Matematisk statistik 2022-08-16 – Lösningar

1. (a) $P(A \cap B \cap C) = P((A \cap C) \cap (B \cap C)) = P(A \cap C)P(B \cap C) = 0.2 \cdot 0.3 = 0.06.$

   (b) Vi söker sannolikheten att endast A inträffar. Om man ritar ett Venn-diagram kan man övertyga sig själv om att denna sannolikhet ges av

   $$P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) = 0.6 - 0.2 - 0.2 + 0.06 = 0.26.$$

   När vi drar bort både $P(A \cap B)$ och $P(A \cap C)$ så drar vi bort ”trippelsnittet” två gånger. Detta måste man kompensera för.

   (c) Låt $p = P(B) = 0.6$. Vi upprepar försöket åtta gånger och räknar antalet X gånger som B inträffar. Det följer att $X \sim \text{Bin}(8, p)$ och vi söker $P(X > 6)$:

   $$P(X > 6) = P(X = 7) + P(X = 8) = \binom{8}{7} 0.6^7 \cdot 0.4 + \binom{8}{8} 0.6^8 = 0.1064.$$

   Svar: (a) 0.06. (b) 0.26. (c) 0.1064.

2. (a) Vi söker konfidensintervall för andelen som svarat att det är OK. Totalt sett $n = 360$ stycken personer, och fått $X = 18$ stycken OK. En naturlig skattning på den verkliga andelen ges av

   $$\widehat{P} = \frac{X}{n},$$

   och vårt observerade värde är $\widehat{p} = 18/360 = 0.05$. Fördelningen för X ges av $X \sim \text{Bin}(n, p)$. Vidare ser vi att

   $$np(1-p) \approx 360 \cdot 0.05 \cdot 0.95 = 17.1,$$

   så vi borde kunna göra en normalapproximation: $X \overset{\text{appr.}}{\sim} N(np, D)$ där $D = \sqrt{np(1-p)}$. Alltså blir $\widehat{P} \overset{\text{appr.}}{\sim} N(p, d)$, där

   $$d = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{17.1/360^2} = 0.0115.$$

   Vår testvariabel blir alltså

   $$Z = \frac{\widehat{P} - p}{d} \overset{\text{appr.}}{\sim} N(0, 1).$$

   Vi stänger in Z:

   $$P(-\lambda_{\alpha/2} \leq Z \leq \lambda_{\alpha/2}) = 1 - \alpha,$$

   där λ_α är normalfördelningens α-kvantil. Vi har $\alpha = 0.05$ och $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Vi löser ut p ur olikheten inuti sannolikhetsmåttet, och erhåller då

   $$\widehat{P} - \lambda_{0.025} d \leq p \leq \widehat{P} + \lambda_{0.025} d.$$

   Detta ger oss konfidensintervallet $I_p = [0.0275, 0.0725]$ om vi ersätter $\widehat{P}$ med det observerade värdet $\widehat{p} = 0.05$.

   (b) Egentligen är vilken siffra vi än hittar på (som är en sannolikhet) en punktskattning, men om dessa är lämpliga är mer tveksamt. Konfidensintervall brukar vara symmetriska (om de är dubbelsidiga, vilket vi antar här), så en vettig punkskattning på andelen p är talet i mitten av intervallet. Vi skattar alltså med $\widehat{p} = (0.039 + 0.062)/2 = 0.0505$.

   Svar: (a) $I_p = [0.0275, 0.0725]$ (b) T.ex. $\widehat{p} = 0.0505$.

3. (a) Vi varje dragning så förändras sannolikheten att dra en vit (eller svart) kula nästa gång. Dragningar efter varandra är alltså inte oberoende.
   (b) Vi konstaterar att $0 \leq p_X(k) \leq 1$ för alla k och att $\sum_k p_X(k) = 1$. Detta är nödvändigt och tillräckligt för att p_X ska vara en sannolikhetsfunktion. Väntevärdena kan beräknas enligt

$$E(X) = \sum_{k=-3}^{3} k p_X(k) = -2 \cdot 0.2 - 1 \cdot 0.2 + 1 \cdot 0.3 + 2 \cdot 0.2 = 0.1$$

   och

$$E(X^2) = \sum_{k=-3}^{3} k^2 p_X(k) = (-2)^2 \cdot 0.2 + (-1)^2 \cdot 0.2 + 1^2 \cdot 0.3 + 2^2 \cdot 0.2 = 2.1.$$

   (c) Endast (i) är en sannolikhetsfunktion. I (ii) så summerar inte sannolikheterna till ett och i (iii) så är $p_X(-1) < 0$.

