Institutionen för matematik och statistik
Introduktion till sannolikhetskalkyl
Lösningsförslag till räkneövning 6
23.2.2015

1. (3:1) Beräkna väntevärdena $\mathcal{E}(X^4)$ och $\mathcal{E}(\sin(2\pi X)$ då $X \sim Tas(0,1)$.
   Lösning: Eftersom $X \sim Tas(0,1)$ så är frekvensfunktionen för X

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{annars.} \end{cases}$$

   Vi vet även att om $g\colon \mathbb{R} \to \mathbb{R}$ är kontinuerlig så är

$$\mathcal{E}(g \circ X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

   (i) Eftersom $g(x) = x^4$ är kontinuerlig så är

$$\mathcal{E}(X^4) = \mathcal{E}(g \circ X) = \int_{-\infty}^{\infty} x^4 f_X(x) dx = \int_0^1 x^4 dx = \frac{1}{5}.$$

   (ii) Eftersom $g(x) = \sin(2\pi x)$ är kontinuerlig så är

$$\mathcal{E}(\sin(2\pi X) = \mathcal{E}(g \circ X) = \int_{-\infty}^{\infty} \sin(2\pi x) f_X(x) dx = \int_0^1 \sin(2\pi x) dx = 0.$$

$\square$

2. Visa att variansen $Var(Y) = \sigma^2$ då $Y \sim N(\mu, \sigma^2)$. *Tips:* visa först att $Var(X) = 1$ då $X \sim N(0,1)$ är standardnormalfördelad genom att partiellt integrera

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx$$

   med uppdelningen $x^2 e^{-x^2/2} = x \cdot (x e^{-x^2/2})$.
   Lösning: Låt $X \sim N(0,1)$. Nu är $\mathcal{E}(X) = 0$ och

$$Var(X) = \mathcal{E}(X^2) - (\underbrace{\mathcal{E}(X)}_{0})^2 = \mathcal{E}(X^2)$$

   Enligt tipset beräknar vi nu variansen av X och erhåller

$$Var(X) = \mathcal{E}(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{2\pi} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1.$$

   Se modellsvaren från övning 5 uppgift 4(ii) för partiell integreringen för att få resultatet

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \sqrt{2\pi}.$$

Vi vet alltså att variansen av en standardfördelad slumpvariabel är 1. Nu om $Y \sim N(\mu, \sigma^2)$ så känner vi till följande egenskap hos Y:

$$Var(aY + b) = a^2 Var(Y), \text{ för alla } a, b \in \mathbb{R}.$$

Detta innebär att variansen hos normalfördelade slumpvariablar inte beror på konstant translationer och är kvadratisk. Eftersom $Y \sim N(\mu, \sigma^2)$ så är

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1).$$

Från detta kan vi erhålla följande implikationskedja:

$$\begin{aligned}
& Var(Z) = 1 \\
\Rightarrow \quad & Var\left(\frac{Y-\mu}{\sigma}\right) = 1 \qquad | \cdot \sigma^2 \\
\Rightarrow \quad & \sigma^2 Var\left(\frac{Y-\mu}{\sigma}\right) = \sigma^2 \\
\Rightarrow \quad & Var(Y - \mu) = \sigma^2 \\
\Rightarrow \quad & Var(Y) = \sigma^2.
\end{aligned}$$

□

3. (3:8) Anta att de stokastiska variablerna X, Y och Z är oberoende, samt att alla har samma väntevärde μ och samma varians σ^2. Beräkna väntevärdet och variansen till följande stokastiska variabler:

   (i) $2X + 3$, (ii) $X - Y$ (iii) $X - \frac{1}{2}Y$, (iv) $X + 2Y + 3Z$, (v) $X \cdot Y$.

   Påminnelse: om X och Y är oberoende, så är också $g(X)$ och $h(Y)$ oberoende för alla kontinuerliga funktioner g och h på $\mathbf{R}$.

   Lösning: Vi vet att $\mathcal{E}(X) = \mathcal{E}(Y) = \mathcal{E}(Z) = \mu$ och $Var(X) = Var(Y) = Var(Z) = \sigma^2$. I uppgiften använder vi oss av att väntevärdet är linjär, alltså att

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y), \text{ för alla } a, b \in \mathbb{R}$$

   då X och Y är slumpvariabler. Därtill vet vi att variansen är kvadratisk, alltså att

$$Var(aX + b) = a^2 Var(X), \text{ för alla } a, b \in \mathbb{R}$$

   då X är en slumpvariabel. Om X och Y är oberoende slumvariabler och $g(x)$ och $h(x)$ är kontinuerliga funktioner, så är även $g \circ X$ och $f \circ Y$ oberoende. Detta innebär att variansen uppfyller följande egenskap:

$$Var(g \circ X + h \circ Y) = Var(g \circ X) + Var(h \circ Y).$$

   (i)

$$\mathcal{E}(2X + 3) = 2\mathcal{E}(X) + 3 = 2\mu + 3,$$
$$Var(2X + 3) = 4Var(X) = 4\sigma^2.$$

