdesriktig, eftersom

$$E\left[1/\bar{X}\right] \neq 1/E[\bar{X}] = 1/(1/\lambda) = \lambda$$

ty $X \sim \text{Exp}(\beta)$ med $\beta = E[X] = 1/\lambda$.

6a) Vännen menar att du ska testa nollhypotesen $H_0 : p \leq 0.2$ mot alternativhypotesen $H_1 : p > 0.2$. b) Testets nivå är sannolikheten att felaktigt förkasta en sann nollhypotes och man ska räkna på värsta fallet som inträffar då $p = 0.2$. Nivån är därför

$$\alpha = P(X \geq 16 \,|\, p = 0.2) \approx P\left(Z \geq \frac{15.5 - 50 \cdot 0.2}{\sqrt{50 \cdot 0.2 \cdot 0.8}}\right) = P(Z \geq 1.94) \approx 0.026$$

där $X \sim \text{Bin}(50, p)$ och $Z \sim \text{N}(0, 1)$. Normalapproximationen är tillåten eftersom $50 \cdot 0.2 = 19 \geq 5$. c) Den sökta sannolikheten är

$$P(X \geq 16 \,|\, p = 0.25) \approx P\left(Z \geq \frac{15.5 - 50 \cdot 0.25}{\sqrt{50 \cdot 0.25 \cdot 0.75}}\right) = P(Z \geq 0.98) \approx 0.064$$

Även här är naturligtvis normalapproximationen OK.

7) Antalet observationer är $n = 7$. Räkna ut medelvärde $\bar{x} = 12.2$ och standardavvikelse $s = 1.686$ samt hämta i tabell över t-fördelningen $t_{0.025} = 2.447$ ($n - 1 = 6$ frihetsgrader). Konfidensintervallet härleds ur

$$0.95 = P\left(-t_{0.025} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0.025}\right)$$

(j f r stycket "Ett normalfördelat stickprov" i formelsamlingen) till att bli

$$\mu = \bar{x} \pm t_{0.025} s/\sqrt{n} \quad (0.95)$$

Instoppning av siffervärden ger att

$$\mu = 12.2 \pm 1.56 \quad (0.95)$$

8) Vi skattar p_A och p_B väntevärdesriktigt enl

$$\widehat{p}_A = \frac{x_A}{n_A} \approx 0.154, \quad \widehat{p}_B = \frac{x_B}{n_B} \approx 0.121$$

Konfidensintervall härleds ur

$$\frac{\widehat{p}_A - \widehat{p}_B - (p_A - p_B)}{\sqrt{\widehat{p}_A(1 - \widehat{p}_A)/n_A + \widehat{p}_B(1 - \widehat{p}_B)/n_B}} \overset{\text{ap}}{\sim} \text{N}(0, 1)$$

(j f r stycket "Jämförelse av proportioner" i formelsamlingen). Vi får

$$p_A - p_B = \widehat{p}_A - \widehat{p}_B \pm z_{\alpha/2}\sqrt{\widehat{p}_A(1 - \widehat{p}_A)/n_A + \widehat{p}_B(1 - \widehat{p}_B)/n_B} \quad (\approx 1 - \alpha)$$

Insättning av siffervärden (jag använde $z_{0.025} = 2$, $z_{0.025} = 1.96$ är naturligtvis också tillåtet) ger

$$p_A - p_B = 0.033 \pm 0.035 \quad (\approx 0.95)$$

eller $p_A - p_B \in (-0.002,\ 0.068)$ med ca 95% konfidens. Det kan inte anses statistiskt säkerställt på nivån 5% att $p_A \neq p_B$ eftersom $0 \in$ konfidensintervallet.