LINKÖPINGS UNIVERSITET
Matematiska institutionen
Johan Thim

Kurskod: TAMS79
Provkod: TEN1
Datum: 2019-01-17
Institution: MAI

Tentamen i matematisk statistik, TAMS79/TEN1 2019-01-17 (4h)

Hjälpmedel är: miniräknare med tömda minnen och formelsamling ”Formel- och tabellsamling i matematisk statistik TAMS65 (Martin Singull)” och/eller formelsamling i matematisk statistik utgiven av matematiska institutionen. Inga anteckningar i formelsamlingarna är tillåtna. Rimliga ordböcker är tillåtna (till exempel svenska och persiska).
Varje uppgift är värd 6 poäng. För godkänd tentamen räcker 15 poäng. För betyg 4/5 räcker 20 respektive 26 poäng (således tillräckliga krav). Noggrann motivering krävs där alla viktiga detaljer skall motiveras. Approximationer är tillåtna om dessa är rimliga och motiveras noggrant. Uppgifterna är i nummerordning.

---

1. (a) Låt A och B vara två händelser så att $P(A) = 0.4$, $P(B) = 0.6$ och $P(A \cap B) = 0.36$. Beräkna $P(A \cup B)$, $P(A \cap B^*)$ samt avgör om A och B är oberoende. (3p)

   (b) Visa att om C och D är oberoende händelser med $P(D) > 0$ så är $P(C \mid D) = P(C)$. (1p)

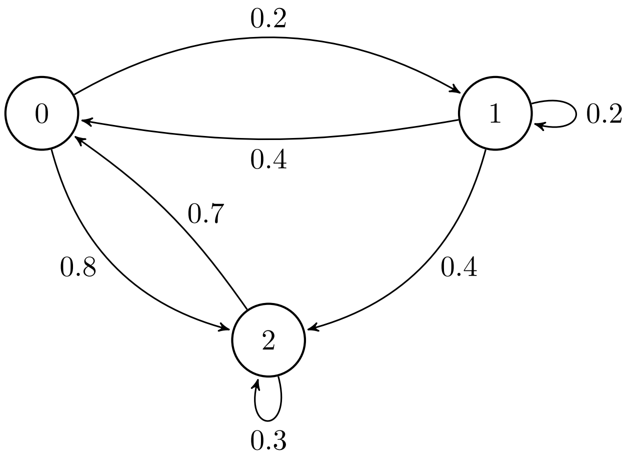
   (c) I tre fabriker M_1, M_2 och M_3 tillverkas en speciell fosforsyraester för tveksamt bruk. Samtliga fabriker tillverkar en liter åt gången (för att minska risken vid en olycka) och förpackar dessa individuellt i behållare (en liter i varje). Felrisken för en tillverkningsomgång (dvs för en liter) vid M_1 är 0.03, felrisk vid M_2 är 0.05 och felrisken för M_3 är 0.065. På en dag tillverkar man 200 liter i M_1, 150 liter i M_2 och 300 liter i M_3. Bestäm sannolikheten att en vid dagens slut slumpmässigt utvald behållare, som visar sig innehålla en felaktig vätska, kommer från M_2. (2p)

2. (a) Låt $X_1 \sim N(1, \sqrt{2})$, $X_2 \sim N(-1, \sqrt{2})$, och $X_3 \sim N(0, 1)$, vara oberoende stokastiska variabler ($V(X_1) = V(X_2) = 2$). Beräkna $P(X_1 + X_2 - X_3 \leq \frac{2}{5})$. (2p)

   (b) Låt $X \sim N(1, 2)$ ($V(X) = 4$). Bestäm $a \in \mathbf{R}$ så att $P(X > a) = 0.98$. (2p)

   (c) Låt $X_i \sim N(1, 2)$, $i = 1, 2, \ldots$, vara en följd av oberoende stokastiska variabler ($V(X_i) = 4$). Hur många termer måste man ta med i ett medelvärde för att få $V(\overline{X}) < \frac{1}{10}$? (2p)

