UPPSALA UNIVERSITET
Matematiska Institutionen
Ingemar Kaj

PROV I MATEMATIK
Sannolikhet och statistik, 1MS321
2016-08-15
Lösningar

1. Vi söker $P(X > Y) = P(X - Y > 0)$. Eftersom X och Y är normalfördelade så är också variabeln $X-Y$ normalfördelad. Väntevärdet är $E(X-Y) = E(X)-E(Y) = -4-(-2) = -2$. Variansen blir $V(X-Y) = V(X) + (-1)^2 V(Y) = 5 + 4 = 9$, eftersom X och Y är oberoende. Det följer att den centrerade och omskalade variabeln $Z = (X - Y - (-2))/3$ har standardnormalfördelningen $N(0, 1)$, och vidare

$$P(X - Y > 0) = P\Big(\frac{X - Y - (-2)}{3} > \frac{0 - (-2)}{3}\Big) = P(Z > 2/3) = 1 - \Phi(2/3).$$

Från tabell fås $\Phi(2/3) \approx 0.747$, så $P(X - Y > 0) \approx 0.253$.

2. a) Skattningens väntevärde är

$$E(\widehat{m}_1) = E[(X + Y)/2] = E(X)/2 + E(Y)/2 = (m_1 - m_2)/2 + (m_1 + m_2)/2 = m_1,$$

vilket ger väntevärdesriktigheten, per definition.

b) På grund av oberoendet är skattningens varians och standardavvikelse

$$V(\widehat{m}_1) = V(\frac{X+Y}{2}) = \frac{V(X) + V(Y)}{4} = \frac{m_1 - m_2 + m_1 + m_2}{4} = \frac{m_1}{2}, \quad D(\widehat{m}_1) = \sqrt{\frac{m_1}{2}}$$

Standardavvikelsen $D(\widehat{m}_1)$ beror alltså på den okända parametern m_1. Baserat på observationer x och y kan vi enligt a) skatta m_1 med $\widehat{m}_1 = (x + y)/2$. Medelfelet $d(\widehat{m}_1)$ är motsvarande skattning av standardavvikelsen, dvs $d(\widehat{m}_1) = \sqrt{\widehat{m}_1/2} = \sqrt{x+y}/2$.

c) Sätt $\widehat{m}_2 = (Y - X)/2$. Som ovan visas att $E(\widehat{m}_2) = m_2$ och $D(\widehat{m}_2) = \sqrt{m_1/2}$, vilket leder till samma medelfel som i föregående uppgift: $d(\widehat{m}_2) = \sqrt{x+y}/2$.

3. Fördelningsfunktionen för X blir

$$F(x) = \int_0^x \frac{y}{2}\,dy = \frac{x^2}{4}, \quad 0 \le x \le 2.$$

Lösningen av ekvationen $F(x) = u$, då $0 \le u \le 1$, ges av $x = 2\sqrt{u}$. Enligt “inversmetoden” för simulering används nu de givna rektangelfördelade slumptalen $u_1, \dots, u_5$ för att få slumptal $x_i = 2\sqrt{u_i}$ från X, vilket ger

$$x = [1.8052 \quad 1.9035 \quad 0.7127 \quad 1.9114 \quad 1.5905]$$

Medelvärdet blir $\bar{x} = 1.5847$, vilket kan jämföras med

$$E(X) = \int_0^2 x \cdot \frac{x}{2}\,dx = \frac{4}{3}.$$

4. Låt X vara payload-storleken och $Y = 20 + X$ totala paketstorleken mätt i bytes.

a) Väntevärdet av Y är

$$E(Y) = 20 + E(X) = 20 + 0.2 \cdot 80 + 0.7 \cdot 200 + 0.1 \cdot 1500 = 20 + 306 = 326$$

b) Protokollets effektivitet tolkas som kvoten $E(X)/E(Y)$, vilket blir $306/326 = 0.9386\dots$ dvs 93.9%.

c) Vi har

$$E(X^2) = 0.2 \cdot 80^2 + 0.7 \cdot 200^2 + 0.1 \cdot 1500^2 = 254\,280$$

så att

$$V(Y) = V(X) = E(X^2) - E(X)^2 = 254\,280 - 306^2 = 160\,644, \quad D(Y) = 400.804\ldots$$

d) Sätt $S = Y_1 + \cdots + Y_{3000}$, där $Y_1, Y_2, \ldots$ är oberoende variabler med samma fördelning som Y. Vi söker $P(S < 1000\,000)$. Enligt a) och c) gäller

$$E(S) = 3000 \cdot 326 = 978\,000 \text{ bytes} = 978 \text{ kbytes}, \quad D(S) = \sqrt{3000 \cdot 160\,644} = 21.9529 \text{ kbytes}$$

