LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (9MA241, STN2)
2013-08-19 kl 08–12

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. Motivera svaren på följande frågor noggrant!

   (a) Låt $f(x) = 2x$ för $0 \leq x < 1$ och $f(x) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (b) Låt $f(x) = \sin x$ för $0 < x < 2\pi$ och $f(x) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (c) Låt $f(x) = 0.25 \sin x$ för $0 < x < 2\pi$ och $f(x) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (d) Låt A och B vara händelser så att $P(A \cap B) = 0.2$ och $P(A^* \cap B^*) = 0.5$. Visa att A och B ej kan vara oberoende samt beräkna sannolikheten att precis en av A och B inträffar. (3p)

2. I ett försök med en ny medicin på fem testpersoner erhölls följande mätdata i lämplig enhet före (x_i) och efter (y_i) behandling:

x_i	11.80	11.88	13.49	13.41	13.42
y_i	8.69	9.57	10.34	13.58	12.77

   Mätningarna är observationer av stokastiska variabler $X_i \sim \mathrm{N}(\mu_i, \sigma)$ och $Y_i \sim \mathrm{N}(\mu_i + \Delta, \sigma)$ där olika personer antas oberoende av varandra.

   (a) Ange en punktskattning för Δ. (2p)

   (b) Finn ett konfidensintervall med konfidensgrad 90% för Δ. Ange en lämplig signifikansnivå och svara på frågan om man kan man förkasta hypotesen att inte är någon skillnad på värden före och efter behandling. (4p)

3. I en stor populationsgrupp så vet man att 4% har en viss gensammansättning som gör att man har högre risk att utveckla en allvarlig sjukdom. Approximationer är OK om dessa motiveras.

   (a) Om man skulle undersöka 3000 personer, vad är sannolikheten att fler än 110 men färre än 125 har denna genmsammansättning? (3p)

   (b) När man testade 1000 personer så hade 50 stycken denna gensammansättning. Ange ett 99% konfidensintervall för andelen som har denna gensammansättning i den större populationen. Är 4% rimligt? (3p)

Vänd!

4. Betrakta en samling vanliga kortlekar (52 kort, 4 färger). Antag att vi har en gigantisk container fylld med sådana kortlekar välblandade tillsammans (vi tänker oss här att det finns oändligt många kort).

   (a) Om man utan att titta tar kort efter kort ur containern på måfå till dess att man lyckas få en spader för första gången, vad är sannolikheten att detta tar fler än 5 försök (inklusive det första spader-kortet man finner)? (3p)

   (b) Om man på måfå drar ett kort, vad är den förväntade valören (siffervärdet) på kortet? (3p)

5. Låt den två-dimensionella stokastiska variabeln (X, Y) vara likformigt fördelad på området i planet där $-2 \leq x \leq 2$ och $0 \leq y < |x|$.

   (a) Bestäm de marginella täthetsfunktionerna $f_X(x)$ och $f_Y(y)$. Är X och Y oberoende? (3p)

   (b) Beräkna $P(X < 1, Y > 1)$. (3p)

6. Låt $f(x) = (a+bx)e^{-2x}$ för $x \geq 0$. Kan man finna reella tal a och b så att f blir en täthetsfunktion för en stokastisk variabel med väntevärde 4? Finn talen a och b eller motbevisa existens. (6p)

Lösningsskisser för matematisk statistik, 9MA241, 2013-08-19

1. (a) Ja. Arean är ett och funktionen är ≥ 0.

   (b) Nej. Funktionen är < 0 bitvis.

   (c) Nej. Funktionen är < 0 bitvis.

   (d) Ur ett Venndiagram finner vi att $P(A^* \cap B^*) = P((A \cup B)^*)$. Alltså är $P(A \cup B) = 1 - 0.5 = 0.5$, och den sökta sannolikheten blir således

   $$P(A \cup B) - P(A \cap B) = 0.5 - 0.2 = 0.3.$$

   Låt $a = P(A)$ och $b = P(B)$. Antag att A och B är oberoende. Då gäller att $P(A \cap B) = ab = 0.2$ och $P(A^* \cap B^*) = (1-a)(1-b) = 0.5$. Alltså blir

   $$ab = 0.2 \quad \Leftrightarrow \quad a^2 - 0.7a + 0.2 = 0,$$

   vilket saknar reella lösningar. Händelserna kan alltså inte vara oberoende.

   Svar: (a) Ja. (b) Nej. (c) Nej. (d) 0.3, beroende händelser.

