LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (9MA241/9MA341, STN2) 2013-01-11 kl 14–18

Hjälpmedel är: miniräknare med tömda minnen och formelbladet bifogat.
Varje uppgift är värd 6 poäng. För godkänd tentamen räcker 16 poäng. Noggrann motivering krävs där alla viktiga detaljer skall motiveras.

För lösningsskisser, se kurshemsidan `www.mai.liu.se/~jothi/kurser/9MA241-stat/` efter skrivningens slut. Lycka till!

---

1. Vid ett slumpförsök undersöker man tre händelser: A, B och C. Man vet att $P(A) = P(B) = 0.6$ och att $P(A \cap B) = P(A \cap C) = 0.2$ samt $P(B \cap C) = 0.3$. Vidare är $A \cap C$ och $B \cap C$ oberoende.

   (a) Beräkna sannolikheten att alla tre händelserna inträffar. (2p)

   (b) Beräkna sannolikheten att *endast* A inträffar. (2p)

   (c) Vi upprepar slumpförsöket (oberoende) åtta gånger. Vad är sannolikheten att B inträffar fler än sex gånger? (2p)

2. Vid spåranalys av ämnen brukar man först skapa en kalibreringslinje där man använder lösningar med känd koncentration, för att sedan från denna linje finna koncentrationen i okända prover.

   (a) Vid mätningar vid olika koncentrationer fann man följande samband mellan den uppmätta intensiteten I och den kända koncentrationen c (i lämpliga enheter):

c	1	2	3	4	5
I	4.40	8.88	13.22	17.41	21.95

   Använd linjär regression med modellen $I_k = ac_k + b + \epsilon_k$, $k = 1, 2, 3, 4, 5$, där $\epsilon_k \sim N(0, \sigma)$ är oberoende och $a, b \in \mathbf{R}$ konstanter, för att bestämma en linje $I = ac + b$ som minimerar kvadratfelet. (3p)

   (b) För ett okänt prov gjorde man 10 upprepade mätningar och fann medelvärdet $\overline{I} = 10.5900$ samt stickprovsvariansen $s^2 = 0.01412$. Antag att dessa mätningar är observationer från oberoende $N(\mu_I, \sigma)$-variabler. Ange ett 99% konfidensintervall för väntevärdet μ_I.

3. Låt $X \sim \text{Re}(-2, 2)$ och Y ha täthetsfunktionen $f_Y(y) = cy^2$ när $0 < y < 3$ och $f_Y(y) = 0$ för övrigt. Vidare antar vi att X och Y är oberoende.

   (a) Bestäm c så att f_Y blir en täthetsfunktion. (1p)

   (b) Bestäm den simultana täthetsfunktionen $f_{X,Y}(x, y)$ och kovariansen $C(X, Y)$. (2p)

   (c) Vad blir sannolikheten $P(X + Y < 1)$? (3p)

Var god vänd!

4. Låt $X \sim \text{Exp}(5)$ (väntevärde 5). Beräkna medianen för X (ledning: en median är ett tal $m \in \mathbf{R}$ så att $P(X \leq m) = P(X \geq m) = 0.5$) samt visa att, för alla $a, t > 0$, så gäller

$$P(X > x + a \,|\, X > a) = P(X > x),$$

det vill säga, visa att X är "minnesslös." (6p)

5. En dator ska generera binomialfördelade slumptal med fixt $p = 0.4$ (och variabelt n).

   (a) Räkna ut sannolikheten att ett slumptal $X \sim \text{Bin}(250, 0.4)$ är större än 115 eller mindre än 85. Approximationer är OK om dessa motiveras. (2p)

   (b) Den verkliga sannolikheten p är okänd, och för att testa datorn genererar man 250 slumptal från $\text{Bin}(1, p)$ och summerar antalet X ettor, vilket visar sig vara 110 stycken. Antag att dessa 250 slumptal är oberoende av varandra och ställ upp ett approximativt 95% konfidensintervall för den verkliga sannolikheten p. (3p)

   (c) Man gjorde yttligare ett test och genererar 300 slumptal från $\text{Bin}(250, p)$ och finner att 7 stycken är mindre än 85 och 6 stycken är större än 115. Hur stämmer detta överens med resultatet i (a)? (1p)

6. Låt $\Theta \sim \text{Re}(0, \pi/4)$. Bestäm först $E(\cos\Theta)$ och sedan täthetsfunktionen för $Y = \cos\Theta$. Beräkna efter det även $E(Y)$ direkt från definitionen.

Lösningsskisser för matematisk statistik, 9MA241, 2013-01-11

1. (a) $P(A \cap B \cap C) = P((A \cap C) \cap (B \cap C)) = P(A \cap C)P(B \cap C) = 0.2 \cdot 0.3 = 0.06$.

