UPPSALA UNIVERSITET | PROV I MATEMATIK
Matematiska Institutionen | Sannolikhet och statistik, 1MS321
IK | 2017-06-01, Lösningsförslag

1. Enligt definition av kovarians gäller

$$C(X^2, 1/X) = E(X^2 \cdot 1/X) - E(X^2)\,E(1/X) = E(X) - E(X^2)E(1/X).$$

För den aktuella fördelningen finner man

$$E(1/X) = \sum_{k=1}^{4} \frac{1}{k}\frac{k}{10} = \frac{1}{10}\big(1+1+1+1\big) = \frac{4}{10}$$

samt

$$E(X) = \sum_{k=1}^{4} k \cdot \frac{k}{10} = \frac{1}{10}\big(1+4+9+16\big) = 3, \quad E(X^2) = \frac{1}{10}\big(1+8+27+64\big) = 10$$

varför

$$C(X^2, 1/X) = E(X) - E(X^2)E(1/X) = 3 - 10 \cdot \frac{4}{10} = -1.$$

2. a) Variabeln X har samma fördelning som antal lyckade om man utför tio oberoende försök som vart och ett lyckas med sannolikhet p, dvs binomialfördelningen $\mathrm{Bin}(10, p)$.

b) Om $X = 0$ finns inget paket att skicka, om $X = 1$ skickas ett paket, om $2 \leq X \leq 10$ uppstår kollision och inget paket överförs. Den sökta sannolikheten är alltså

$$P(X = 1) = \binom{10}{1} p^1 (1-p)^9 = 10p(1-p)^9$$

c) Systemets prestanda beskrivs exempelvis av

$$Z = \text{antal paket som överförs under en slot} = 1_{\{X=1\}}$$

och det genomsnittliga antalet

$$E(Z) = E(1_{\{X=1\}}) = 1 \cdot P(X = 1) = 10p(1-p)^9.$$

Genom att skissera $E(Z)$ som funktion av p ser man att det finns ett optimalt $p = p_{\max}$ där $E(Z)$ är maximal. Exakt värde på $p_{\max}$ fås genom derivering:

$$\frac{d}{dp} 10p(1-p)^9 = 10(1-p)^9 - 90p(1-p)^8 = 0 \quad \Rightarrow \quad (1-p) = 9p \quad \Rightarrow p_{\max} = 1/10$$

d) För Y = antal paket per slot som fördröjs, fås

$$\begin{aligned} E(Y) &= 2P(X = 2) + 3P(X = 3) + \cdots + 10P(X = 10) \\ &= E(X) - P(X = 1) = 10p - 10p(1-p)^9 \end{aligned}$$

vilket för $p = 0.05$ ger ca 0.185 fördröjda paket per slot.

3. Inversmetoden innebär att man väljer slumptalen x_k så att $F(x_k) = u_k$, där F är fördelningsfunktionen för den önskade slumpvariabeln X. Den angivna beräkningsmetoden kan skrivas om på formen $u_k = 1 - (1 - x_k/6)^2$, vilket innebär $F(x) = 1 - (1 - x/6)^2$. Täthetsfunktionen för X är alltså

$$f(x) = F'(x) = \frac{1}{3}(1 - \frac{x}{6}), \quad 0 \leq x \leq 6.$$

4. Vi antar att variablerna

$$X_i = \text{vikt av person } i \sim N(70, 100), \quad i = 1, 2, \ldots$$

är oberoende. Eftersom en summa av normalfördelade variabler fortfarande är normalfördelad, gäller Y_n = totalvikt av n personer $= \sum_{i=1}^n X_i \sim N(70n, 100n)$. Villkoret är att $P(Y_n > 1000) \leq 0.01$, dvs

$$P(Y_n \leq 1000) = P(\frac{Y_n - 70n}{10\sqrt{n}} \leq \frac{1000 - 70n}{10\sqrt{n}}) = \Phi(\frac{1000 - 70n}{10\sqrt{n}}) > 0.99,$$

vilket är samma sak som $\frac{1000-70n}{10\sqrt{n}} \geq \lambda_{0.01} = 2.326$. Prövning med tänkbara n ger villkoret $n \leq 13$.

