LINKÖPINGS UNIVERSITET
Matematiska institutionen
Johan Thim

Kurskod: TAMS79
Provkod: TEN1
Datum: 2019-04-24
Institution: MAI

Tentamen i matematisk statistik, TAMS79/TEN1 2019-04-24 (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. I ett vackert skogsparti växer 5 olika sorters bär. Alla är ungefär lika vanliga, men en av sorterna är ganska giftig.

   (a) Om man på måfå plockar 6 stycken bär, vad är sannolikheten att minst två är giftiga? (2p)

   (b) LD50 (d v s dosen 50% av testpopulation dör av) för den giftiga sorten ligger vid 25 stycken bär. Om hungrige Håkan mumsar i sig 150 bär, vad är sannolikheten att han uppnår LD50-gränsen? (2p)

   (c) Håkan vandrar vidare in i skogspartiet och stöter på Nyarlathotep. Håkan frågar om det finns fler bär någonstans och blir förflyttad till en annan värld där det finns 50 olika sorters bär. Dock är 1 av dessa sorter dödligt giftig (det räcker med ett bär och man dör omedelbums efter inmundigandet). Om Håkan äter ett bär i taget tills han plötsligt avlider, hitta det högsta antalet bär Håkan kan mumsa i sig för att med sannolikheten 0.5 fortfarande leva. (2p)

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

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

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

   (c) Låt $Y = X_1^2 + X_2^2$. Hitta fördelningen för Y. (2p)

3. Låt en Markovkedja ha övergångsmatrisen

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{10} & \frac{9}{10} \end{pmatrix}.$$

   Klassificera beständigheten hos samtliga tillstånd och finn alla stationära fördelningar. Finns det en asymptotisk fördelning (motivera)? Om så är fallet, vilken? (6p)

Vänd!

4. Låt den diskreta stokastiska variabeln (X, Y) ha sannolikhetsfunktionen

	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	0.02	0.04	0.01	0.03
$Y=1$	0.08	0.16	0.04	?
$Y=2$	0.1	?	0.05	0.15

(a) Antag att X och Y är oberoende. Bestäm de okända sannolikheterna och hitta sannolikhetsfunktionen för $Z = X + Y$. (4p)

(b) Antag istället att $C(X, Y) = 0.05$. Bestäm om möjligt de okända sannolikheterna i detta fall. Motivera annars varför det är omöjligt. (2p)

5. Den stokastiska variabeln X har täthetsfunktionen $f_X(x) = c \sin x$ för $0 \leq x \leq \pi$, där c är en lämplig konstant.

(a) Bestäm c och beräkna $E(X)$ samt medianen m för X. (3p)

(b) Låt $X_0, X_1, X_2, \ldots$ vara en oberoende följd stokastiska variabler med täthetsfunktionen f_X beskriven ovan. Antag att $N(t)$ är en Poissonprocess med intensiteten $\lambda > 0$. Låt

$$Y = \sum_{k=0}^{N(t)} X_k$$

vara summan av de $N(t)$ första variablerna i sekvensen och bestäm väntevärdet $E(Y)$ vid tidpunkten $t > 0$. (3p)

6. (a) Låt $X_1, X_2, X_3, \ldots$ vara en följd oberoende likafördelade stokastiska variabler med ändligt väntevärde μ och ändlig varians. Visa att

$$\lim_{n \to \infty} E\left(\left(\mu - \frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right) = 0.$$

(Konvergens av denna typ brukar kallas "convergence in mean square"). (4p)

(b) Antag att X är en icke-negativ kontinuerlig stokastisk variabel med en täthetsfunktion. Visa att

$$P(X \geq a) \leq \frac{E(X)}{a}, \quad a > 0.$$

(Detta är Markovs olikhet). (2p)

Lösningsskisser

1. Vi låter X vara antal giftiga bär bland n plockade på måfå. Då är $X \sim \mathrm{Bin}(n, 1/5)$ eftersom varje bär har en sannolikhet på $1/5$ att vara giftigt.

   (a) I denna deluppgift är $n = 6$ och vi är intresserade av sannolikheten att $X \geq 2$. Vi får

   $$P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - 0.65536 \approx 0.345$$

   ifrån tabell.