   (b) Vi söker sannolikheten att endast A inträffar. Om man ritar ett Venn-diagram kan man övertyga sig själv om att denna sannolikhet ges av

   $$P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) = 0.6 - 0.2 - 0.2 + 0.06 = 0.26.$$

   När vi drar bort både $P(A \cap B)$ och $P(A \cap C)$ så drar vi bort ”trippelsnittet” två gånger. Detta måste man kompensera för.

   (c) Låt $p = P(B) = 0.6$. Vi upprepar försöket åtta gånger och räknar antalet X gånger som B inträffar. Det följer att $X \sim \text{Bin}(8, p)$ och vi söker $P(X > 6)$:

   $$P(X > 6) = P(X = 7) + P(X = 8) = \binom{8}{7} 0.6^7 \cdot 0.4 + \binom{8}{8} 0.6^8 = 0.1064.$$

Svar: (a) 0.06. (b) 0.26. (c) 0.1064.

2. (a) Vi räknar ut koefficienterna b_0 och b_1 (se formelbladet):

   $$b_0 = 0.08300 \qquad \text{och} \qquad b_1 = 4.363.$$

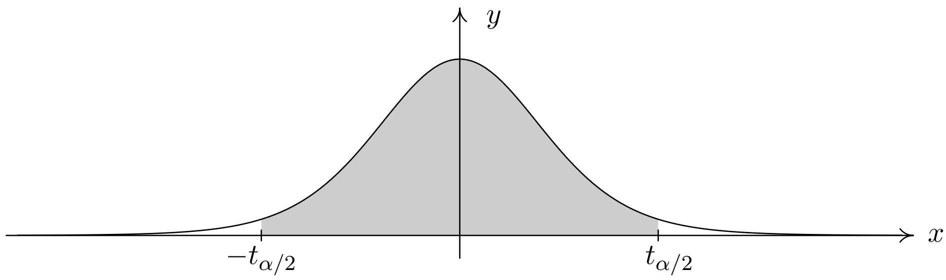
   Vår regressionslinje blir alltså $I = b_0 + b_1 c = 0.083 + 4.363c$.

   (b) Vi har $n = 10$ och bildar testvariabeln

   $$T = \frac{\overline{I} - \mu_I}{S_I/\sqrt{n}} \sim t(9)$$

   och ser att

   $$P(-t_{\alpha/2}(9) \leq T \leq t_{\alpha/2}(9)) = 1 - \alpha.$$



Figur 1: Den markerade arean innehåller 99% av sannolikheten för $t(9)$-fördelningen.

   Vi löser ut μ_I ur intervallet i sannolikhetsmåttet och får att

   $$\overline{I} - \frac{S_x}{\sqrt{n}} t_{\alpha/2}(9) \leq \mu_I \leq \overline{I} + \frac{S_x}{\sqrt{n}} t_{\alpha/2}(9).$$

   Vi ersätter $\overline{I}$ med den observerade punktskattningen $\overline{I}^*_{\text{obs}} = 10.59$ och S_x med $s_x = \sqrt{0.0142}$. Då erhåller vi ett konfidensintervall med konfidensgrad 99%:

   $$I_\mu = [10.47, 10.71],$$

   där vi använt $\alpha = 0.01$ och $t_{0.005}(9) = 3.2498$.

Svar: (a) $I = 0.083 + 4.363c$. (b) $I_\mu = [10.47, 10.71]$.

3. (a) Vi bestämmer c:

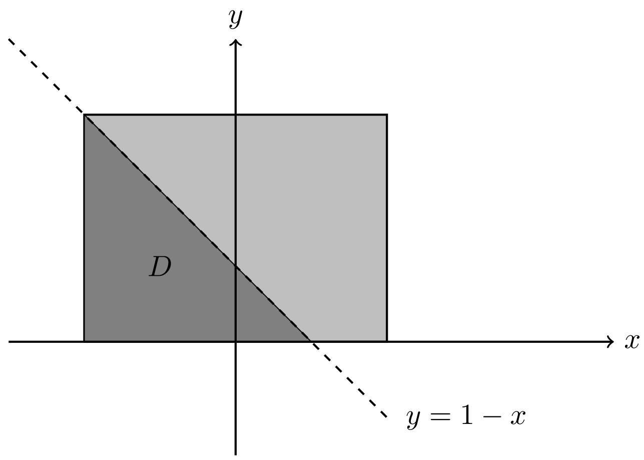
$$1 = \int_0^3 cy^2\, dy = c\left[\frac{y^3}{3}\right]_0^3 = 9c \quad \Rightarrow \quad c = \frac{1}{9}.$$

(b) Den simultana täthetsfunktionen ges av $f_X(x)f_Y(y)$ eftersom variablerna är oberoende, så

$$f_{X,Y}(x,y) = \begin{cases} \frac{y^2}{36}, & -2 \le x \le 2,\ 0 < y < 3, \\ 0, & \text{för övrigt.} \end{cases}$$

Eftersom variablerna är oberoende så blir $C(X,Y) = 0$!

