LINKÖPINGS UNIVERSITET
Matematiska institutionen
Johan Thim

Utbildningskod: TAMS79
Modul: TEN1
Datum: 2019-08-26
Institution: MAI

Tentamen i matematisk statistik, TAMS79/TEN1 2019-08-26 (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.
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. Eva tillverkar egen ammunition till sitt handeldvapen. Mängden drivmedel som hamnar i varje patron har visat sig ha täthetsfunktionen

$$f_X(x) = \begin{cases} c\,(x-4)(9-x), & 4 \leq x \leq 9, \\ 0, & \text{för övrigt.} \end{cases}$$

Om mängden drivmedel i patronen understiger 5 grain[1] kommer kulan att få för låg hastighet och om vikten överstiger 8.5 grain så blir trycket för högt för Evas vapen så att pipan går sönder. Laddningen i varje patron anses oberoende av tidigare tillverkade patroner.

(a) Bestäm konstanten c. (1p)

(b) Om man på måfå plockar en patron, vad är sannolikheten att den har en laddning som inte ger för låg hastighet men inte heller spränger pipan? (1p)

(c) Evas vapen är en halvautomatisk pistol med en magasinkapacitet på 15 patroner. Vad är sannolikheten för att minst 2 av kulorna har för låg hastighet om man skjuter 15 skott? (2p)

(d) Eva köper en ask med 50 patroner av rätt kaliber och laddning. Eva blandar dessa med de 10 patroner hon tillverkat under eftermiddagen och drar sedan på måfå ut 15 patroner och laddar ett magasin. Vad är sannolikheten att magasinet nu innehåller fler än 3 egentillverkade patroner? (2p)

2. Låt $X_1 \sim N(0,1)$ och $X_2 \sim N(0,\sqrt{2})$ vara två oberoende stokastiska variabler (där varianserna alltså ges av $V(X_1) = 1$ och $V(X_2) = 2$).

(a) Hitta fördelningen för $Z = 1 + 2X_1 + 3X_2$ och beräkna $P(2Z < 1)$. (2p)

(b) Beräkna $P(\max\{X_1,\, X_2\} < 0)$. (2p)

(c) Låt $Y = \sqrt{|X_1|}$. Bestäm fördelningen för Y. (2p)

Vänd!

[1] 1 grain är 64.79891 mg.

3. Låt en Markovkedja ha övergångsmatrisen $P = \begin{pmatrix} 0.3 & 0.7 \\ 0.2 & 0.8 \end{pmatrix}$.

   (a) Om vi startar i tillstånd noll, vad är sannolikheterna att befinna sig i de olika tillstånden efter fyra steg? (2p)

   (b) Visa att $(2/9,\ 7/9)$ är en stationär sannolikhetsvektor till P. (2p)

   (c) Antag att en Markovkedja har kontinuerlig tid och att tiden T mellan en viss övergång och nästa är exponentialfördelad med väntevärde μ. Visa att $P(T > t+t_0 \mid T > t_0) = P(T > t)$. (2p)

4. Låt (X, Y) ha täthetsfunktionen

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{3}{32}xy, & 0 \le x \le 4,\ 0 \le y \le \sqrt{x}, \\ 0, & \text{för övrigt.} \end{cases}$$

   (a) Finn de marginella täthetsfunktionerna. (2p)

   (b) Beräkna $E(Y \mid X = x)$ och $V(Y \mid X = x)$. (2p)

   (c) Antag att vi har 100 oberoende variabler (X_i, Y_i), $i = 1, 2, \ldots, 100$, var och en med fördelningen ovan. Låt $Z_i = (Y_i \mid X_i = 3)$, $i = 1, 2, 3, \ldots, 100$, och definiera

$$Z = \sum_{i=1}^{100} Z_i.$$

   Vad blir sannolikheten $P(Z > 112)$? (2p)

5. (a) Låt X och Y vara två oberoende stokastiska variabler med

$$p_X(k) = \begin{cases} 0.1, & k = 0, \\ 0.4, & k = 1, \\ 0.5, & k = 2, \end{cases} \qquad \text{och} \qquad p_Y(k) = \begin{cases} 0.3, & k = -1, \\ 0.3, & k = 0, \\ 0.4, & k = 1. \end{cases}$$

