Matematisk statistik
Matematikcentrum
Lunds tekniska högskola
Lunds universitet

Lösning till tentamen: 2009–01–08 kl 08^{00}–13^{00}
FMS 012 — Matematisk statistik för I, CDE, FΠN 9 hp
MAS B03 — Matematisk statistik för fysiker, 9 hp
(Även FMS022, FMS121, FMS233 för CDE, I, resp. fysiker)

---

1. Inför händelse-beteckningarna

$$A = \text{"rådjuret äter upp en given tulpan"}$$
$$G = \text{"en tulpan är gul"},\quad R = \text{"en tulpan är röd"},\quad S = \text{"en tulpan är svart"}$$

(a) Vi har givet färgsannolikheterna $\mathsf{P}(G) = 0.3$, $\mathsf{P}(R) = 0.5$, och $\mathsf{P}(S) = 0.2$, samt $\mathsf{P}(A \mid G) = 1.0$, $\mathsf{P}(A \mid R) = 0.6$, och $\mathsf{P}(A \mid S) = 0.1$. Sannolikheten att rådjuret äter upp en tulpan den träffar på är då, enligt satsen om total sannolikhet, (4p)

$$\begin{aligned}\mathsf{P}(A) &= \mathsf{P}(A \mid G)\,\mathsf{P}(G) + \mathsf{P}(A \mid R)\,\mathsf{P}(R) + \mathsf{P}(A \mid S)\,\mathsf{P}(S)\\ &= 1 \cdot 0.3 + 0.6 \cdot 0.5 + 0.1 \cdot 0.2 = 0.62.\end{aligned}$$

(b) Inför beteckningen T = "rådjuret träffar på en tulpan". Nu är det inte säkert att rådjuret träffar på tulpanen, och vi noterar att svaret i (a) i själva verket är den *betingade* sannolikheten $\mathsf{P}(A \mid T) = 0.62$. Från uppgiften vet vi att $\mathsf{P}(T) = 0.8$. Den sökta sannolikheten är nu (6p)

$$\mathsf{P}(T \mid \text{tulpanen står kvar}) = \frac{\mathsf{P}(T \,\&\, \text{tulpanen står kvar})}{\mathsf{P}(\text{tulpanen står kvar})} = \frac{\mathsf{P}(\text{tulpanen står kvar} \mid T)\,\mathsf{P}(T)}{\mathsf{P}(\text{tulpanen står kvar})}$$

Att tulpanen *inte* står kvar kan bara inträffa om rådjuret dök upp, och dessutom åt upp tulpanen, och vi får

$$\mathsf{P}(T \mid \text{tulpanen står kvar}) = \frac{(1 - \mathsf{P}(A \mid T))\,\mathsf{P}(T)}{1 - \mathsf{P}(T)\,\mathsf{P}(A \mid T)} = \frac{(1 - 0.62) \cdot 0.8}{1 - 0.8 \cdot 0.62} = \frac{0.304}{0.504} \approx 0.6032.$$

Alternativt kan man i nämnaren använda satsen om total sannolikhet ännu en gång, genom

$$\begin{aligned}\mathsf{P}(\text{tulpanen står kvar}) &= \mathsf{P}(\text{tulpanen står kvar} \mid T)\,\mathsf{P}(T) + \mathsf{P}(\text{tulpanen står kvar} \mid T^*)\,\mathsf{P}(T^*)\\ &= (1 - 0.62) \cdot 0.8 + 1 \cdot (1 - 0.8) = 0.504.\end{aligned}$$

2. (a) Den sammanlagda betjäningstiden för en kund vid lager A är $X + Y + Z$, med väntevärde $\mathsf{E}(X + Y + Z) = \mathsf{E}(X) + \mathsf{E}(Y) + \mathsf{E}(Z) = 2 + 3 + 6 = 11$, varians $\mathsf{V}(X + Y + Z) = \mathsf{V}(X) + \mathsf{V}(Y) + \mathsf{V}(Z) = 2^2 + 3^2 + 6^2 = 49$ och standarddavvikelse $\mathsf{D}(X + Y + Z) = \sqrt{\mathsf{V}(X + Y + Z)} = 7$. (4p)