3. Låt en Markovkedja ges av följande diagram:



   (a) Låt ingångssannolikheterna ges av $p(0) = (\,0,\ 0.5,\ 0.5\,)$. Bestäm sannolikheten att efter tre steg befinna sig i tillstånd två. (3p)

(b) Bestäm sannolikheterna att befinna sig i de tre tillstånden efter lång tid. (3p)

4. Den stokastiska variabeln (X, Y) har täthetsfunktionen

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{3}(xy+1)e^{-y}, & 0 < x < 1,\ 0 < y < \infty, \\ 0, & \text{för övrigt.} \end{cases}$$

(a) Beräkna de marginella tätheterna $f_X(x)$ och $f_Y(y)$. Är X och Y oberoende och vad blir den betingade fördelningen $f_{X\,|\,Y=y}(x)$? (3p)

(b) Antag att vi har 10 oberoende variabler (X_i, Y_i), $i = 1, 2, \ldots, 10$, var och en med fördelningen ovan. Vad är sannolikheten att högst 8 av dessa uppfyller att $Y_i > X_i$? (3p)

5. Låt den stokastiska variabeln X ha täthetsfunktionen $f_X(x) = 4(2x - x^3)/3$ för $0 \leq x \leq 1$ och $f_X(x) = 0$ för övrigt.

(a) Låt Y vara medelvärdet av 50 oberoende variabler med fördelningen ovan. Vad är sannolikheten att medelvärdet understiger 0.6? (4p)

(b) Låt $\lambda = 1/E(X)$ och definiera Poissonprocessen $Z(t) \sim \text{Po}(\lambda t)$. Hitta det minsta talet $a > 0$ så att $P(Z(t) > 33) \geq 0.90$ då $t \geq a$. (2p)

6. Låt $X_1, X_2, \ldots$ vara en följd oberoende stokastiska variabler och låt $X_i \sim \text{Exp}(\lambda)$ för $i = 1, 2, \ldots$ Definiera $Y_n = \max\{X_1, X_2, \ldots, X_n\}$ för $n = 1, 2, 3, \ldots$

(a) Hitta täthetsfunktionen för Y_n (om den existerar). (3p)

(b) Visa att $Y_n - \dfrac{\ln n}{\lambda}$ konvergerar i fördelning (ange även den asymptotiska fördelningen). (3p)

Lösningsskisser

1. (a) Om A och B är oberoende så är $P(A\cap B)=P(A)P(B)=0.24$, men då $P(A\cap B)=0.36$ så kan *inte* A och B vara oberoende. Vidare gäller att (se ett Venndiagram till exempel)

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.4+0.6-0.36=0.64.$$

Ur till exempel ett Venndiagram till följer det att

$$P(A\cap B^*)=P(A)-P(A\cap B)=0.4-0.36=0.04.$$

(b) Direkt från definitionen av betingad sannolikhet ser vi att

$$P(C\,|\,D)=\frac{P(C\cap D)}{P(D)}=\frac{P(C)P(D)}{P(D)}=P(C)$$

ty C och D är oberoende så att $P(C\cap D)=P(C)P(D)$.

(c) Det tillverkas totalt sett 650 liter på en dag. Från detta erhåller vi sannolikheterna att en enhet producerats av en viss maskin: $P(M_1)=200/650=0.308$, $P(M_2)=150/650=0.231$ och $P(M_3)=300/650=0.461$. Låt T vara händelsen att enheten vi väljer ut är defekt. Enligt uppgiften gäller nu att $P(T\,|\,M_1)=0.03$, $P(T\,|\,M_2)=0.05$ och $P(T\,|\,M_3)=0.065$. Lagen om total sannolikhet ger att händelsen att en enhet vid dagens slut är trasig blir

$$P(T)=0.308\cdot 0.03+0.231\cdot 0.05+0.461\cdot 0.065=0.0508.$$

Vi söker sannolikheten $P(M_2\,|\,T)$. Vi använder Bayes sats för ”vända” på betingningen:

$$P(M_2\,|\,T)=\frac{P(T\,|\,M_2)P(M_2)}{P(T)}=\frac{0.05\cdot 0.231}{0.0508}=0.227.$$

Svar: (a) 0.64, ej oberoende, samt 0.04 (b) se ovan (c) 0.227.