5. a) $$E(X) = \int_0^r x f_X(x)\, dx = \frac{2}{r^2}\int_0^r x^2\, dx = \frac{2}{r^2}\frac{r^3}{3} = \frac{2r}{3}$$

b) $$E(X^2) = \int_0^r x^2 \frac{2x}{r^2}\, dx = \frac{2}{r^2}\frac{r^4}{4} = \frac{r^2}{2}, \quad D(X) = \sqrt{E(X^2) - E(X)^2} = \frac{r}{\sqrt{18}}$$

c) Vi vet att $\bar{x}$ är en punktskattning av $E(X) = 2r/3$ och därför blir $\widehat{r} = 3\bar{x}/2$ en väntevärdesriktig punktskattning av r. Med $\bar{x} = 3.4$ km fås $\hat{r} = 5.1$ km.

d) Motsvarande estimator är $\widehat{r} = 3\bar{X}/2$. Medelfelet $d[\widehat{r}]$ kan beräknas enligt

$$D(\widehat{r}) = \frac{3}{2}D(\bar{X}) = \frac{3}{2}\frac{D(X)}{\sqrt{20}} = \frac{3}{2}\frac{r}{\sqrt{18}}\frac{1}{\sqrt{20}} = \frac{r}{4\sqrt{10}}, \quad d[\widehat{r}] = \frac{5.1}{4\sqrt{10}} = 0.403$$

e) Centrala gränsvärdessatsen ger ett ungefärligt 95% konfidensintervall för r på formen

$$I_r = [\widehat{r} \pm \lambda_{0.025} \cdot d[\widehat{r}]] = [5.1 \pm 1.96 \cdot 0.403] = [4.31\,,\, 5.89] \quad \text{(enhet km)}$$

f) Undersökningen påvisar att leverantörens påstående kan bekräftas, eller åtminstone "inte avvisas", eftersom parametern r bedöms ligga i ett intervall där alla värden överstiger 4 km.

6. a) Begynnelsefördelningen ges av radvektorn $p^{(0)} = (1/2 \;\; 0 \;\; 1/2 \;\; 0)$. Efter att Markovkedjan gjort två övergångar får vi fördelningen $p^{(2)}$ enligt regeln

$$p^{(2)} = p^{(0)}\mathbf{P}^2 = (1/2 \;\; 0 \;\; 1/2 \;\; 0)\begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 5/6 & 1/6 \\ 0 & 1/2 & 1/4 & 1/4 \\ 2/3 & 0 & 1/3 & 0 \end{pmatrix} = (0 \;\; 1/4 \;\; 3/8 \;\; 3/8)$$

dvs sannolikhet $3/8 = 0.375$ att kedjan befinner sig i tillstånd 4 efter två steg. Alternativt tänker man efter utan matrisberäkning hur kedjan kan ta sig från 1 till 4 på två steg (sannolkhet 1/2) eller från 3 till 4 på två steg (sannolikhet 1/4) och viktar dessa möjligheter lika.

b) När det finns en asymptotisk fördelning $\pi = (\pi_1 \; \pi_2 \; \pi_3 \; \pi_4)$ så hittar men den som lösning till ekvationssystemet $\pi\,\mathbf{P} = \pi$, dvs ekvationerna

$$\frac{2}{3}\pi_2 = \pi_1, \quad \pi_4 = \pi_2, \quad \pi_1 + \frac{1}{3}\pi_2 + \frac{1}{2}\pi_3 = \pi_3, \quad \frac{1}{2}\pi_3 = \pi_4$$

samt $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. Lös, till exempel, genom

$$\pi_3 = 2\pi_4,\; \pi_2 = \pi_4,\; \pi_1 = \frac{2}{3}\pi_2 = \frac{2}{3}\pi_4 \;\;\Rightarrow\;\; 1 = (\frac{2}{3} + 1 + 2 + 1)\pi_4 \;\;\Rightarrow\;\; \pi = (\frac{2}{14} \; \frac{3}{14} \; \frac{6}{14} \; \frac{3}{14})$$

7. Vi har observationer $x_1, \ldots, x_6$ från en normalfördelning $\mathrm{N}(\mu, \sigma^2)$ och beräknar punktskattningar $\bar{x}$ av μ och s av σ, enligt

$$\bar{x} = \frac{1}{6}(x_1 + \cdots + x_6) = 11.3567, \quad s^2 = \frac{1}{5}\sum_{k=1}^{6}(x_k - \bar{x})^2 = 2.8623, \quad s = 1.6918$$

Ett (exakt) konfidensintervall för μ med konfidensgrad 0.95 erhålles nu från

$$I_\mu = [\bar{x} \pm t_{0.025}(5) \cdot \frac{s}{\sqrt{6}}] = [11.3567 \pm 2.57 \cdot \frac{1.6918}{\sqrt{6}}] = [11.3567 \pm 1.7751] \;\;\Rightarrow\;\; I_\mu = [9.58\,,\, 13.14]$$

Eftersom intervallet täcker över den efterfrågade tidsgränsen 12 sek är slutsatsen att mjukvaran inte kan anses uppfylla kriteriet, med konfidensgrad 0.95.