   Bestäm sannolikhetsfunktionen för $Z = XY$ samt $E(XY)$. (3p)

   (b) Låt $Y = (X_1, X_2, X_3)$, där $p_{X_k}(k) = p_X(k)$ är sannolikhetsfunktionen för X i föregående deluppgift. Låt vidare X_1, X_2 och X_3 vara oberoende och definiera $p = P(Y = (0,1,2))$. Antag nu att vi har 10000 oberoende variabler Y_i med samma fördelning som Y. Vid en simulering så har man precis $10000 \cdot p$ stycken vektorer $(0,1,2)$ bland utfallen av de 10000 variablerna Y_i. Ifrån dessa 10000 variabler väljer man ut n stycken. Vilket är det minsta talet n för att sannolikheten att få med åtminstone 6 stycken vektorer $(0,1,2)$ är åtminstone 95%? (3p)

6. (a) Låt $X_1, X_2, X_3, \ldots$ vara en följd oberoende stokastiska variabler sådana att $X_k \sim \mathrm{Re}\left(-\dfrac{1}{k}, \dfrac{1}{k}\right)$.

   Visa att $X_k \xrightarrow{\mathrm{P}} 0$ då $k \to \infty$. (3p)

   (b) Antag att $X_n \xrightarrow{\mathrm{D}} X$ och att alla $X_1, X_2, X_3, \ldots$ samt X har täthetsfunktioner $f_{X_1}, f_{X_2}, f_{X_3}, \ldots$ samt f_X. Innebär detta att $f_{X_k}(x) \to f_X(x)$ då $k \to \infty$? Bevisa eller ange motexempel. (3p)

Lösningsskisser

1. (a) För att vara en täthetsfunktion så måste $f_X(x) \geq 0$ och

$$1 = \int_4^9 c(x-4)(9-x)\,dx = \cdots = c \cdot \frac{125}{6},$$

så $c = 6/125$.

(b) Den eftersöka sannolikheten finner vi genom kalkylen

$$\int_5^{8.5} \frac{6}{125}(x-4)(9-x)\,dx = \cdots = 0.868.$$

(c) Sannolikheten att en kula håller för låg hastighet ges av

$$p = \int_4^5 \frac{6}{125}(x-4)(9-x)\,dx = \cdots = 0.104.$$

Låt $X \sim \mathrm{Bin}(15, p)$. Vi söker sannolikheten att $X \geq 2$:

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - (1-0.104)^{15} - 15 \cdot 0.104 \cdot (1-0.104)^{14} = 0.4721.$$

(d) Vi har 60 patroner totalt sett och ska välja ut 15 stycken. Av dessa 60 är 10 egentillverkade. Låt X vara antalet av de 15 som är egentillverkade. Situationen är hypergeometriskt fördelad: $X \sim \mathrm{Hyp}(60, 15, p = 10/60)$. Vi söker sannolikheten

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - \left(\frac{\binom{10}{0}\binom{50}{15}}{\binom{60}{15}} + \frac{\binom{10}{1}\binom{50}{14}}{\binom{60}{15}} + \frac{\binom{10}{2}\binom{50}{13}}{\binom{60}{15}} \right) \\ &= 0.4812. \end{aligned}$$

Svar: (a) $c = 6/125$ (b) ca 0.868 (c) 0.4721 (d) 0.4812.

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

$$Z = 1 + 2X_1 + 3X_2 \sim N\left(1+0+0, \sqrt{2^2 \cdot 1 + 3^2(\sqrt{2})^2}\right) = N(1, \sqrt{22}).$$

Vidare gäller att

$$P(2Z < 1) = P\left(\frac{Z-1}{\sqrt{22}} < \frac{1/2-1}{\sqrt{22}}\right) = \Phi\left(-\frac{1}{2\sqrt{22}}\right) = 1 - \Phi\left(0.1066\right) = 0.4576$$

(b) Eftersom väntevärdet för både X_1 och X_2 är noll så är $P(X_1 < 0) = P(X_2 < 0) = 0.5$ och

$$P(\max\{X_1, X_2\} < 0) = P(X_1 < 0)P(X_2 < 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