(ii) Eftersom X och Y är oberoende så är även X och $-Y$ oberoende (ty $g(x) = -x$ är kontinuerlig).

$$\mathcal{E}(X - Y) = \mathcal{E}(X) - \mathcal{E}(Y) = \mu - \mu = 0,$$

$$Var(X - Y) = Var(X + (-Y)) = Var(X) + Var(-Y) = Var(X) + Var(Y) = 2\sigma^2.$$

(iii) Eftersom X och Y är oberoende så är även X och $-1/2Y$ oberoende (ty $g(x) = -1/2x$ är kontinuerlig).

$$\mathcal{E}(X - \frac{1}{2}Y) = \mathcal{E}(X) - \frac{1}{2}\mathcal{E}(Y) = \frac{1}{2}\mu,$$

$$Var(X - \frac{1}{2}Y) = Var(X + (-\frac{1}{2}Y)) = Var(X) + \frac{Var(Y)}{4} = \frac{5}{4}\sigma^2.$$

(iv) Eftersom X, Y och Z är oberoende så är även X, $2Y$ och $3Z$ oberoende.

$$\mathcal{E}(X + 2Y + 3Z) = \mathcal{E}(X) + 2\mathcal{E}(Y) + 3\mathcal{E}(Z) = 6\mu,$$

$$\begin{aligned} Var(X + 2Y + 3Z) &= Var(X) + Var(2Y) + Var(3Z) \\ &= Var(X) + 4Var(Y) + 9Var(Z) \\ &= 14\sigma^2. \end{aligned}$$

(v) Eftersom X och Y är oberoende och så kan väntevärdet av produkten delas upp i produkten av väntevärden. Vi använder oss även av formeln

$$Var(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2.$$

Då får vi att

$$\mathcal{E}(X \cdot Y) = \mathcal{E}(X) \cdot \mathcal{E}(Y) = \mu^2$$

och

$$\begin{aligned} Var(X \cdot Y) &= \mathcal{E}((X \cdot Y)^2) - (\mathcal{E}(X \cdot Y))^2 \\ &= \mathcal{E}(X^2) \cdot \mathcal{E}(Y^2) - \mu^4 \\ &= (Var(X) + \mathcal{E}(X)^2)(Var(Y) + \mathcal{E}(Y)^2) - \mu^4 \\ &= (\sigma^2 + \mu^2)^2 - \mu^4 \\ &= \sigma^4 + 2\sigma^2\mu^2. \end{aligned}$$

□

4. (3:15) Anta att X och Y är stokastiska variabler med (de ändliga) varianserna $Var(X)$ och $Var(Y)$. Visa att följande villkor är ekvivalenta:

(i) $Cov(X, Y) = 0$,

(ii) $\mathcal{E}(X \cdot Y) = \mathcal{E}(X) \cdot \mathcal{E}(Y)$,

(iii) $Var(X + Y) = Var(X) + Var(Y)$.

Lösning: För att visa att dessa tre egenskaper är ekvivalenta bör vi visa att det existerar en implikationskedja som binder samman dem. Det måste vara möjligt att från vilken som helst av de tre egenskaper kunna härleda vilken som helst annan. Detta kan göras på många olika sätt. I detta lösningsförslag visar vi ekvivalensen (i) $\Leftrightarrow$ (ii) och ekvivalensen (ii) $\Leftrightarrow$ (iii). Låt X och Y vara slumpvariabler med ändliga varianser $Var(X)$ och $Var(Y)$, $\mathcal{E}(X) = \mu_1$ och $\mathcal{E}(Y) = \mu_2$. Nu är

$$\begin{aligned} Cov(X, Y) &= \mathcal{E}((X - \mu_1)(Y - \mu_2)) \\ &= \mathcal{E}(XY - X\mu_2 - \mu_1 Y + \mu_1\mu_2) \\ &= \mathcal{E}(XY) - \mu_2\mathcal{E}(X) - \mu_1\mathcal{E}(Y) + \mu_1\mu_2 \\ &= \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) \end{aligned}$$

