UPPSALA UNIVERSITET
Matematiska institutionen
Raimundas Gaigalas

STATISTIK IP
2004-01-15

Lösningar till tenta 040115

1. Inför händelser A, B och C att respektive väg är plogad. Då är A, B och C oberoende och $P(A) = 0.3$, $P(B) = 0.4$, $P(C) = 0.7$.

   (a) Sannolikheten att någeon av vägarna är plogad:

   $$\begin{aligned} P(A \cup B \cup C) &= 1 - P(A^* \cap B^* \cap C^*) = 1 - P(A^*)P(B^*)P(C^*) \\ &= 1 - 0.7 \cdot 0.6 \cdot 0.3 = 0.874. \end{aligned}$$

   (b) Söks: $P(B|A \cup B \cup C) = \dfrac{P(B \cap (A \cup B \cup C))}{P(A \cup B \cup C)} = \dfrac{P(B)}{P(A \cup B \cup C)} = \dfrac{0.4}{0.874} = 0.4577.$

2. (a) X = antal felaktiga batterier i partiet $\sim \mathrm{Bin}(14, 0.15)$. Sannolikheten att minst 2 är felaktiga:

   $$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.3567 = 0.6433.$$

   (b) X = antal felaktiga batterier i partiet $\sim \mathrm{Bin}(160, 0.15)$. Eftersom $npq = 160 \cdot 0.15 \cdot 0.85 = 20.4 > 10$, kan normalapproximation användas. Därmed $X \overset{\cdot}{\sim} \mathrm{N}(24, 20.4)$ och den sökta sannolikheten

   $$\begin{aligned} P(X \geq 20) &= 1 - P(X \leq 19) \approx 1 - \Phi(\frac{19 - 24 + 0.5}{\sqrt{20.4}}) \\ &= 1 - \Phi(-1.00) = \Phi(1.00) = 0.8413. \end{aligned}$$

3. X = matchtid $\sim \mathrm{N}(1.5, 0.49)$, Y = paustid $\sim \mathrm{N}(0.5, 0.01)$.

   (a) Sannolikheten att en match är slut efter 1.85 timmar:

   $$P(X \leq 1.85) = \Phi(\frac{1.85 - 1.5}{\sqrt{0.49}}) = \Phi(0.5) = 0.6915.$$

   (b) Z = den sammanlagda tiden för 3 matcher och 2 pauser. Då $Z = X_1 + X_2 + X_3 + Y_1 + Y_2 \sim \mathrm{N}(3 \cdot 1.5 + 2 \cdot 0.5, 3 \cdot 0.49 + 2 \cdot 0.01) = \mathrm{N}(5.5, 1.49)$. Vi söker

   $$P(Z > 6.5) = 1 - \Phi(\frac{6.5 - 5.5}{\sqrt{1.49}}) = 1 - \Phi(0.82) = 1 - 0.7939 = 0.2061.$$

   (c) Matcherna slutar oberoende av varandra. Därmed

   $$P(X_1 \leq 1.85, X_2 \leq 1.85, X_3 \leq 1.85) = (P(X \leq 1.85))^3 = 0.6915^3 = 0.3307.$$

4. (a) $P(X > 0) = p_X(1) + p_X(2) = 0.1 + 0.3 = 0.4.$

   (b) $$\begin{aligned} E(X) &= \sum_{k=-2}^{2} k p_X(k) = (-2) \cdot 0.1 + (-1) \cdot 0.3 + 0 \cdot 0.2 + 1 \cdot 0.1 \\ &\quad + 2 \cdot 0.3 = 0.2, \\ E(X^2) &= \sum_{k=-2}^{2} k^2 p_X(k) = (-2)^2 \cdot 0.1 + (-1)^2 \cdot 0.3 + 0^2 \cdot 0.2 \\ &\quad + 1^2 \cdot 0.1 + 2^2 \cdot 0.3 = 2, \end{aligned}$$

   $\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 2 - 0.04 = 1.96,$

   $D(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{1.96} = 1.4.$

(c) $P(X=2|X>0) = \dfrac{P(X=2\cap X>0)}{P(X>0)} = \dfrac{P(X=2)}{P(X>0)} = \dfrac{0.3}{0.4} = 0.75.$

5. $X =$ svarstid med musik $\sim \mathrm{N}(\mu_1, \sigma^2)$, $Y =$ svarstid utan musik $\sim \mathrm{N}(\mu_2, \sigma^2)$. $x_1, \ldots, x_6$ observationer av X, $y_1, \ldots, y_5$ observationer av Y. Vi testar $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 < \mu_2$.