(c) Låt $Y = \sqrt{|X_1|}$. Vi vet att $Y < 0$ med sannolikhet 0 (varför?). För $y \geq 0$ gäller att

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\sqrt{|X_1|} \leq y) = P(|X_1| \leq y^2) = P(-y^2 \leq X_1 \leq y^2) \\ &= \Phi(y^2) - \Phi(-y^2). \end{aligned}$$

För att hitta täthetsfunktionen så deriverar vi och utnyttjar att $\Phi'(x) = \varphi(x)$, så

$$f_Y(y) = \frac{d}{dy}\left(\Phi(y^2) - \Phi(-y^2)\right) = 2y\varphi(y^2) + 2y\varphi(-y^2) = 4y\varphi(y^2), \quad y \geq 0,$$

ty $\varphi(-t) = \varphi(t)$ för alla $t \in \mathbf{R}$. Explicit har vi nu visat att

$$f_Y(y) = \begin{cases} 0, & y < 0, \\ \dfrac{4y}{\sqrt{2\pi}} \exp\left(-\dfrac{y^4}{2}\right), & y \geq 0. \end{cases}$$

Svar: (a) $N(1, \sqrt{22})$ och 0.46 (b) 1/4 (c) se ovan.

3. (a) Vi räknar ut övergångsmatrisen av ordning fyra:

$$P^4 = \begin{pmatrix} 0.2223 & 0.7777 \\ 0.2222 & 0.7778 \end{pmatrix}.$$

Vi startar i tillstånd noll, så ingångssannolikheterna ges av $p(0) = (1\ 0)$ och

$$p(4) = p(0) \cdot P^4 = \begin{pmatrix} 0.2223 & 0.7777 \end{pmatrix}.$$

(b) Vi visar att $\underline{\pi} = \underline{\pi}P$:

$$\begin{pmatrix} 2/9 & 7/9 \end{pmatrix} \cdot P = \begin{pmatrix} 2/9 & 7/9 \end{pmatrix}.$$

Alltså är detta en stationär vektor!

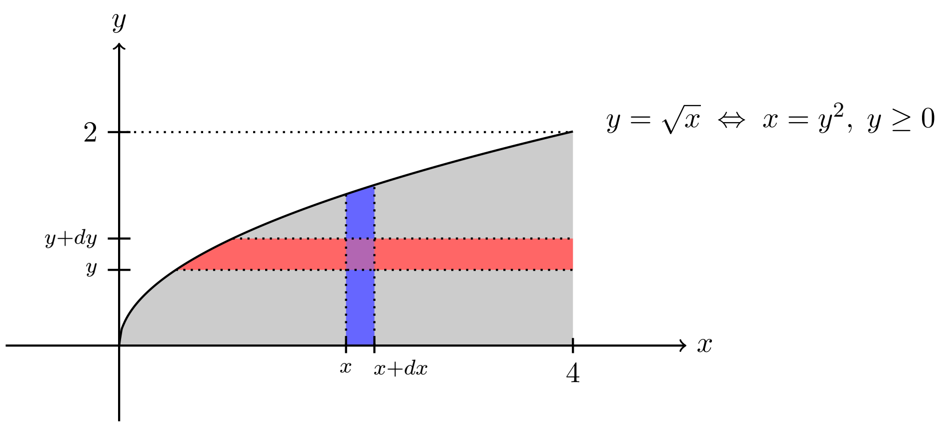
(c) Eftersom $F_T(t) = 1 - e^{-\lambda t}$ för $t \geq 0$ och något $\lambda > 0$ så gäller att

$$P\big(T > t + t_0 \mid t > t_0\big) = \frac{P(\{T > t + t_0\} \cap \{T > t_0\})}{P(T > t_0)} = \frac{P(T > t + t_0)}{P(T > t_0)} = \frac{e^{-\lambda(t_0+t)}}{e^{-\lambda t_0}}$$
$$= e^{-\lambda t} = P(T > t),$$

vilket var vad vi skulle visa.

Svar: (a) 0.22 respektive 0.78 (b) Se ovan. (c) se ovan.