2. (a) Variablerna är oberoende och normalfördelade, så

$$Z=X_1+X_2-X_3\sim N\left(1+(-1)+0,\ \sqrt{(\sqrt{2})^2+(\sqrt{2})^2+1}\right).$$

Vi beräknar:

$$P(Z\le 2/5)=\Phi\left(\frac{2/5-0}{\sqrt{5}}\right)=\Phi(2/(5\sqrt{5}))\approx\Phi(0.179)=0.571.$$

(b) Låt $Z=(X-1)/\sqrt{4}\sim N(0,1)$. Vi söker nu $a\in\mathbf{R}$ så att

$$\begin{aligned}0.98=P(X>a)&=P\left(\frac{X-1}{2}>\frac{a-1}{2}\right)=P\left(Z>\frac{a-1}{2}\right)=1-\Phi\left(\frac{a-1}{2}\right)\\&=\Phi\left(-\frac{a-1}{2}\right),\end{aligned}$$

och ur tabell finner vi att $-(a-1)/2=2.055$, eller $a=-3.11$.

(c) Vi vet att

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\sim N(0,\sqrt{\frac{4}{n}}),$$

så $V(\overline{X})=4/n$. Det följer att $V(\overline{X})<1/10$ om $n>40$. Det duger alltså med $n=41$.

Svar: (a) 0.571 (b) $a=-3.11$ (c) $n=41$.

3. Vi ställer upp P och beräknar P^3. Vi behöver denna matris för att räkna ut $p(3) = p(0)P^3$ (Chapman-Kolmogorov). Alltså,

$$P = \begin{pmatrix} 0.0 & 0.2 & 0.8 \\ 0.4 & 0.2 & 0.4 \\ 0.7 & 0.0 & 0.3 \end{pmatrix} \qquad P^2 = \begin{pmatrix} 0.64 & 0.04 & 0.32 \\ 0.36 & 0.12 & 0.52 \\ 0.21 & 0.14 & 0.65 \end{pmatrix} \qquad P^3 = \begin{pmatrix} 0.240 & 0.136 & 0.624 \\ 0.412 & 0.096 & 0.492 \\ 0.511 & 0.070 & 0.419 \end{pmatrix}$$

vilket medför att

$$p(3) = p(0)P^3 = \begin{pmatrix} 0.0 & 0.5 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0.240 & 0.136 & 0.624 \\ 0.412 & 0.096 & 0.492 \\ 0.511 & 0.070 & 0.419 \end{pmatrix} = \begin{pmatrix} 0.4615 & 0.0830 & 0.4555 \end{pmatrix}.$$

Eftersom kolumn två har alla element skilda från noll så måste den asymptotiska fördelningen vara en stationär fördelning, och dessa finner vi ur ekvationen $\underline{\pi}P = \underline{\pi}$. Vi erhåller då ekvationssystemet (om vi lägger till kravet att $\underline{\pi}$ är en stokastisk vektor)

$$\begin{cases} -\pi_0 + 0.4\pi_1 + 0.7\pi_2 = 0 \\ 0.2\pi_0 - 0.8\pi_1 = 0 \\ 0.8\pi_0 + 0.4\pi_1 - 0.7\pi_2 = 0 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = 4\pi_1 \\ -3.6\pi_1 + 0.7\pi_2 = 0 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = 0.3944 \\ \pi_1 = 0.0986 \\ \pi_2 = 0.5070 \end{cases}$$

Svar: (a) 45.6%. (b) Se $\underline{\pi}$ ovan.