(b) Betjäningstiderna för olika kunder vid vart och ett av lagren är likafördelade och oberoende, så att de sammanlagda betjäningstiderna, $T_A = \sum_{k=1}^{100}(X_k + Y_k + Z_k)$ respektive $T_B = \sum_{k=1}^{100} W_k$, är approximativt normalfördelade enligt centrala gränsvärdessatsen. Vi får att (6p)

$$T_A \underset{\sim}{\in} N\left(100\mathsf{E}(X + Y + Z),\ \sqrt{100}\mathsf{D}(X + Y + Z)\right)$$
$$T_B \underset{\sim}{\in} N\left(100\mathsf{E}(W),\ \sqrt{100}\mathsf{D}(W)\right)$$

och $\mathsf{E}(T_A) = 1100$, $\mathsf{D}(T_A) = 70$, $\mathsf{E}(T_B) = 1000$, $\mathsf{D}(T_A) = 60$. Approximativt blir den sökta sannolikheten

$$\begin{aligned}\mathsf{P}(T_A < T_B) &= \mathsf{P}\left(T_A - T_B < 0\right) = \mathsf{P}\left(\frac{T_A - T_B - (1100 - 1000)}{\sqrt{70^2 + 60^2}} < \frac{0 - (1100 - 1000)}{\sqrt{70^2 + 60^2}}\right)\\ &\approx \Phi(-1.0847) = 1 - \Phi(1.0847) \approx 1 - 0.86 = 0.14.\end{aligned}$$

3. Uppgiften har två aternativa lösningar: direkt resonemang eller direkt räkning: (10p)

Alt. 1: En exponentialfördelad variabel multiplicerad med en positiv konstant är också exponentialfördelad. Därför är $2X$ exponentialfördelad, med väntevärde $2a$, vilket är samma fördelning som för Y. Eftersom variablerna är oberoende måste $\mathsf{P}(2X < Y) = \mathsf{P}(Y < 2X)$ av symmetriskäl, så att $\mathsf{P}(2X \leq Y) = 1 - \mathsf{P}(Y < 2X) = 1 - \mathsf{P}(2X < Y) = 1 - \mathsf{P}(2X \leq Y)$, där den sista likheten följer av att variablerna är kontinuerliga. Alltså måste $\mathsf{P}(2X \leq Y) = 1/2$.

Alt. 2: Tätheterna för X och Y ges av $f_X(x) = \frac{1}{a}e^{-x/a}$, $x \geq 0$, respektive $f_Y(y) = \frac{1}{2a}\mathrm{e}^{-x/(2a)}$, $y \geq 0$. Den sökta sannolikheten fås genom integration av den simultana tätheten, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, över området $0 \leq 2x \leq y$:

$$\begin{aligned}\mathsf{P}(2X \leq Y) &= \int_0^\infty \int_{2x}^\infty f_X(x)f_Y(y)dy\,dx = \int_0^\infty \frac{1}{a}\mathrm{e}^{-x/a}\left(\int_{2x}^\infty \frac{1}{2a}\mathrm{e}^{-y/(2a)}dy\right)dx \\ &= \int_0^\infty \frac{1}{a}\mathrm{e}^{-x/a}\left[-\mathrm{e}^{-y/(2a)}\right]_{y=2x}^\infty dx = \int_0^\infty \frac{1}{a}\mathrm{e}^{-x/a}\mathrm{e}^{-2x/(2a)}dx \\ &= \int_0^\infty \frac{1}{a}\mathrm{e}^{-2x/a}dx = \left[-\frac{1}{2}\mathrm{e}^{-2x/a}\right]_{x=0}^\infty = \frac{1}{2}.\end{aligned}$$