4. Vi börjar med att rita hur området ser ut.



(a) Vi finner de marginella täthetern genom att ”integrera bort” den variabel vi inte vill ha kvar:

$$f_X(x) = \int_0^{\sqrt{x}} \frac{3}{32} xy\,dy = \frac{3x}{32}\left[\frac{y^2}{2}\right]_0^{\sqrt{x}} = \frac{3x^2}{64}, \quad 0 \leq x \leq 4,$$

respektive

$$f_Y(y) = \int_{y^2}^{4} \frac{3}{32} xy\,dy = \cdots = \frac{3y}{64}\left(16 - y^4\right) \quad 0 \leq y \leq 2.$$

Notera att kraven på x och y är viktiga att ha med!

(b) För ett fixt tal $x \in (0,4)$ så gäller att

$$f_{Y \mid X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{3xy/32}{3x^2/64} = \frac{2y}{x}, \quad 0 \le y \le \sqrt{x}.$$

Således kan vi beräkna de betingade väntevärdet enligt

$$E(Y \mid X=x) = \int_0^{\sqrt{x}} y\frac{2y}{x}\,dy = \cdots = \frac{2\sqrt{x}}{3}.$$

Steiners sats gäller även för betingade väntevärden, så

$$V(Y \mid X=x) = E((Y - E(Y \mid X=x))^2 \mid X=x) = E(Y^2 \mid X=x) - (E(Y \mid X=x))^2.$$

Vidare gäller att

$$E(Y^2 \mid X=x) = \int_0^{\sqrt{x}} y^2\frac{2y}{x}\,dy = \cdots = \frac{x}{2},$$

så

$$V(Y \mid X=x) = \frac{x}{2} - \frac{4x}{9} = \frac{x}{18}.$$

(c) Eftersom vi summerar många termer (betydligt fler än 30) så kommer summan Z att vara approximativt normalfördelad:

$$Z \overset{\text{appr.}}{\sim} N\left(100 \cdot \frac{2\sqrt{3}}{3}, \sqrt{100 \cdot \frac{1}{6}}\right) = N(115.47, \sqrt{16.67}).$$

Den eftersökta sannolikheten ges alltså approximativt av

$$P(Z > 112) = 1 - P(Z \le 112) = 1 - \Phi\left(\frac{112 - 115.47}{\sqrt{16.67}}\right) = \Phi\left(\frac{3.47}{\sqrt{16.67}}\right) = 0.8023.$$

Svar: (a) $f_X(x) = \frac{3x^2}{64}$, $0 \le x \le 4$, $f_Y(y) = \frac{3y}{64}\left(16 - y^4\right)$, $0 \le y \le 2$, (b) $E(Y \mid X=x) = \frac{2\sqrt{x}}{3}$ och $V(Y \mid X=x) = \frac{x}{18}$, (c) Ca 80%.

5. (a) Möjliga utfall för $Z = XY$ är $-2, -1, 0, 1, 2$. Vi ställer upp sannolikheterna direkt

	p_Z
$Z=-2$	$0.5 \cdot 0.3 = 0.15$
$Z=-1$	$0.4 \cdot 0.3 = 0.12$
$Z=0$	$0.1 + 0.3 - 0.1 \cdot 0.3 = 0.37$
$Z=1$	$0.4^2 = 0.16$
$Z=2$	$0.5 \cdot 0.4 = 0.2$

I tabellen har vi helt enkelt summerat sannolikheterna för de utfall (X, Y) som ger rätt summa (vilket är OK då dessa händelser är disjunkta men se upp för $Z = 0$..). En snabb kontroll visar att detta summerar till 1.

