UPPSALA UNIVERSITET
Matematiska institutionen
Örjan Stenflo

Sannolikhet och statistik
2013-06-03

Sannolikhet och statistik, Tenta 2013-06-03
Lösningsförslag

1. (a) Då A och B är oberoende är $P(A \cap B) = P(A)P(B) = 0.1 \cdot 0.3 = 0.03$. Detta ger

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.10 + 0.30 - 0.03 = 0.37.$$

(b) Låt A_2 vara händelsen att första talet är delbart med 2, A_5 vara händelsen att första talet är delbart med 5, B_2 vara händelsen att andra talet är delbart med 2, B_5 vara händelsen att andra talet är delbart med 5. Vi har $P(A_2) = P(B_2) = 4/9$ och $P(A_5) = P(B_5) = 1/9$.

$$\begin{aligned} P(\text{produkten är delbar med 10}) &= P(A_2 \cap B_5) + P(A_5 \cap B_2) \\ &= P(A_2)P(B_5) + P(A_5)P(B_2) \\ &= 2 \cdot \frac{1}{9} \cdot \frac{4}{9} = \frac{8}{81} \approx 0.099. \end{aligned}$$

(c) Antag att vi storleksordnat datamaterialet så att $x_1 \leq x_2 \leq x_3$. Då är $\tilde{x} := x_2$ medianen och $\bar{x} := (x_1 + x_2 + x_3)/3$ medelvärdet. Om t.ex. $x_1 = -2$, $x_2 = 2$, $x_3 = 3$ blir $\tilde{x} = 2$ och $\bar{x} = 1$ alltså är medelvärdet mindre än medianen för detta datamaterial.

2. Inför händelserna $F =$"fusktärningen väljs", och $'6'=$"tärningen visar 6". Vi har $P(F) = 0.5$, $P(F^c) = 1 - 0.5 = 0.5$, $P('6'|F^c) = 1/6$, $P('6'|F) = 1$.

Enligt lagen om total sannolikhet är

$$P('6') = P('6'|F^c)P(F^c) + P('6'|F)P(F) = (1/6) \cdot (1/2) + 1 \cdot (1/2) = 7/12.$$

Enligt Bayes sats är $P(F|'6') = P('6'|F)P(F)/P('6') = (1/2)/(7/12) = 6/7 \approx 0.857$.

3. (a) Per definition är

$$E(X) = \sum_k kp(k) = (-1) \cdot \frac{5}{36} + 0 \cdot \frac{26}{36} + 1 \cdot \frac{5}{36} = 0,$$

och

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = E(X^2) = \sum_k k^2 p(k) \\ &= (-1)^2 \cdot p(-1) + 0^2 \cdot p(0) + 1^2 \cdot p(1) = p(-1) + p(1) = 10/36 = 5/18. \end{aligned}$$

(b) Vi har

$$E(S) = E(\sum_{i=1}^{500} X_i) = \sum_{i=1}^{500} \underbrace{E(X_i)}_{=0} = 0.$$

$$V(S) = V(\sum_{i=1}^{500} X_i) = \sum_{i=1}^{500} \underbrace{V(X_i)}_{5/18} = \frac{2500}{18}.$$

Från centrala gränsvärdessatsen (CGS) följer att $\frac{S-E(S)}{\sqrt{V(S)}}$ är ungefär $N(0,1)$-fördelad. Då S är en heltalsvärd slumpvariabel ser vi från CGS att

$$\begin{aligned} P(S \leq 10) &= P(S \leq 10.5) = P(\frac{S - E(S)}{\sqrt{V(S)}} \leq \frac{10.5 - 0}{\sqrt{2500/18}}) \\ &\approx \Phi(\underbrace{10.5 \cdot \sqrt{18}/50}_{0.891}) \underbrace{=}_{tabell} 0.8135, \end{aligned} \qquad (1)$$

(där $\Phi(x)$ betecknar fördelningsfunktionen för en $N(0,1)$-fördelad slumpvariabel, vars värden ges av tabell.)