2. Modellen är stickprov i par. Vi betraktar skillnaden som ett enda stickprov:

   $$z_i = x_i - y_i \quad 3.11 \quad 2.31 \quad 3.15 \quad -0.17 \quad 0.65$$

   och skattar väntevärdet med

   $$\widehat{\delta} = \frac{1}{5}\sum_{i=1}^{5}(x_i - y_i) = 1.81$$

   och variansen med

   $$s = \left(\frac{1}{5-1}\sum_{i=1}^{5}(x_i - y_i - \bar{\Delta})^2\right)^{1/2} = 1.50.$$

   (a) En bra skattning av Δ är medelvärdet 1.81.

   (b) Låt

   $$\widehat{\Delta} = \frac{1}{5}\sum_{i=1}^{5}(X_i - Y_i).$$

   Vi bildar testvariabeln

   $$T = \frac{\widehat{\Delta} - \Delta}{S/\sqrt{n}} \sim t(4)$$

   där vi skattar S med s, så $s/\sqrt{n} = 0.6709$ och

   $$P(-t_{\alpha/2}(4) < T < t_{\alpha/2}(4)) = 1 - \alpha.$$

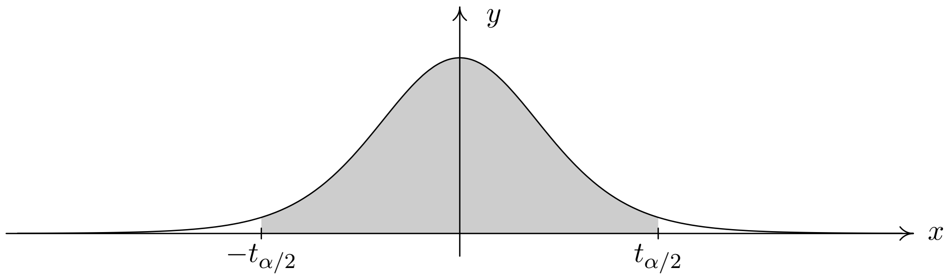
   Vi löser ut Δ ur intervallet i sannolikhetsmåttet och får att

   $$\widehat{\Delta} - 0.6709 t_{\alpha/2}(4) < \Delta < \widehat{\Delta} + 0.6709 t_{\alpha/2}(4).$$

   Vi ersätter $\widehat{\Delta}$ med den observerade punktskattningen $\widehat{\delta} = 1.81$ och erhåller ett konfidensintervall av konfidensgrad 90%:

   $$I_\Delta = [0.38, 3.24],$$

   där vi använt $\alpha = 0.10$ och $t_{0.05}(4) = 2.132$. Eftersom $0 \notin I_\Delta$ kan vi förkasta hypotesen (med ett dubbelsidigt test). Hade ett enkelsidigt varit smartare?



Figur 1: Den markerade arean innehåller 90% av sannolikheten för $t(4)$-fördelningen.

Svar: (a) T.ex. 1.81 (eller $e^{\sqrt{\pi}}$?). (b) $[0.38, 3.24]$; Hypotesen kan förkastas på 10%-nivån.

3. (a) Vi ser här att $X \sim \mathrm{Bin}(n, p)$. Med $p = 0.04$ och $n = 3000$ så är $np(1-p) = 115.2$ betydligt större än 10. Normalapproximation är OK. Således erhåller vi

$$\begin{aligned} P(110 < X < 125) &= P(111 \le X \le 124) \\ &\approx \Phi((124-120)/\sqrt{115.2}) - \Phi((111-120)/\sqrt{115.2}) = 0.44. \end{aligned}$$

Verklig sannolikhet (direkt via binomialfördelningen) är 0.451.

(b) Vi söker konfidensintervall för andelen som har denna gensammansättning. Totalt sett frågat $n = 1000$ stycken personer, och uppmätt är $X = 50$. En naturlig skattning på den verkliga andelen ges av

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

och vårt observerade värde är $\widehat{p} = 50/1000 = 0.05$. Fördelningen för X ges av $X \sim \mathrm{Bin}(n, p)$. Vidare ser vi att

$$np(1-p) \approx 1000 \cdot 0.05 \cdot 0.95 = 47.5,$$

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

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{47.5/1000^2} = 0.0069.$$

Vår testvariabel blir alltså

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

Vi stänger in Z:

$$P(-\lambda_{\alpha/2} \le Z \le \lambda_{\alpha/2}) = 1 - \alpha,$$

där λ_α är normalfördelningens α-kvantil. Vi har $\alpha = 0.01$ och $\lambda_{\alpha/2} = \lambda_{0.005} = 2.576$. Vi löser ut p ur olikheten inuti sannolikhetsmåttet, och erhåller då

$$\widehat{P} - \lambda_{0.005} d \le p \le \widehat{P} + \lambda_{0.005} d.$$

Detta ger oss konfidensintervallet $I_p = [0.0322, 0.0678]$ om vi ersätter $\widehat{P}$ med det observerade värdet $\widehat{p} = 0.05$.