Om vi nu antar att (i) är sann så följer (ii) direkt, och om vi antar att (ii) är sann så följer (i) också direkt, alltså gäller (i) $\Leftrightarrow$ (ii). För den andra ekvivalensen härleder vi följande likhet:

$$\begin{aligned} Var(X + Y) &= \mathcal{E}((X + Y)^2) - (\mathcal{E}(X + Y))^2 \\ &= \mathcal{E}(X^2) + 2\mathcal{E}(XY) + \mathcal{E}(Y^2) - (\mathcal{E}(X) + \mathcal{E}(Y))^2 \\ &= \underbrace{\mathcal{E}(X^2) - \mathcal{E}(X)^2}_{Var(X)} + 2\mathcal{E}(XY) + \underbrace{\mathcal{E}(Y^2) - \mathcal{E}(Y)^2}_{Var(Y)} - 2\mathcal{E}(X)\mathcal{E}(Y) \\ &= Var(X) + Var(Y) + 2(\mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y)). \end{aligned}$$

Om vi nu antar (ii) så följer (iii) direkt ur denna likhet. Om vi antar (iii) så följer det även att (ii) är sann, alltså gäller (ii) $\Leftrightarrow$ (iii). Därmed är de villkoren ekvivalenta.

$\square$

5. (3:11) Jämför den uppskattning Chebyshevs olikhet

$$P(\{|X - \mu| \geq k\sigma\}) \leq \frac{1}{k^2}$$

ger med de exakta sannolikheterna i fallen $k = 2$ och $k = 3$, då

(i) $X \sim Exp(\lambda)$, (ii) $X \sim N(\mu, \sigma^2)$.

Lösning: Chebyshevs olikhet ger en övre gräns på sannolikheten av att en slumpvariabel ger värden som varierar från väntevärdet med högst k stycken standardavvikelser.

(i) Eftersom $X \sim Exp(\lambda)$ så vet vi att $\mathcal{E}(X) = 1/\lambda$ och $Var(X) = 1/\lambda^2$ (alltså $\sigma = \mu$). Då $k = 2$ och $k = 3$ så får vi med Chebyshevs olikhet följande uppskattningar:

$$P\left(|X - \frac{1}{\lambda}| \geq 2 \cdot \frac{1}{\lambda}\right) \leq \frac{1}{4},$$
$$P\left(|X - \frac{1}{\lambda}| \geq 3 \cdot \frac{1}{\lambda}\right) \leq \frac{1}{9}.$$

Nu beräknar vi de exakta sannolikheterna då $k = 2, 3$.

$$\begin{aligned}
P\left(|X - \frac{1}{\lambda}| \geq \frac{k}{\lambda}\right) &= P\left(\{X - \frac{1}{\lambda} \geq \frac{k}{\lambda}\} \cup \{X - \frac{1}{\lambda} \leq -\frac{k}{\lambda}\}\right) \\
&= P\left(X \geq \frac{k+1}{\lambda}\right) + P\left(X \leq \frac{-k+1}{\lambda}\right) \\
&= 1 - P\left(X \leq \frac{k+1}{\lambda}\right) + P\left(X \leq \frac{-k+1}{\lambda}\right) \\
&= 1 - F_X\left(\frac{k+1}{\lambda}\right) + F_X\left(\frac{-k+1}{\lambda}\right)
\end{aligned}$$

Vi vet att fördelningsfunktionen för en $Exp(\lambda)$-fördelad slumvariabel är

$$F_X(x) = \begin{cases} 0, & \text{då } x \leq 0 \\ 1 - e^{-x\lambda}, & \text{då } x > 0 \end{cases}$$

vilket betyder att

$$P\left(|X - \frac{1}{\lambda}| \geq \frac{k}{\lambda}\right) = 1 - \underbrace{F_X\left(\frac{k+1}{\lambda}\right)}_{1-e^{-(k+1)}} + \underbrace{F_X\left(\frac{-k+1}{\lambda}\right)}_{0} = e^{-(k+1)}$$

då $k = 2, 3$. Detta betyder att

$$\begin{aligned}
P\left(|X - \frac{1}{\lambda}| \geq 2 \cdot \frac{1}{\lambda}\right) &= e^{-3} \approx 0{,}04979 \leq \frac{1}{4}, \\
P\left(|X - \frac{1}{\lambda}| \geq 3 \cdot \frac{1}{\lambda}\right) &= e^{-4} \approx 0{,}01832 \leq \frac{1}{9}.
\end{aligned}$$

(ii) Låt $X \sim N(\mu, \sigma^2)$. Då $k = 2$ och $k = 3$ så får vi med Chebyshevs olikhet följande uppskattningar:

$$\begin{aligned}
P\left(|X - \mu| \geq 2\sigma\right) &\leq \frac{1}{4}, \\
P\left(|X - \mu| \geq 3\sigma\right) &\leq \frac{1}{9}.
\end{aligned}$$