4. (a) Vi beräknar de marginella tätheterna:

$$f_X(x) = \frac{2}{3}\int_0^\infty (xy+1)e^{-y}\,dy = /\text{ PI }/ = \frac{2}{3}\left(\left[-(xy+1)e^{-y}\right]_0^\infty + x\int_0^\infty e^{-y}\,dy\right)$$
$$= \frac{2}{3}(1+x), \quad 0 < x < 1,$$

samt

$$f_Y(y) = \frac{2}{3}\int_0^1 (xy+1)e^{-y}\,dx = \frac{2}{3}e^{-y}\left[\frac{x^2y}{2} + x\right]_0^1 = \frac{2}{3}e^{-y}\left(\frac{y}{2}+1\right), \quad 0 < y < \infty.$$

Vi ser att

$$f_X(x)f_Y(y) = \frac{4}{9}e^{-y}(1+x)\left(\frac{y}{2}+1\right) = \frac{2}{9}e^{-y}(y+2+xy+2x) \neq f(x,y)$$

för de flesta punkter (x, y) i området, så variablerna är inte oberoende. Den betingade fördelningen ges enligt definition av

$$f_{X\,|\,Y=y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{xy+1}{\frac{y}{2}+1} = \frac{2xy+2}{y+2},$$

för $0 < x < 1$ (och $0 < y < \infty$). Även här kan vi tydligt se att variablerna är beroende då den betingade fördelningen endast reducerar till $f_X(x)$ då $y = 1$.

(b) Vi börjar med att beräkna sannolikheten att $Y > X$. För det ändamålet åker vi på lite partialintegration:

$$P(Y > X) = \int_0^1\int_x^\infty \frac{2}{3}(xy+1)\,e^{-y}\,dydx = \frac{2}{3}\int_0^1\left(\left[-(xy+1)e^{-y}\right]_x^\infty - \int_x^\infty xe^{-y}\,dy\right)dx$$
$$= \frac{2}{3}\int_0^1 (x^2+x+1)e^{-x}\,dx = \cdots = \frac{2}{3}\left(4 - 8e^{-1}\right) \approx 0.705.$$

Låt $p = P(Y > X) \approx 0.705$ och låt N vara antalet av de 10 variablerna där $Y > X$. Eftersom följden är oberoende så kommer $N \sim \text{Bin}(10, p)$ och

$$P(N \leq 8) = 1 - P(N \geq 9) = 1 - P(N = 9) - P(N = 10) = 0.8428$$

Svar: (a) Se ovan (b) 0.843.

5. (a) Vi använder definitionen och finner att

$$E(X) = \int_0^1 \frac{4x(2x - x^2)}{3}\,dx = \cdots = \frac{28}{45}$$

och

$$E(X^2) = \int_0^1 \frac{4x^2(2x - x^2)}{3}\,dx = \cdots = \frac{4}{9}.$$

Steiners sats implicerar nu att $V(X) = E(X^2) - E(X)^2 = 116/2025 \approx 0.0573$. Enligt centrala gränsvärdessatsen gäller att

$$Y := \frac{1}{50}\sum_{k=1}^{50} X_k \overset{\text{appr.}}{\sim} \mathrm{N}\left(\frac{28}{45}, \sqrt{\frac{116}{2025 \cdot 50}}\right) \approx \mathrm{N}(0.622, 0.0338).$$

Vi erhåller nu att

$$P(Y < 0.6) \approx \Phi\left(\frac{0.6 - 0.622}{0.0338}\right) = 1 - \Phi(0.651) = 0.2575.$$

(b) Poissonprocessen $Z(t)$ har alltså intensiteten $\lambda = 1/E(X) = 45/28 \approx 1.607$. Vi söker ett tal a så att om $t \geq a$ ska $P(Z(t) > 33) = 0.90$. Om vi antar att t är tillräckligt stort för att tillåta oss att normalapproximera fördelningen för $Z(t)$ så erhåller vi att

$$\begin{aligned} 0.90 = P(Z(t) > 33) &= 1 - P(Z(t) \leq 32) = 1 - P\left(\frac{Z(t) - \lambda t}{\sqrt{\lambda t}} \leq \frac{32 - \lambda t}{\sqrt{\lambda t}}\right) \\ &\approx 1 - \Phi\left(\frac{32 - \lambda t}{\sqrt{\lambda t}}\right), \end{aligned}$$

så vi försöker lösa

$$1 - \Phi\left(\frac{32 - \lambda t}{\sqrt{\lambda t}}\right) = 0.90 \quad \Leftrightarrow \quad \Phi\left(-\frac{32 - \lambda t}{\sqrt{\lambda t}}\right) = 0.90 \quad \Leftrightarrow \quad -\frac{32 - \lambda t}{\sqrt{\lambda t}} = 1.28.$$