4. (a) Eftersom X och Y är oberoende,

$$P(X < 0, Y > 0) = P(X < 0)P(Y > 0) = \Phi(0)(1-P(Y \leq 0)) = 0.5(1-\Phi(1/4)) = 0.20.$$

   (b) Låt $Z = X - 2Y$. Då är $Y \sim \mathrm{N}(-1, \sqrt{25 + 4 \cdot 16})$ (där $V(Z) = 89$) och

$$P(X + 2Y > 0) = P(Z > 0) = 1 - P(Z \leq 0) = 1 - \Phi(1/\sqrt{89}) = 0.458.$$

   (c) Eftersom variablerna är oberoende kan vi summera varianserna och därmed erhålla att

$$V(\overline{X}) = \frac{1}{n^2} V\left(\sum_{k=1}^{n} X_k\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

   Då $\sigma \leq 2$ ser vi att $V(\overline{X}) \leq 4/n$, så kravet att $V(\overline{X}) < 0.1$ är uppfyllt då $n \geq 41$ (41 eftersom det är strikt olikhet!).

5. Vi låter H_0 vara hypotesen att antagandet om exponentialfördelning stämmer och H_1 att antagandet inte är sant. Om vi antar H_0 så gäller att täthetsfunktionen för livslängden hos en komponent X ges av $f(x) = \mu^{-1} \exp(-\mu^{-1} x)$, så

$$P(a \leq X < b) = \int_a^b \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right) dx = \exp\left(-\frac{a}{\mu}\right) - \exp\left(-\frac{b}{\mu}\right).$$

   Med siffrorna ovan ser vi att

$$P(X \in I_k) = \begin{cases} p_1 = 0.2212, & k = 1, \\ p_2 = 0.1723, & k = 2, \\ p_3 = 0.2387, & k = 3, \\ p_4 = 0.1447, & k = 4, \\ p_5 = 0.2231, & k = 5. \end{cases}$$

   Vi finner här sannolikheterna att en komponent går sönder i I_1 respektive I_4: 0.2212 respektive 0.1447.

Teststorheten vi använder för ett χ^2-test kommer nu ges av

$$q = \sum_{j=1}^{5} \frac{(x_j - np_j)^2}{np_j} = \frac{(11 - 50 \cdot 0.2212)^2}{50 \cdot 0.2212} + \cdots + \frac{(11 - 50 \cdot 0.2231)^2}{50 \cdot 0.2231} = 0.1276.$$

Om H_0 är sann så kommer q vara en observation av $Q \overset{\text{appr.}}{\sim} \chi^2(5-1) = \chi^2(4)$, så med det kritiska området $C = (0,\ c)$ där $c = 13.28$, ser vi att vi inte kan förkasta H_0. Säljaren kan mycket väl ha rätt.

6. (a) Eftersom $X \sim \text{Bin}(n,p)$ så är $E(\widehat{P}) = \dfrac{1}{n}E(X) = \dfrac{np}{n} = p$. Således väntevärdesriktig!

   (b) Allt (ja, varenda en du kan komma på!) är skattningar av p^2. Men de flesta är så klart väldigt dåliga. Är $\widehat{P}^2$ väntevärdesriktig? Vi undersöker:

$$\begin{aligned} E(\widehat{P}^2) &= \frac{E(X^2)}{n^2} = \frac{V(X) + E(X)^2}{n^2} = \frac{np(1-p) + (np)^2}{n^2} \\ &= p^2\left(1 - \frac{1}{n}\right) + \frac{p}{n}. \end{aligned}$$

   Alltså kommer inte väntevärdet för skattningsvariabeln att bli p^2. Detta är alltså inte en väntevärdesriktig skattning! Tycker du den verkar vettig ändå?

   (c) Från uttrycket i föregående deluppgift kan vi enkelt se att

$$\lim_{n\to\infty} E(\widehat{P}^2) = p^2\,(1-0) + 0 = p^2,$$

   så enligt definitionen i uppgiften är detta en s.k. asymptotiskt väntevärdesriktig skattning av p^2.

Svar: Se ovan.