(c) Vi skisserar området som (X,Y) är definierad på.



Figur 2: Den ljust skuggade rektangeln är hela mängden där $f_{X,Y} \neq 0$, och den mörkare skuggade triangeln är delområdet D där $x + y < 1$.

$$P(X + Y < 1) = \int_{-2}^{1}\int_0^{1-x} \frac{y^2}{36}\, dydx = \cdots = 3/16.$$

Svar: (a) $c = 1/9$ (b) Se ovan. (c) $3/16$.

4. Vi räknar ut fördelningsfunktionen för X. Om $x > 0$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}.$$

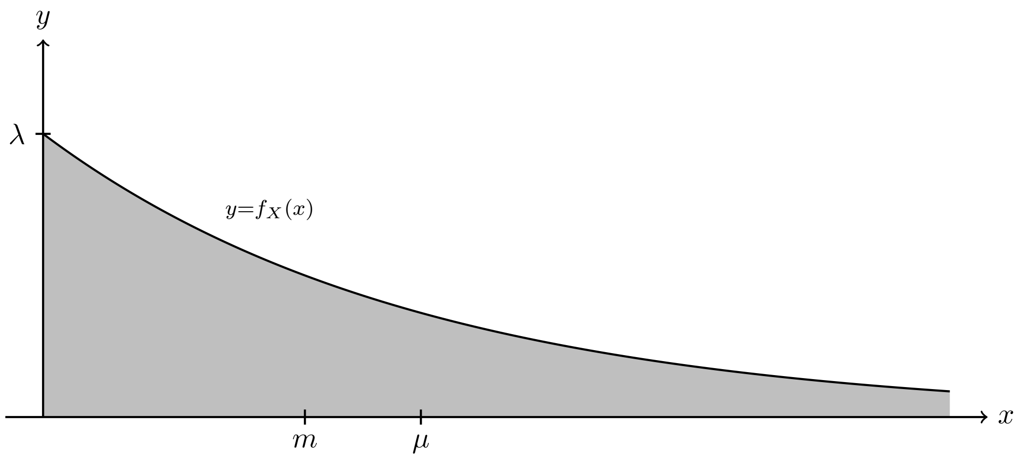
Medianen finner vi ur ekvationen $F_X(m) = 1/2$, dvs

$$1 - e^{-m/5} = \frac{1}{2} \quad \Leftrightarrow \quad e^{-m/5} = \frac{1}{2} \quad \Leftrightarrow \quad m = \frac{\ln 2}{1/5} \approx 3.47.$$

Jämför gärna detta med väntevärdet för X:

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = 0 - 0 + \left[-\frac{e^{-\lambda x}}{\lambda}\right]_0^\infty = \frac{1}{\lambda} = 5.$$

Medianen och väntevärdet behöver alltså *inte* vara samma sak!



Vi visar nu att Exponentialfördelningen är minnesslös: vi använder definitionen av betingad sannolikhet och erhåller

$$\begin{aligned} P(X > x + a \,|\, X > a) &= \frac{P(\{X > x+a\} \cap \{X > a\})}{P(X > a)} = \frac{P(X > x+a)}{P(X > a)} \\ &= \frac{\frac{1}{5}\int_{x+a}^{\infty} e^{-t/5}\,dt}{\frac{1}{5}\int_{a}^{\infty} e^{-t/5}\,dt} = \frac{e^{-(x+a)/5}}{e^{-a/5}} = e^{-x/5}. \end{aligned}$$

Observera att denna sannolikhet är samma som

$$P(X > x) = \frac{1}{100}\int_x^{\infty} e^{-t/100}\,dt = e^{-x/5}.$$

Svar: $m = 3.47$.