   (b) I det här fallet är $n = 150$, så vi är inte hjälpta av varken tabell eller direkta kalkyler (alldeles för stora siffror). Men, eftersom $np(1-p) = 24 \geq 10$ så försöker vi med en normalapproximation. Med andra ord, X är approximativt $N(30, \sqrt{24})$-fördelad. Alltså,

   $$P(X \geq 25) = 1 - P(X \leq 24) \approx 1 - \Phi\left(\frac{24-30}{\sqrt{24}}\right) \approx 1 - \Phi(-1.22) = \Phi(1.22) \approx 0.89.$$

   Sannolikheten är alltså ca 89%. Om vi använder halvkorrigering så får vi istället

   $$P(X \geq 25) \approx 1 - \Phi\left(\frac{24 + 0.5 - 30}{\sqrt{24}}\right) \approx \Phi(1.12) \approx 0.87.$$

   (c) Låt $X \sim \mathrm{Ffg}(p = 1/50)$. Vi söker det största talet n så att $P(X \geq n) = 0.5$. Tolkningen är då att Håkan dör vid bär nummer n eller högre (med sannolikhet 0.5). Alltså,

   $$P(X \geq n) = \sum_{k=n}^{\infty} p(1-p)^{k-1} = p(1-p)^{n-1} \sum_{k=0}^{\infty} (1-p)^k = p(1-p)^{n-1}\frac{1}{p} = (1-p)^{n-1},$$

   så

   $$P(X \geq n) = 0.5 \quad \Leftrightarrow \quad (1-p)^{n-1} = 0.5 \quad \Leftrightarrow \quad n = 1 + \frac{\ln 0.5}{\ln(1-p)} = 35.31.$$

   Håkan kan alltså äta 34 bär i en följd innan han passerar gränsen.

   Svar: (a) 0.34 (b) ca 0.89 (c) 34.

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

   $$Z = 2 + 2X_1 + 2X_2 \sim N\left(2 + 0 + 0,\ \sqrt{2^2(\sqrt{2})^2 + 2^2(\sqrt{2})^2}\right) = N(2, 4).$$

   Vidare gäller att

   $$P(Z < 1) = P\left(\frac{Z-2}{4} < \frac{1-2}{4}\right) = \Phi\left(-\frac{1}{4}\right) = 1 - \Phi\left(\frac{1}{4}\right) = 0.4013$$

   (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

   $$\begin{aligned} P(\min\{X_1, X_2\} < 0) &= 1 - P(\min\{X_1, X_2\} \geq 0) = 1 - P(X_1 \geq 0)P(X_2 \geq 0) \\ &= 1 - (1 - P(X_1 < 0))(1 - P(X_2 < 0)) = 1 - 1/4 = 3/4. \end{aligned}$$

   (c) Låt $Z = X_1^2 + X_2^2$. Vi vet att $Z < 0$ med sannolikhet 0. Eftersom X_1 och X_2 är oberoende ges den simultana täthetsfunktionen av

   $$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{4\pi}\exp\left(-\frac{x_1^2 + x_2^2}{4}\right), \quad (x_1, x_2) \in \mathbf{R}^2.$$

För $z \geq 0$ gäller att

$$\begin{aligned}
F_Z(z) &= P(Z \leq z) = \iint_{x_1^2+x_2^2 \leq z} f_{X_1,X_2}(x_1,x_2)\,dx_1dx_2 \\
&= \frac{1}{4\pi} \iint_{x_1^2+x_2^2 \leq z} \exp\left(-\frac{x_1^2+x_2^2}{4}\right) dx_1dx_2 = \frac{1}{4\pi}\int_0^{\sqrt{z}}\int_0^{2\pi} r\exp\left(-\frac{r^2}{4}\right) d\theta dr \\
&= \frac{1}{2}\int_0^{\sqrt{z}} r\exp\left(-\frac{r^2}{4}\right) dr = \frac{1}{2}\int_0^{\sqrt{z}} r\exp\left(-\frac{r^2}{4}\right) dr \\
&= \frac{1}{2}\left[-2\exp\left(-\frac{r^2}{4}\right)\right]_0^{\sqrt{z}} = 1 - e^{-z/4},
\end{aligned}$$

vilket är fördelningsfunktionen för en Exp($\lambda = 1/4$)-fördelad variabel. Vi kan även erhålla detta resultat genom att transformera om variablerna till $N(0,1)$-variabler och nyttja att kvadratsumman av oberoende $N(0,1)$-variabler är χ^2-fördelad.

Svar: (a) $N(2,4)$ och 0.40 (b) 3/4 (c) Exp($\lambda = 1/4$).