Låt $u = \sqrt{\lambda t} > 0$. Vi vill då lösa

$$u^2 - 1.28u - 32 = 0 \quad \Leftrightarrow \quad u \approx 6.33 \text{ eller } u \approx -5.$$

Endast $u = 6.33$ är möjlig och resulterar i att $t = \dfrac{6.33^2}{\lambda} = 24.9$. För $t \geq 24.9$ så bör sannolikheten vara uppfylld. Vi noterar också att $\lambda t \approx 6.33^2 \geq 15$ så normalapproximationen var OK.

Svar: (a) Ca 0.258. (b) $a = 24.9$.

6. (a) Eftersom $X_i \sim \text{Exp}(\lambda)$ så ges fördelningsfunktionen av $F_X(x) = 1 - e^{-\lambda x}$, $x \geq 0$, för alla $i = 1, 2, 3, \ldots$ Vi söker fördelningsfunktionen $F_Y(y)$ för Y och finner den genom

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\max\{X_1, X_2, \ldots, X_n\} \leq y) \\ &= P(X_1 \leq y \text{ och } X_2 \leq y \text{ och } \cdots \text{ och } X_n \leq y) \\ &= P(X_1 \leq y)P(X_2 \leq y)\cdots P(X_n \leq y) = \left(1 - e^{-\lambda y}\right)^n, \end{aligned}$$

eftersom variablerna X_i är oberoende. Vi kan nu derivera fram täthetsfunktionen:

$$f_Y(y) = \frac{d}{dy}\left(1 - e^{-\lambda y}\right)^n = n\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right)^{n-1}.$$

(b) Låt $Z_n = Y - \dfrac{\ln n}{\lambda}$. Vi vill visa att Z_n konvergerar i fördelning. Alltså,

$$\begin{aligned}
F_{Z_n}(z) &= P(Z_n \le z) = P\left(Y - \frac{\ln n}{\lambda} \le z\right) = P\left(Y \le z + \frac{\ln n}{\lambda}\right) = F_Y\left(z + \frac{\ln n}{\lambda}\right) \\
&= \left(1 - e^{-\lambda\left(z + \frac{\ln n}{\lambda}\right)}\right)^n = \left(1 - e^{-\lambda z - \ln n}\right)^n = \left(1 - \frac{1}{n}e^{-\lambda z}\right)^n \\
&= \exp\left(n \ln\left(1 - \frac{1}{n}e^{-\lambda z}\right)\right) = \exp\left(n\left(-\frac{1}{n}e^{-\lambda z} + O\left(\frac{1}{n^2}\right)\right)\right) \\
&\to \exp\left(-e^{-\lambda z}\right),
\end{aligned}$$

då $n \to \infty$ för varje $z \in \mathbf{R}$. Vi har nu visat att $Z_n \xrightarrow{\text{D}} Z$, där Z har fördelningsfunktionen

$$F_Z(z) = \exp\left(-e^{-\lambda z}\right) \text{ för } z \in \mathbf{R}.$$

Att detta är en fördelningsfunktion följer av att F_Z är växande, $F_Z(z) \to 0$ då $z \to -\infty$, samt att $F_Z(z) \to 1$ då $z \to \infty$.

Svar: (a) $f_Y(y) = n\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right)^{n-1}$
(b) Se ovan; asymptotisk fördelning: $F(z) = \exp\left(-e^{-\lambda z}\right)$, $z \in \mathbf{R}$.