4. (a) Detta är typexempel på stickprov i par, och lämplig modell ges av att z_i, $i = 1, \ldots, 10$, är oberoende observationer av en stokastisk variabel $Z \in N(\Delta,\ \sigma_z)$. En skattning av Δ ges av $\Delta^* = \bar{z} = -0.46$. Skattningens fördelning är $N\left(\Delta,\ \sigma_z/\sqrt{10}\right)$. (Om vi haft fler observationer kunde nöjt oss med att bara likafördelade observationer, så att skattningen är ungefär normalfördelad enligt CGS.) En skattning av σ_z är $s_z = \sqrt{0.1404} = 0.3747$, och medelfelet för Δ^* blir $d(\Delta^*) = s_z/\sqrt{10} = 0.1185$. Frihetsgraderna är $10 - 1 = 9$, och vi får ett 95%-igt konfidensintervall genom (10p)

$$I_\Delta = \Delta^* \pm t_{0.025}(9)d(\Delta^*) = -0.46 \pm 2.26 \cdot 0.1185 = (-0.728,\ -0.192).$$

(b) Nu kan vi inte längre använda stickprov i par, utan måste använda de två stickproven direkt. En enkel modell blir nu att x_i och y_i, $i = 1, \ldots, 10$, är oberoende observationer av två stokastiska variabler $X \in N\left(\mu,\ \sigma_x\right)$ respektive $X \in N\left(\mu + \Delta,\ \sigma_y\right)$. Vi antar dessutom att $\sigma_x = \sigma_y = \sigma$. Stickprovsmedelvärdena $\bar{x}$ och $\bar{y}$ har fördelning $N\left(\mu,\ \sigma/\sqrt{10}\right)$ respektive $N\left(\mu + \Delta,\ \sigma/\sqrt{10}\right)$, och skattningen $\Delta^* = \bar{y} - \bar{x}$ har fördelning $N\left(\Delta,\ \sigma\sqrt{\frac{1}{10} + \frac{1}{10}}\right)$, där σ skattas med $s = \sqrt{\frac{(10-1)s_x^2 + (10-1)s_y^2}{10-1+10-1}} = 5.19$. Medelfelet blir $d(\Delta^*) = s\sqrt{2/10} = 2.32$, och frihetsgraderna är $10 - 1 + 10 - 1 = 18$. Konfidensintervallet ges av (10p)

$$I_\Delta = \Delta^* \pm t_{0.025}(18)d(\Delta^*) = -0.46 \pm 2.10 \cdot 2.32 = (-5.34,\ 4.32).$$

Intervallet är nu oanvändbart brett!

5. Uppgiften är egentligen felformulerad; det är signifikansnnivån 1% som önskas. Ett test på nivån 99% skulle nästan alltid förkasta nollhypotesen! Naturligtvis är det också så att när accelerationen är noll är planet horisontellt, inte tvärtom.

(a) Skattningarna av β-parametrarna fås genom (4p)

$$\boldsymbol{\beta}^* = \begin{bmatrix}\beta_0^* \\ \beta_1^* \\ \beta_2^*\end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y} = \begin{bmatrix}-0.1390 \\ 2.328 \\ -0.1647\end{bmatrix}$$

och σ skattas av $s = \sqrt{Q_0/(21-3)} = 0.164$.

(b) Vi använder konfidensmetoden, och förkastar $H_0 : \beta_2 = 0$ till förmån för $H_1 : \beta_2 \neq 0$ om intervallet inte täcker 0. Medelfelet för β_2^* fås genom $d(\beta_2^*) = s \cdot \sqrt{c_2}$, där c_2 är diagonalelementet i $(\mathbf{X}^T\mathbf{X})^{-1}$ motsvarande β_2, dvs $c_2 = 0.3529$. Intervallet blir (8p)

$$I_{\beta_2} = \beta_2^* \pm t_{0.005}(21-3)d(\beta_2^*) = -0.1647 \pm 2.88 \cdot 0.0974 = (-0.445,\ 0.116),$$

och vi kan alltså inte förkasta noll-hypotesen, vilket tyder på att planet inte lutar.