Nu beräknar vi de exakta sannolikheterna då $k = 2, 3$. Låt

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Nu kan vi erhålla

$$\begin{aligned}
P\left(|X - \mu| \geq k\sigma\right) &= P\left(X - \mu \geq k\sigma\right) + P\left(X - \mu \leq -k\sigma\right) \\
&= P\left(\frac{X - \mu}{\sigma} \geq k\right) + P\left(\frac{X - \mu}{\sigma} \leq -k\right) \\
&= P\left(Z \geq k\right) + P\left(Z \leq -k\right) \\
&= 1 - P\left(Z \leq k\right) + P\left(Z \leq -k\right) \\
&= 1 - \Phi(k) + \Phi(-k) \\
&= 1 - \Phi(k) + 1 - \Phi(k)
\end{aligned}$$

då $k = 2, 3$. Detta betyder att

$$P(|X - \mu| \geq 2\sigma) \approx 0,04550 \leq \frac{1}{4},$$
$$P(|X - \mu| \geq 3\sigma) \approx 0,02700 \leq \frac{1}{9}.$$

$\square$

6. Man avrundar de reella talen $x_1, \ldots, x_n$ till heltal $y_1, \ldots, y_n$ när man beräknar det aritmetiska medeltalet av talen. Anta att avrundningsfelen är stokastiska variabler som är oberoende av varandra samt $Tas(-\frac{1}{2}, \frac{1}{2})$-fördelade. Låt $X = \frac{1}{n}\sum_{k=1}^{n}(x_k - y_k)$ vara det aritmetiska medeltalet av felen. Uppskatta med hjälp av Chebyshevs olikhet hur stort n bör vara för att sannolikheten

$$P(\{|X| \geq 0.01\}) < 0.05.$$

Lösning: Låt oss beteckna $X_k := x_k - y_k$. Vi vet att $X_k \sim Tas(-1/2, 1/2)$ vilket betyder att

$$f_{X_k}(x) = \begin{cases} 1, & \text{då } -1/2 < x < 1/2 \\ 0, & \text{annars.} \end{cases}$$

Då märker vi att

$$\mathcal{E}(X_k) = \int_{-1/2}^{1/2} x f_X(x)dx = \int_{-1/2}^{1/2} x dx = 0, \quad \text{för alla } k \in \{1, \ldots, n\}.$$

Eftersom väntevärdet är en linjär operation kan vi erhålla.

$$\mathcal{E}(X) = \mathcal{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathcal{E}(X_k) = \frac{1}{n}\cdot n \cdot 0 = 0$$
$$\Rightarrow \mu = 0.$$

Nu har vi allt för att beräkna variansen av X. Först beräknar vi

$$Var(X_k) = \mathcal{E}({X_k}^2) - (\underbrace{\mathcal{E}(X_k)}_{0})^2 = \int_{-1/2}^{1/2} x^2 f_X(x)dx = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$$

och där efter erhåller vi variansen (notera att X_k är oberoende för alla $k \in \{1, \ldots, n\}$):

$$Var(X) = Var(\frac{1}{n}\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} Var(\frac{1}{n}X_k) = \sum_{k=1}^{n}\frac{1}{n^2}Var(X_k) = \frac{1}{n^2}\cdot n \cdot \frac{1}{12} = \frac{1}{12n}$$
$$\Rightarrow \sigma = \sqrt{\frac{1}{12n}}.$$

Nu får vi genom att använda Chebyshevs olikhet uppskattningen

$$P\left(\{|X| \geq k\sqrt{\frac{1}{12n}}\}\right) \leq \frac{1}{k^2}$$

För att kunna använda denna olikhet för att hitta svaret till vår fråga måste vi välja vårt k på passande sätt. Låt oss välja $k = \sqrt{12n}/100$ eftersom då är

$$P\left(\{|X| \geq 0,01\}\right) \leq \frac{1}{k^2}$$

och då räcker det för oss att beräkna när $1/k^2 < 0,05$ och från det hitta en uppskattning för vad n skall vara.

$$\begin{aligned}
& \frac{1}{k^2} < 0,05 \\
\Rightarrow \quad & k^2 > \frac{1}{0,05} = 20 \\
\Rightarrow \quad & \frac{12n}{10000} > 20 \\
\Rightarrow \quad & n > \frac{20 \cdot 10000}{12} = 16666\frac{2}{3}
\end{aligned}$$

Detta betyder att om $n \geq 16667$ så gäller olikheten

$$P(\{|X| \geq 0.01\}) < 0.05.$$

$\square$