5. (a) Låt $X \sim \mathrm{Bin}(250, 0.4)$. Eftersom $250 \cdot 0.4 \cdot 0.6 = 60 \gg 10$ så är $X \overset{\text{appr.}}{\sim} \mathrm{N}(100, \sqrt{60})$. Då blir

$$\begin{aligned} P(\{X > 115\} \cup \{X < 85\}) &= P(X > 115) + P(X < 85) = 1 - P(X \leq 115) + P(X \leq 84) \\ &\approx 1 - \Phi((115-100)/\sqrt{60}) + \Phi((84-100)/\sqrt{60}) \\ &= 2 - \Phi(1.94) - \Phi(2.07) = 0.045. \end{aligned}$$

(b) Vi söker konfidensintervall för p. Om man summerar n oberoende $\mathrm{Bin}(1,p)$ variabler så erhåller vi en $\mathrm{Bin}(n,p)$ variabel. Med $n = 250$ och observerade antalet $X = 110$ ges en naturlig skattning på den verkliga andelen av

$$\widehat{P} = \frac{X}{n},$$

och vårt observerade värde är $\widehat{p} = 110/250 = 0.44$. Vi vet att variabeln X är binomialfördelad: $X \sim \mathrm{Bin}(n,p)$. Vidare ser vi att

$$np(1-p) \approx 250 \cdot 0.4 \cdot 0.6 = 60,$$

så vi borde kunna göra en normalapproximation: $X \overset{\text{appr.}}{\sim} N(np, D)$ där $D = \sqrt{np(1-p)}$. Alltså blir $\widehat{P} \overset{\text{appr.}}{\sim} N(p, d)$, där

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{0.44 \cdot 0.56/250} = 0.0314.$$

Vår testvariabel blir alltså

$$Z = \frac{\widehat{P} - p}{d} \overset{\text{appr.}}{\sim} N(0,1).$$

Vi stänger in Z:

$$P(-\lambda_{\alpha/2} \leq Z \leq \lambda_{\alpha/2}) \approx 1-\alpha,$$

där λ_α är normalfördelningens α-kvantil. Vi har $\alpha = 0.05$ och $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Vi löser ut p ur olikheten inuti sannolikhetsmåttet, och erhåller då

$$\widehat{P} - \lambda_{0.025} d \leq p \leq \widehat{P} + \lambda_{0.025} d.$$

Detta ger oss konfidensintervallet $I_p = [0.38, 0.50]$ om vi ersätter $\widehat{P}$ med det observerade värdet $\widehat{p} = 0.44$.

(c) Vi hittar alltså $7 + 6 = 13$ stycken utanför intervallet $[85, 115]$. Vi vet att sannolikheten att hamna utanför detta intervall (om fördelningen stämmer) är 0.045. Vi testar $13/300 = 0.043$, vilket ligger ganska nära. Mer precist än så kan vi inte säga utan att göra ett konfidensintervall eller hypotestest.

Svar: (a) 0.045 (b) $I_p = [0.38, 0.50]$. (c) Se ovan.

6. Det enklaste sättet är att använda satsen för $E(g(X))$:

$$E(Y) = E(\cos\Theta) = \int_{-\infty}^{\infty} \cos\theta f_\Theta(\theta)\,d\theta = \frac{4}{\pi}\int_0^{\pi/4} \cos\theta\,d\theta = \frac{4}{\pi}\left(\frac{\sqrt{2}}{2} - 0\right) = \frac{2\sqrt{2}}{\pi}.$$

Den andra varianten börjar med beräkning av täthetsfunktionen för $Y = \cos\Theta$. Vi ställer upp fördelningsfunktionen:

$$F_Y(y) = P(Y \leq y) = P(\cos\Theta \leq y) = P(\Theta \geq \arccos y) = 1 - P(\Theta < \arccos y) = 1 - F_\Theta(\arccos y).$$

Här har vi uttnyttjat att $\cos\theta$ är avtagande för $\theta \in [0, \pi/4]$. Vi kan nu derivera fram $f_Y(y)$:

$$\begin{aligned} f_Y(y) = F_Y'(y) &= -F_\Theta'(\arccos y) \cdot \frac{-1}{\sqrt{1-y^2}} = \frac{f_\Theta(\arccos y)}{\sqrt{1-y^2}} \\ &= \begin{cases} \frac{4}{\pi}\frac{1}{\sqrt{1-y^2}}, & 0 \leq \arccos y \leq \frac{\pi}{4} \quad \Leftrightarrow \quad \frac{\sqrt{2}}{2} \leq y \leq 1 \\ 0, & \text{för övrigt.} \end{cases} \end{aligned}$$

Vi kan nu beräkna $E(Y)$ enligt definitionen:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \frac{4}{\pi}\int_0^{\sqrt{2}/2} \frac{y}{\sqrt{1-y^2}}\,dy = \frac{4}{\pi}\left[-\sqrt{1-y^2}\right]_0^{\frac{\sqrt{2}}{2}} = \frac{2\sqrt{2}}{\pi}.$$

Svar: Se ovan.