Svar: (a) Ca 44%. (b) $[0.0322, 0.0678]$. Andelen $p = 0.04$ förefaller inte orimlig.

4. (a) Sannolikheten att dra ett spaderkort är konstant $p = 0.25$ (med oändligt många kort, annars ändras sannolikheten hela tiden). Låt X vara antalet dragningar till och med vi drar spader första gången. Då är $X \sim \mathrm{Ffg}(0.25)$ och

$$P(X \ge 6) = \sum_{k=6}^{\infty} 0.25 \cdot 0.75^{k-1} = 0.25 \cdot 0.75^5 \sum_{k=0}^{\infty} 0.75^k = 0.237,$$

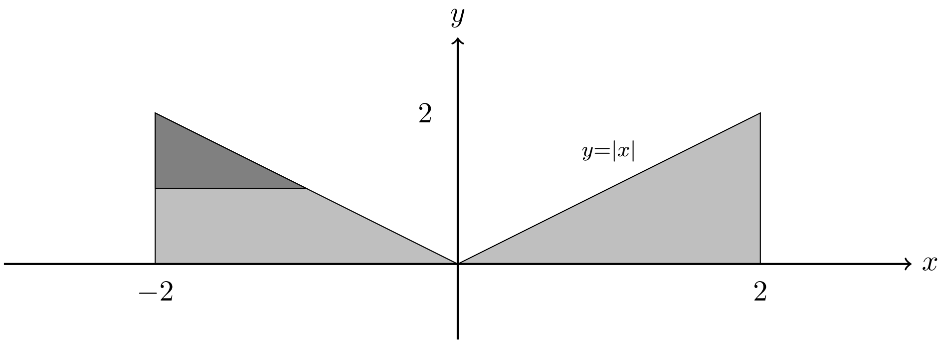
enligt formel för geometrisk serie. Ogillar man serier går det utmärkt att betrakta komplementhändelsen istället, blir lite bökigare men ger samma svar.

(b) Låt Y vara valören på ett kort som dras på måfå ur containern. Då kan Y anta de olika värderna $1, 2, \ldots, 13$ och sannolikheten är konstant lika med $1/13$. Vi söker $E(Y)$, och kan räkna ut detta med hjälp av definitionen:

$$E(Y) = \sum_{k=1}^{13} kP(Y = k) = \frac{1}{13}\sum_{k=1}^{13} k = \frac{13(1+13)}{2 \cdot 13} = 7.$$

Svar: (a) 0.237. (b) Väntevärdet är 7.

5. Vi skissar det intressanta området:



Figur 2: Det ljust skuggade området är där $f_{X,Y}$ inte är noll, och det mörkare den delmängd där $x < 1$ och $y > 1$.

Eftersom området har area 4 så kommer den simultana täthetsfunktionen att ha värdet $f_{X,Y}(x, y) = 1/4$ för (x, y) i det skuggade området (och lika med noll för övrigt).

(a) Vi räknar ut de marginella täthetsfunktionerna:

$$f_X(x) = \int_0^{|x|} \frac{1}{4}\,dy = |x|/4, \quad -2 \le x \le 2,$$

och

$$f_Y(y) = \int_{-2}^{-y} \frac{1}{4}\,dx + \int_y^2 \frac{1}{4}\,dx = \frac{2-y}{2}, \quad 0 \le y \le 2.$$

Vi ser direkt att $f_X(x)f_Y(y) \neq f_{X,Y}(x, y)$ för de flesta punkter, så X och Y är beroende.

(b) Den del av det skuggade området där $X < 1$ och $Y > 1$ har area $1/2$, så eftersom vi har en likformig fördelning blir det sökta sannolikheten

$$P(X < 1,\ Y > 1) = \frac{1/2}{4} = \frac{1}{8}.$$

Svar: (a) Se ovan. Variablerna är beroende. (b) 1/8.

6. Kraven i uppgiften ger (efter partialintegration) att

$$4 = E(X) = \int_0^{\infty} x \cdot (a + bx)e^{-2x}\,dx = \cdots = \frac{a+b}{4}$$

och

$$1 = \int_0^\infty (a + bx)e^{-2x}\,dx = \cdots = \frac{a}{2} + \frac{b}{4}.$$

Dessa ekvationer medför att $a = -12$ och att $b = 28$. Så fungerar detta? Enligt konstruktion så måste integralkraven ovan vara uppfyllda, men blir f en täthetsfunktion? Vad gäller kravet att $f \geq 0$? Med $a = -12$ ser vi att när x är nära noll så blir f negativ! Alltså är dessa krav omöjliga!

Svar: Går inte.