4. Vi har $X \sim Bin(500, 1/6)$ och $Y \sim Bin(500, 1/6)$ där X och Y är oberoende. Eftersom $V(X) = V(Y) = 500 \cdot 1/6 \cdot 5/6 > 5$ är X och Y ungefär normalfördelade, vilket gör att $X - Y$ också är ungefär normalfördelad. Enligt formelsamlingen är

$$E(X - Y) = E(X) - E(Y) = 500 \cdot 1/6 - 500 \cdot 1/6 = 0,$$

och

$$V(X - Y) = V(X) + V(-Y) = V(X) + V(Y) = 1000 \cdot 1/6 \cdot 5/6 = 2500/18.$$

Alltså är $X - Y$ approximativt $N(0, 2500/18)$-fördelad (d.v.s. $X - Y$ har samma approximativa normalfördelning som S i uppgift 3.)[1]

Detta ger

$$P(X > Y + 10) = P(X - Y > 10) = 1 - P(X - Y \leq 10) \underbrace{\approx}_{se\ (1)\ i\ uppgift\ 3} 1 - 0.8135 = 0.1865.$$

5. (a) Låt $X \sim N(7.5, 0.8^2)$ vara vikten (i kg) av en slumpmässigt utvald 6 månader gammal flicka.

$$P(X > 7.0) = 1 - P(X \leq 7.0) = 1 - \Phi\Big(\underbrace{\frac{7.0 - 7.5}{0.8}}_{-0.625}\Big) = \Phi(0.625) \underbrace{\approx}_{tabell} 0.734,$$

(där $\Phi(x)$ betecknar fördelningsfunktionen för en $N(0,1)$-fördelad slumpvariabel, vars värden ges av tabell.)

[1] Kuriosa: $X - Y$ har *exakt* samma fördelning som S i uppgift 3. (Övning: Visa detta.)

(b) Om Y = antal utvalda 6 månader gamla flickor som väger mindre än 7.0 kg, gäller $Y \sim Bin(5, p)$, där $p = P(X < 7.0) = P(X \leq 7.0) \approx 1 - 0.734 = 0.266$.

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y < 2) = 1 - P(Y \leq 1) = 1 - P(Y = 0) - P(Y = 1) \\ &= 1 - \binom{5}{0} p^0 (1-p)^5 - \binom{5}{1} p^1 (1-p)^4 = 1 - (1-p)^5 - 5p(1-p)^4 \\ &\approx 1 - 0.213 - 0.386 = 0.401. \end{aligned}$$

6. Låt p vara andelen röstberättigade som vid tidpunkten för undersökningen skulle ha röstat på folkpartiet. Baserat på undersökningen skattar vi p med $\hat{p} = 0.068$. Undersökningen är baserad på $n = 1941$ intervjuer. Skattningen $\hat{p}$ kan antas vara en observation av $\hat{P} := X/n$, där $X \sim Bin(n, p)$. Då tumregeln $n\hat{p}(1 - \hat{p}) \approx 123 > 5$ är uppfylld är X approximativt $N(np, n\hat{p}(1 - \hat{p}))$-fördelad, vilket medför att $\hat{P}$ är approximativt $N(p, \frac{\hat{p}(1-\hat{p})}{n})$-fördelad. Från detta följer att

$$\begin{aligned} 1 - \alpha &\approx P(-\lambda_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq \lambda_{\alpha/2}) \\ &= P\Big(\hat{P} - \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{P} + \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\Big), \end{aligned} \tag{3}$$

där för ett givet $0 < \alpha < 1$, $\lambda_{\alpha/2}$ betecknar $\alpha/2$-kvantilen för standardnormalfördelningen, d.v.s. det tal som uppfyller $\Phi(\lambda_{\alpha/2}) = 1 - \alpha/2$.

Ett konfidensintervall för p med approximativ konfidensgrad $1 - \alpha$ ges därför av

$$I_p = \Big(\hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\Big).$$

Om vi vill ha ett 99 % konfidensintervall använder vi $\alpha = 1 - 0.99 = 0.01$. Enligt tabell är $\lambda_{\alpha/2} = \lambda_{0.005} = 2.5758$, som ger konfidensintervallet

$$I_p = \Big(0.068 \pm 2.5758\sqrt{\frac{0.068(1 - 0.068)}{1941}}\Big) \approx (0.053, 0.083).$$

7. Om vi antar att $x_1 = 6.8$, $x_2 = 7.7$,...,$x_{10} = 21.5$ är observationer av X_1,...,X_{10} där $X_i \sim N(\mu_i, \sigma_1^2)$ och $y_1 = 9.5$, $y_2 = 9.0$,...,$y_{10} = 22.0$ är observationer av Y_1,...,Y_{10}, där $Y_i \sim N(\mu_i + \Delta, \sigma_2^2)$, där Δ och μ_i och σ_i är konstanter, och alla X_i och Y_i variabler är oberoende så blir $z_i = y_i - x_i$ ett $N(\Delta, \sigma_1^2 + \sigma_2^2)$-fördelat stickprov. Detta stickprov, alltså $z_1 = 9.5 - 6.8 = 2.7$,...,$z_{10} = 22.0 - 21.5 = 0.5$, har stickprovsmedelvärde $\bar{z} = 1.54$ och stickprovsstandardavvikelse $s_z = \sqrt{(\sum_{i=1}^{10} z_i^2 - 10\bar{z}^2)/(10-1)} \approx 2.4098$.