3. Vi börjar med att klassificera tillstånden. Låt f_{ii}^n vara sannolikheten att för första gången återvända till tillstånd i efter n steg och låt $f_{ii} = \sum_{n=0}^{\infty} f_{ii}^n$. Vi har

$$\begin{aligned}
f_{00} &= \frac{3}{4} + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \cdots = \frac{3}{4} + \frac{1}{8}\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k = \frac{3}{4} + \frac{1}{4} = 1, \\
f_{11} &= \frac{1}{2} + \frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{4} + \cdots = \frac{1}{2} + \frac{1}{8}\sum_{k=0}^{\infty}\left(\frac{3}{4}\right)^k = \frac{1}{2} + \frac{1}{2} = 1, \\
f_{22} &= \frac{1}{3}, \\
f_{33} &= \frac{9}{10}.
\end{aligned}$$

Från detta ser vi att 2 och 3 är obeständiga tillstånd. Både 0 och 1 är beständiga, och eftersom det rör sig om geometriska serier följer det att $\sum_{n=0}^{\infty} n f_{ii}^n < \infty$, så dessa tillstånd är positivt beständiga.

Vidare ser vi om vi räknar ut P^2 (t.ex.) att alla element i kolumn ett är skilda från noll, så en asymptotisk fördelning existerar. Denna erhåller vi genom att hitta alla stationära fördelningar ur ekvationen $\underline{\pi}(P - I) = 0$ med bivillkoret att $\underline{\pi}$ är en stokastisk vektor. Vi ställer upp ekvationssystemet:

$$\begin{cases}
-\frac{1}{4}\pi_0 + \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 = 0 \\
\frac{1}{4}\pi_0 - \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 = 0 \\
-\frac{2}{3}\pi_2 + \frac{1}{10}\pi_3 = 0 \\
-\frac{1}{10}\pi_3 = 0 \\
\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1
\end{cases}
\Leftrightarrow
\begin{cases}
\pi_0 = 2\pi_1 \\
\pi_2 = \pi_3 = 0 \\
2\pi_1 + \pi_1 = 1
\end{cases}
\Leftrightarrow
\begin{cases}
\pi_0 = \frac{2}{3} \\
\pi_1 = \frac{1}{3} \\
\pi_2 = \pi_3 = 0
\end{cases}$$

Eftersom vi bara finner en stationär fördelning är detta den asymptotiska fördelningen.

Svar: Se ovan.

4. Vi kan räkna ut de marginella sannolikhetsfunktionerna och fylla på tabellen:

	$X = 0$	$X = 1$	$X = 2$	$X = 3$	p_Y
$Y = 0$	0.02	0.04	0.01	0.03	0.1
$Y = 1$	0.08	0.16	0.04	b	$0.28 + b$
$Y = 2$	0.1	a	0.05	0.15	$0.3 + a$
p_X	0.2	$0.2 + a$	0.1	$0.18 + b$	

Vi ser direkt att $a+b = 1-0.68 = 0.32$ är nödvändigt för att dessa ska bli sannolikhetsfunktioner.

(a) Om X och Y är oberoende så måste $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ gälla för alla x och y. Genom att välja till exempel $Y = 0$ och $X = 1$ så ser vi att

$$0.04 = (0.2 + a) \cdot 0.1 \quad \Leftrightarrow \quad a = 0.2.$$

Då blir $b = 0.32 - 0.2 = 0.12$. Möjliga utfall för $Z = X + Y$ är $0, 1, 2, 3, 4, 5$. Vi ställer upp sannolikheterna direkt

	p_Z
$Z = 0$	0.02
$Z = 1$	$0.04 + 0.08 = 0.12$
$Z = 2$	$0.01 + 0.1 + 0.16 = 0.27$
$Z = 3$	$0.03 + 0.04 + a = 0.27$
$Z = 4$	$b + 0.05 = 0.17$
$Z = 5$	0.15

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). En snabb kontroll visar att detta summerar till 1.