(c) Vi ska konstruera ett konfidensintervall för $\mu_0 = \mu(1.5)$, och sätter $\mathbf{x}_0 = [1,\ 1.5,\ 1.5^2]$. Punktskattningen ges då av $\mu_0^* = \mathbf{x}_0\boldsymbol{\beta}^* = 2.9828$, med medelfel $d(\mu_0^*) = s\sqrt{\mathbf{x}_0(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_0^T} = 0.0470$. Ett 99%-igt konfidensintervall (Anmärkning: uppgiften specificerade inte konfidensgraden, så vi är fria att välja själva!) ges av (8p)

$$I_{\mu_0} = \mu_0^* \pm t_{0.005}(21-3)d(\mu_0^*) = 2.9828 \pm 2.88 \cdot 0.0470 = (2.847,\ 3.118).$$

6. (a) Vi vet att $f_X(x) = \frac{1}{\mu}\,\mathrm{e}^{-x/\mu}$ för $x \geq 0$ och $f_Y(y) = \frac{1}{2\mu}\,\mathrm{e}^{-y/2\mu}$ för $y \geq 0$. Det ger (6p)
$L(\mu) = f_X(x)\cdot f_Y(y) = \frac{1}{\mu}\,\mathrm{e}^{-x/\mu}\cdot\frac{1}{2\mu}\,\mathrm{e}^{-y/2\mu} = \frac{1}{2}\cdot\mu^{-2}\cdot\mathrm{e}^{-(x+y/2)/\mu}$.
$l(\mu) = \ln L(\mu) = \ln\frac{1}{2} - 2\ln\mu - \frac{x+y/2}{\mu}$.
$\frac{dl(\mu)}{d\mu} = -\frac{2}{\mu} + \frac{x+y/2}{\mu^2} = 0 \Rightarrow \mu^*_{\mathrm{ML}} = \frac{x+\frac{y}{2}}{2}$

(b) Vi vet, eftersom det är exponentialfördelning, att $\mathsf{E}(X) = \mu$, $\mathsf{V}(X) = \mu^2$, $\mathsf{E}(Y) = 2\mu$ och $\mathsf{V}(Y) = (2\mu)^2$. Det ger (6p)
$\mathsf{E}(Z) = \mathsf{E}(\frac{1}{2}\left(X + \frac{Y}{2}\right)) = \frac{1}{2}\left(\mathsf{E}(X) + \frac{\mathsf{E}(Y)}{2}\right) = \frac{1}{2}\left(\mu + \frac{2\mu}{2}\right) = \mu$,
$\mathsf{V}(Z) = \frac{1}{2^2}\left(\mathsf{V}(X) + \frac{\mathsf{V}(Y)}{2^2}\right) = \frac{1}{2^2}\left(\mu^2 + \frac{(2\mu)^2}{2^2}\right) = \frac{1}{2}\,\mu^2$.

(c) $\mathsf{V}(\mu^*) = \frac{1}{2^2}\left(\mathsf{V}(X) + \frac{1}{4^2}\mathsf{V}(Y+Y_2)\right) = \frac{1}{4}\left(\mu^2 + \frac{1}{16}(\mathsf{V}(Y) + \mathsf{V}(Y_2) + 2\mathsf{C}(Y,Y_2)\right) =$ (8p)
$= \frac{1}{4}\left(\mu^2 + \frac{1}{16}((2\mu)^2 + (2\mu)^2 + 2\mu^2)\right) = \frac{13}{32}\,\mu^2$.
Ja, det blir bättre eftersom variansen minskar från $\frac{1}{2}\,\mu^2 = 0.5\mu^2$ till $\frac{13}{32}\,\mu^2 \approx 0.41\mu^2$.