   Punktskattning $\widehat{\mu_1 - \mu_2} = \bar{x} - \bar{y}$. Spridningsmått $\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$. σ^2 är okänd, skattas med den sammanvägda skattningen $s^2_{(2)} = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$. Testvariabeln $\dfrac{\bar{X}-\bar{Y}}{s_{(2)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ är $t(n_1+n_2-2)$-fördelad om H_0 är sann.

   Test: om $\frac{\bar{x}-\bar{y}}{s_{(2)}\sqrt{\frac{1}{6}+\frac{1}{5}}} < -t_{0.05}(9) = -1.833$, förkasta H_0. I vårt fall $\bar{x} = 14.1333$, $\bar{y} = 14.56$, $s_1^2 = 0.078667$, $s_2^2 = 0.083$, $s_{(2)} = 0.28389$, $\frac{\bar{x}-\bar{y}}{s_{(2)}\sqrt{\frac{1}{6}+\frac{1}{5}}} = -2.4820 < -1.833$, H_0 förkastas. Med 95 % säkerhet finns det en Carolaeffekt.

6. $X =$ antal kunder som anländer under 3 timmar $\sim \mathrm{Po}(3\lambda)$, $x = 57$ är en observation av X. Punktskattning $\widehat{3\lambda} = x = 57 > 15$, därmed kan normalapproximation tillämpas. Spridningsmått $D(X) = \sqrt{\lambda}$ okänd, skattas med $\sqrt{x}$. Testvariabeln $\dfrac{X-3\lambda}{\sqrt{X}}$ är approximativt $N(0,1)$-fördelad. Konfidensintervall för 3λ:

$$I_{3\lambda} = (x \pm z_{0.025}\sqrt{x}) = (57 \pm 1.96\sqrt{57}) = (42.2023, 71.7977).$$

   Konfidensintervall för λ: $I_\lambda = (42.2023/3, 71.7977/3) = (14.07, 23.93)$.

7. $X =$ år, $Y =$ befolkning. Linjär regressionsmodell: $x_1, \ldots, x_6$ är givna, $y_1, \ldots, y_6$ är observationer av $Y_i = \alpha + \beta x_i + \epsilon_i$, där ϵ_i är $\mathrm{N}(0, \sigma^2)$-fördelade.

   (a) Den skattade regressionslinjen: $y = \hat{\alpha} + \hat{\beta}x$, $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ och $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. $S_{xy} = 0.902$, $S_{xx} = 17.5$, $\bar{x} = 1992.5$, $\bar{y} = 8.7207$, $\hat{\beta} = 0.05154$, $\hat{\alpha} = -93.9785$. Därmed $y = -93.9785 + 0.05154x$.

   (b) Korrelationskoefficienten $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$. $S_{yy} = 0.04699$ medför $r = 0.9947$. Ett starkt linjärt samband.

   (c) Insättning av $y = 9$ i den skattade linjens ekvation ger en skattning $x = \frac{9+93.9785}{0.05154} = 1997.9$ för tidpunkt när befolkningen blir 9 miljoner.

8. $X =$ antalet rätta svar bland 16 försök $\sim \mathrm{Bin}(16, p)$, p okänd. Vi testar $H_0:\ p = 0.5$ mot $H_1:\ p > 0.5$. Direktmetoden:

$$P = P(X \geq 13,\ \text{om } p = 0.5) = 1 - P(X \leq 12) = 1 - 0.9894 = 0.0106.$$

   Eftersom $P < 0.05$, förkastas H_0. Med 95 % säkerhet känner personen skillnaden.