(b) Kovariansen $C(X, Y)$ kan beräknas enligt $E(XY) - E(X)E(Y)$ så vi kan ta fram dessa storheter först:

$$E(XY) = 1 \cdot 0.16 + 2 \cdot 0.04 + 3 \cdot b + 2 \cdot a + 4 \cdot 0.05 + 6 \cdot 0.15 = 1.34 + 2a + 3b,$$

$$E(X) = 1 \cdot (0.2 + a) + 2 \cdot 0.1 + 3 \cdot (0.18 + b) = 0.94 + a + 3b$$

samt

$$E(Y) = 1 \cdot (0.28 + b) + 2 \cdot (0.3 + a) = 0.88 + 2a + b.$$

Vi erhåller nu att

$$C(X, Y) = 1.34 + 2a + 3b - (0.94 + a + 3b)(0.88 + 2a + b).$$

Vidare måste fortfarande $a + b = 0.32$. Ekvationen $C(X, Y) = 0.05$ reduceras nu till

$$2b^2 - 0.78b + 0.648 = 0.05 \quad \Leftrightarrow \quad b = 0.02 \text{ eller } b = 0.37.$$

Eftersom $a+b = 0.32$ är det andra alternativet omöjligt. Alltså ges de sökta sannolikheterna av $b = 0.02$ och $a = 0.3$. En kontroll visar att dessa gör att vi erhåller en sannolikhetsfunktion och att kovariansen blir 0.05.

Svar: (a) $a = 0.2$ samt $b = 0.12$ (b) Se ovan. (c) $a = 0.3$ samt $b = 0.02$.

5. (a) För att ha en täthetsfunktion måste sannolikhetsmassan bli ett. Således gäller att

$$1 = \int_0^\pi c \sin x \, dx = c\left[-\cos x\right]_0^\pi = 2c \quad \Leftrightarrow \quad c = \frac{1}{2}.$$

Vi använder definitionen för väntevärdet och finner att

$$E(X) = \int_0^\pi x \cdot \frac{1}{2} \sin x \, dx = \cdots = \frac{\pi}{2}.$$

På liknande sätt kan vi se att medianen ges av samma numeriska värde:

$$0.5 = P(X \leq m) = \int_0^m \frac{1}{2} \sin x \, dx = \frac{1}{2}(1 - \cos m) \quad \Leftrightarrow \quad \cos m = 0,$$

så $m = \pi/2$ är den enda giltiga möjligheten.

(b) Vi vill räkna ut

$$E(Y) = E\left(\sum_{k=0}^{N(t)} X_k\right).$$

Eftersom den övre gränsen är en stokastisk variabel för varje t kan vi inte direkt flytta in väntevärdet i summan utan vi får försöka oss på lagen om totalt väntevärde. Alltså,

$$\begin{aligned} E\left(\sum_{k=0}^{N(t)} X_k\right) &= \sum_{n=0}^{\infty} E\left(\sum_{k=0}^{N(t)} X_k \,\middle|\, N(t) = n\right) P(N(t) = n) = \sum_{n=0}^{\infty} E\left(\sum_{k=0}^{n} X_k\right) P(N(t) = n) \\ &= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} E(X_k)\right) P(N(t) = n) = \sum_{n=0}^{\infty} \frac{\pi}{2} \cdot nP(N(t) = n) \\ &= \frac{\pi}{2} \sum_{n=0}^{\infty} nP(N(t) = n) = \frac{\pi}{2} E(N(t)) = \frac{\lambda \pi t}{2}. \end{aligned}$$

Svar: (a) $c = 1/2$ och $E(X) = m = \pi/2$ (b) $\dfrac{\lambda \pi t}{2}$.

6. (a) Låt $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. Vi expanderar kvadraten och observerar att

$$\begin{aligned} E\left((\mu - Y_n)^2\right) &= E\left(\mu^2 - 2\mu Y_n + Y_n^2\right) = \mu^2 - 2\mu E(Y_n) + E(Y_n^2) \\ &= \mu^2 - 2\mu^2 + V(Y_n) + E(Y_n)^2 = V(Y_n), \end{aligned}$$

där vi utnyttjat att väntevärdesoperatorn är linjär, Steiners sats, samt att $E(Y_n) = \mu$. Vidare gäller att

$$V(Y_n) = V\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{\sigma^2}{n}$$

eftersom vi har en sekvens av oberoende likafördelade variabler. Vi kan nu se att

$$E\left((\mu - Y_n)^2\right) = \frac{\sigma^2}{n} \to 0,$$

då $n \to \infty$, vilket var precis vad vi ville visa.

(b) Eftersom $X \geq 0$ så gäller att

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_0^{\infty} x f_X(x)\, dx \geq \int_a^{\infty} x f_X(x)\, dx \\ &\geq a \int_a^{\infty} f_X(x)\, dx = aP(X \geq a) \end{aligned}$$

då täthetsfunktionen givetvis är icke-negativ och $a > 0$. Lite ommöblering visar att

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Svar: (a) se ovan (b) se ovan.