(b) Den simultana sannolikhetsfunktionen för Y ges av produkten

$$P(Y = (x_1, x_2, x_3)) = p_X(x_1)p_X(x_2)p_X(x_3)$$

eftersom komponenterna i Y är oberoende. Således gäller att

$$p = P(Y = (0,1,2)) = p_X(0)p_X(1)p_X(2) = 0.1 \cdot 0.4 \cdot 0.5 = 0.02.$$

Vi har 10000 utfall och bland dessa finns enligt uppgift precis $10000 \cdot 0.02 = 200$ stycken vektorer $(0, 1, 2)$. Vi väljer ut n stycken av de 10000. Låt Z vara antalet av dessa n som är $(0, 1, 2)$. Då är $Z \sim \text{Hyp}(N = 10000, n, p = 0.02)$.
För att få något som är lite enklare att hantera försöker vi oss på en normalapproximation. Vi kan göra detta i två steg genom att först anta att $n/N \leq 0.1$ för att säga att vi kan betrakta Z som binomialfördelad: $Z \overset{\text{appr.}}{\sim} \text{Bin}(n, p)$. Vi måste återkomma och kontrollera detta villkor senare! Sen antar vi att $np(1-p) \geq 10$ (med $p = 0.02$). Vi måste även kontrollera detta senare. Vi kan givetvis även direkt approximera (se formelsamlingen). Alltså har vi

$$P(Z \geq 6) = 1 - P(Z \leq 5) \approx 1 - \Phi\left(\frac{5 - np}{\sqrt{np(1-p)}}\right) = 1 - \left(1 - \Phi\left(\frac{np - 5}{\sqrt{np(1-p)}}\right)\right),$$

så vi löser

$$\Phi\left(\frac{np - 5}{\sqrt{np(1-p)}}\right) \geq 0.95 \quad \Leftrightarrow \quad \frac{np - 5}{\sqrt{np(1-p)}} \geq 1.645 \quad \Leftrightarrow \quad (np-5)^2 \geq np(1-p) \cdot 1.645^2.$$

Efter att ha stuvat om lite och ersatt $p = 0.02$ så ser vi att vi söker lösningar till

$$n^2 - 632.595n + 62500 = 0.$$

Vi finner två möjligheter: $n = 510.06$ och $n = 122.53$. Det andra alternativet är för litet för att approximationerna ska fungera (testa att räkna ut $np(1-p)$), medan $n = 511$ uppfyller både att $n/N \leq 0.1$ och att $np(1-p) \geq 10$.

Svar: (a) Se ovan. (b) $n \approx 511$.

Anmärkning. Verkligt värde blir $n = 516$, så det går givetvis att testa sig fram om man har tillgång till dator (eller mer avancerade miniräknare).

6. (a) Om $X_k \sim \text{Re}(-1/k, 1/k)$ så är

$$f_{X_k}(x) = \begin{cases} \frac{k}{2}, & -\frac{1}{k} \leq x \leq \frac{1}{k}, \\ 0, & \text{för övrigt.} \end{cases}$$

Vi ska visa att $X_k \xrightarrow{\text{P}} 0$ då $k \to \infty$. Med andra ord vill vi, för varje $\epsilon > 0$, visa att

$$\lim_{k \to \infty} P(|X_k - 0| \geq \epsilon) = 0.$$

Låt $\epsilon > 0$. Antag att k är så stort att $1/k < \epsilon$ (detta är OK då vi tänker låta $k \to \infty$ och ϵ inte förändras). Då erhåller vi att

$$P(|X_k - 0| \geq \epsilon) = 1 - P(-\epsilon < X_k < \epsilon) = 1 - \int_{-1/k}^{1/k} \frac{2}{k}\, dx = 1 - \frac{2}{k}\left(\frac{1}{k} + \frac{1}{k}\right) = 1 - 1 = 0.$$

Alltså måste $X_k \xrightarrow{\text{P}} 0$ då $k \to \infty$.

(b) Påståendet är inte sant. Låt följden ha fördelningsfunktionerna $F_{X_n}(x) = x - \dfrac{\sin(2n\pi x)}{2n\pi}$ för $0 \leq x \leq 1$. För $0 \leq x \leq 1$ så ser vi att $F_{X_n}(x) \to x$ då $n \to \infty$ eftersom sin-termen är begränsad. Vi noterar att $F_X(x) = x$ för $0 \leq x \leq 1$ är fördelningsfunktionen för en likformig fördelning på $[0, 1]$, så täthetsfunktion för X existerar. Däremot så ser vi att

$$f_{X_n}(x) = \frac{d}{dx} F_{X_n}(x) = 1 - \cos(2\pi n x) \to ???$$

då $n \to \infty$ för alla $0 < x < 1$. Gränsvärdet existerar inte ens.

Svar: (a) se ovan (b) Behöver ej vara sant; se ovan.