Enligt formelsamlingen vet vi att referensvariabeln $\frac{\bar{Z}-\Delta}{S_z/\sqrt{10}} \sim t(10-1)$, där $\bar{Z}$ och S_z betecknar motsvarande stickprovsvariabler korresponderande till $\bar{z}$ och s_z. Ett 95% konfidensintervall för Δ ges därför av

$$I_\Delta = \bar{z} \pm t_{0.025}(9)s_z/\sqrt{10} = 1.54 \pm 2.2622 \cdot 2.4098/\sqrt{10} \approx (-0.18, 3.26),$$

där $t_{0.025}(9) = 2.2622$ kommer från tabell. Då $0 \in I_\Delta$ kan vi inte påvisa någon statistisk säkerställd systematisk effekt på halten av ämnet.

8. (a) Vi har följande par av observationer:

$$(x_1, y_1) = (840, 1000), (x_2, y_2) = (750, 910), (x_3, y_3) = (600, 800),$$

$$(x_4, y_4) = (780, 950), (x_5, y_5) = (980, 1300), (x_6, y_6) = (550, 680),$$

$$(x_7, y_7) = (560, 710), (x_8, y_8) = (1200, 1400), (x_9, y_9) = (735, 990), (x_{10}, y_{10}) = (675, 910).$$

Vi vill skatta parametrarna α och β för den teoretiska regressionslinjen $y = \alpha + \beta x$. Enligt formelsamlingen, avsnitt 2.3, sid 4, skattas β med $\beta^\star = S_{xy}/S_{xx}$, α med $\alpha^\star = \bar{y} - \beta^\star \bar{x}$, och σ med $\sigma^\star = \sqrt{(S_{yy} - \frac{S_{xy}^2}{S_{xx}})/(10-2)}$, där

$$\begin{aligned}
\bar{y} &= \frac{\sum_{i=1}^{10} y_i}{10} = 965 \\
\bar{x} &= \frac{\sum_{i=1}^{10} x_i}{10} = 767, \\
S_{xy} &= \sum_{i=1}^{10} (x_i \cdot y_i) - 10 \cdot \bar{x} \cdot \bar{y} = 7811000 - 10 \cdot 767 \cdot 965 = 409450, \\
S_{xx} &= \sum_{i=1}^{10} x_i^2 - 10 \cdot \bar{x}^2 = 6248850 - 10 \cdot (767)^2 = 365960 \\
S_{yy} &= \sum_{i=1}^{10} y_i^2 - 10 \cdot \bar{y}^2 = 9795300 - 10 \cdot (965)^2 = 483050.
\end{aligned}$$

Detta ger

$$\beta^\star = S_{xy}/S_{xx} = 409450/365960 \approx 1.1188$$

och

$$\alpha^\star = \bar{y} - \beta^\star \bar{x} = 965 - (409450/365960) \cdot 767 \approx 106.8512,$$

och

$$\sigma^\star = \sqrt{(S_{yy} - \frac{S_{xy}^2}{S_{xx}})/(10-2)} = \sqrt{(483050 - \frac{409450^2}{365960})/8} = 55.8365.$$

(b) Enligt formelsamlingen är $\frac{\alpha^\star - \alpha}{\sigma^\star\sqrt{\frac{1}{10}+\frac{\bar{x}^2}{S_{xx}}}}$, $t(\underbrace{10-2}_{8})$-fördelad.

Ett 95 % konfidensintervall för α ges därför av

$$I_\alpha = \left(\alpha^\star - t_{0.025}(8)\sigma^\star\sqrt{\frac{1}{10}+\frac{\bar{x}^2}{S_{xx}}}, \alpha^\star + t_{0.025}(8)\sigma^\star\sqrt{\frac{1}{10}+\frac{\bar{x}^2}{S_{xx}}}\right)$$

där $t_{0.025}(8) \approx 2.306$ är 0.025-kvantilen hos $t(8)$-fördelningen som ges av tabell, alltså

$$\begin{aligned} I_\alpha &= \left(106.8512 - 2.306 \cdot 55.8365\sqrt{\frac{1}{10}+\frac{767^2}{365960}}, 106.8512 + 2.306 \cdot 55.8365\sqrt{\frac{1}{10}+\frac{767^2}{365960}}\right) \\ &\approx (-61.4, 275.1). \end{aligned}$$

Då $0 \in I_\alpha$ kan vi inte, på 5 % nivån, förkasta hypotesen att $\alpha = 0$. Det kan därför inte anses vara statistiskt säkerställt att $\alpha \neq 0$.