Med S uttryckt i enheten kilobytes följer nu, enligt centrala gränsvärdessatsen

$$P(S < 1000) = P\Big(\frac{S - 978}{21.9529} < \frac{1000 - 978}{21.9529}\Big) \approx \Phi(1.0021) \approx 0.841$$

Om man istället tolkar Mbytes som 2^{20} bytes fås en högre sannolikhet, ganska nära 1.

5. a) Sannolikheterna för alla vägar ut från en nod måste summera till ett. Alltså fås $\star = 2/3$, $\star\star = 1/4$. Motsvarande övergångsmatris blir

$$\mathbf{P} = \begin{pmatrix} 3/4 & 0 & 1/4 \\ 2/3 & 1/3 & 0 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

b)

c) Den stationära fördelningen $\pi = (\pi_0\ \pi_1\ \pi_2)$, fås från balansekvationerna $\pi = \pi\mathbf{P}$, vilka utskrivna i detta fall blir

$$\begin{aligned} \frac{3}{4}\pi_0 + \frac{2}{3}\pi_1 + \frac{1}{4}\pi_2 &= \pi_0 \\ \frac{1}{3}\pi_1 + \frac{1}{2}\pi_2 &= \pi_1 \\ \frac{1}{4}\pi_0 + \frac{1}{4}\pi_2 &= \pi_2 \end{aligned}$$

dvs $\pi_1 = \pi_0/4$, $\pi_2 = \pi_0/3$. Normering enligt $\pi_0 + \pi_1 + \pi_2 = 1$ ger nu lösningen $\pi = (12/19\ \ 3/19\ \ 4/19)$.

6. Som punktskattningar av väntevärde m och varians σ^2 använder vi stickprovsmedelvärde och stickprovsvarians, dvs $\widetilde{m} = \bar{x} = 212.4$ och $\widetilde{\sigma^2} = s^2 = 18.45^2$. Variansen för $\widetilde{m}$ är $V(\widetilde{m}) = V(\bar{X}) = \sigma^2/16$, och medelfelet således $d(\widetilde{m}) = s/4$. Ett (exakt) 95% konfidensintervall för m ges då av

$$\begin{aligned} I_p: \quad & [\bar{x} - t_{0.025}(15)\,18.45/4\ ,\ \bar{x} + t_{0.025}(15)\,18.45/4] \\ & = [212.4 - 2.13 \cdot 4.6125\ ,\ 212.4 + 2.13 \cdot 4.6125] \approx [202.57\ ,\ 222.23] \end{aligned}$$

Med antagande om känt $\sigma^2 = 20$ fås ett (exakt) konfidensintervall med samma konfidensgrad direkt från normalfördelningens egenskaper, och blir

$$\begin{aligned} I_p: \quad & [\bar{x} - \lambda_{0.025}(15)\,20/4\ ,\ \bar{x} + \lambda_{0.025}\,20/4] \\ & = [212.4 - 1.96 \cdot 5\ ,\ 212.4 + 1.96 \cdot 5] \approx [202.6\ ,\ 222.2] \end{aligned}$$

7. a) Ankomsterna sker enligt en Poisson process N_t sådan att det förväntade antalet under 15 minuter är $E(N_{15}) = 15/10 = 1.5$, och därför $N_{15} \sim \text{Po}(1.5)$. Sannolikheten för minst en inloggning under perioden blir

$$P(N_{15} \geq 1) = 1 - P(N_{15} = 0) = 1 - e^{-15/10} = 0.7769$$

b) Modellen som beskrivs är M/M/∞. I jämvikt är tillståndsvariabeln X = antal användare, Poissonfördelad med väntevärde som är produkten av intensiteten $\lambda = 1/10$ per minut och förväntade systemtiden som var 30 min, dvs $30/10 = 3$ och $X \sim \text{Po}(3)$.

c) Jämviktssannolikheten för att systemet är tomt, dvs $P(X = 0) = e^{-3} = 0.0498$.