Lösningar till Matematisk statistik för Z, del A den 29/10-02

1. (a) Nej, ty $P(A|X=1) \neq P(A|X=2)$.

   (b) Ur formelsamling (se stycket om Binomialfördelningen) fås först att

$$P(X=0) = \left(\frac{2}{3}\right)^2 = \frac{4}{9}, \quad P(X=1) = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}, \quad P(X=2) = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$$

   Genom att dela upp händelsen A i de tre disjunkta bitarna $X=0$, $X=1$ och $X=2$, fås sedan

$$\begin{aligned} P(A) &= P(X=0, A) + P(X=1, A) + P(X=2, A) \\ &= P(X=0)P(A|X=0) + P(X=1)P(A|X=1) + P(X=2)P(A|X=2) \\ &= \frac{4}{9} \cdot \frac{1}{4} + \frac{4}{9} \cdot \frac{1}{4} + \frac{1}{9} \cdot 1 = \frac{1}{3} \end{aligned}$$

   (c)

$$P(X=2|A) = \frac{P(X=2, A)}{P(A)} = \frac{P(X=2)P(A|X=2)}{P(A)} = \frac{\frac{1}{9} \cdot 1}{\frac{1}{3}} = \frac{3}{9} = \frac{1}{3}$$

2. Låt N_t beteckna antalet impulser i tidsintervallet $[0, t]$. Då gäller, för varje $t \geq 0$, att $T > t \Leftrightarrow N_t = 0$. Detta ger $P(T > t) = P(N_t = 0) = e^{-\lambda t}$. Så T:s fördelningsfunktion är $F_T(t) = 1 - e^{-\lambda t}$ och medelst derivering följer att T:s täthet är $f_T(t) = \lambda e^{-\lambda t}$. Detta är tätheten för en Exponentialfördelning med väntevärdet $\beta = 1/\lambda$ (j.f.r formelsamlingen eller Beta), vilket skulle visas.

3. Låt $X \sim \mathrm{N}(5.5,\ 2.2)$ vara ett typiskt mätresultat och låt N vara antalet observationer i intervallet $[3.3,\ 7.7]$. Då är $N \sim \mathrm{Bin}(3, p)$, där $p = P(X \in [3.3,\ 7.7]) = P(Z \in [-1,\ 1]) = 0.6826$ (obs $Z \sim \mathrm{N}(0,1)$). Den sökta sannolikheten är därför

$$P(N \geq 2) = P(N=2) + P(N=3) = 3p^2(1-p) + p^3 \approx 0.76$$

4. (a) För $t > 0$ får vi

$$f_T(t) = \int_{-\infty}^{\infty} f(x,t)\,dx = \int_0^t \frac{1}{t}\,e^{-t}\,dx = e^{-t}$$

   (b) Låt $H(x,t) = x$. Då

$$\begin{aligned} E[X] = E[H(X,T)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,t) f(x,t)\,dx\,dt \\ &= \int_0^{\infty}\int_0^t x\,\frac{1}{t}\,e^{-t}\,dx\,dt = \frac{1}{2}\int_0^{\infty} e^{-t}\,dt = \frac{1}{2} \end{aligned}$$

5. (a) Det gäller att visa att $E[S^2] = \sigma^2$. Vi behöver använda att

$$(n-1)S^2 = \sum_i X_i^2 - n\bar{X}^2$$

   Linjäriteten hos väntevärdesoperatorn ger då direkt

$$E[(n-1)S^2] = \sum_i E[X_i^2] - nE[\bar{X}^2]$$

   Notera

$$\begin{aligned} E[X_i^2] &= \mathrm{Var}[X_i] + E[X_i]^2 = \sigma^2 + \mu^2 \\ E[\bar{X}^2] &= \mathrm{Var}[\bar{X}] + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

   Härur fås

$$E[(n-1)S^2] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2$$

   och $E[S^2] = \sigma^2$ följer.

(b) Det gäller att visa att $E[\bar{X}] = \mu$. Så här går det till:

(c)

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[X_i\right] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Den 2:a likheten följer av variansen är ett kvadratiskt mått och den 3:e av oberoendet.

(d) Låt $\epsilon = k\sigma/\sqrt{n}$ i Chebyshevs olikhet. Då

$$P(|\bar{X} - \mu| > \epsilon) = P(|\bar{X} - \mu| > k\sigma/\sqrt{n}) \leq \frac{1}{k^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$$

då $n \to \infty$. Enl vad vi just visat är ju medelvärdets väntevärde lika med μ och standardavvikelse lika med $\sigma/\sqrt{n}$.

6. (a) Man ska testa nollhypotesen $H_0 : p \geq 0.05$ mot alternativet $H_1 : p < 0.5$ (alternativet ska ju vara det man vill visa). Man ska förkasta då $f \leq c$, ty ju mindre f är, desto mer tyder det på att p är litet, och kravet $\alpha = P(f \leq c|p = 0.05)$, leder till att $f \leq c$ är ekvivalent med att kräva att $\frac{f-10}{\sqrt{9.5}} \leq -z_\alpha$. Detta är naturligtvis samma sak som att kräva att $\frac{10-f}{\sqrt{9.5}} \geq z_\alpha$ eller att $f \leq 10 - z_\alpha\sqrt{9.5}$, där z_α ges av att $P(Z > z_\alpha) = \alpha$ för $Z \sim \mathrm{N}(0, 1)$. Ty då $p = 0.05$ är f enl c.g.s approximativt normalfördelad med väntevärde $200 \cdot 0.05 = 10$ och standardavvikelse $\sqrt{200 \cdot 0.05 \cdot 0.95} = \sqrt{9.5}$

(b) $f = 4 \Rightarrow \frac{10-f}{\sqrt{9.5}} = 1.94$. I tabell över normalfördelningskvantiler ses att $z_{0.05} = 1.645$ $\Rightarrow H_0 : p \geq 0.05$ förkastas. På nivån 0.05 kan det alltså anses säkerställt att $p < 0.05$. Processen behöver då ej justeras.

7. Konfidensintervallet bör beräknas ur faktumet $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, som gäller exakt om mätvärdena är normalfördelade, annars approximativt enl c.g.s. Man får då

$$\mu = \bar{x} \pm t_{29,0.025} S/\sqrt{n} = 63.42 \pm 2.04523 \cdot 10.451/\sqrt{30} = 63.42 \pm 3.90$$

eller $\mu \in [59.52,\ 67.32]$ med konfidensen ca 95%.

8. (a) Lika varians innebär att vi kan väga ihop variansskattningarna till

$$s_p^2 = \frac{29 \cdot 10.451^2 + 31 \cdot 9.378^2}{60} = 98.231 \;\Rightarrow\; s_p = 9.911$$

Konfidensintervallet bör beräknas ur faktumet $\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$, där indexeringen refererar till de olika mätserierna och som gäller exakt om mätvärdena är normalfördelade, annars som ovan approximativt enl c.g.s. Man får då

$$\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2 \pm t_{60,0.05} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 63.42 - 59.34 \pm 1.67065 \cdot 9.911\sqrt{\frac{1}{30} + \frac{1}{32}} = 4.08 \pm 4.21$$

eller $\mu \in [-0.13,\ 8.29]$ med konfidensen ca $\approx 90\%$.

(b) Nej, ty $0 \in [-0.13,\ 8.29]$ gäller med konfidensen ca 90%.