LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (9MA241, STN2)
2013-05-28 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. (a) I ett slumpförsök betraktar vi två händelser A och B där $P(A) = 0.3$ och $P(B \mid A) = 0.4$. Beräkna $P(A \cap B)$. (2p)

   (b) Vi upprepar försöket sju gånger. Vad är sannolikheten att både A och B inträffar i högst tre av försöken? (2p)

   (c) Om vi istället upprepar försöket tills dess att A och B inträffar samtidigt för första gången, vad är sannolikheten att detta tar högst 2 upprepningar? (2p)

2. I en telefonundersökning tillfrågas personer om det är lämpligt att ha tropiska giftormar som husdjur i lägenhet. Personerna i undersökningen har valts ut slumpmässigt.

   (a) Vid tillfället frågade man 360 personer, och av dessa tyckte 18 stycken att det var okej. Hitta ett 95% konfidensintervall för andelen av befolkningen som tycker det är okej. (4p)

   (b) Låt $[0.039, 0.062]$ vara ett 99% konfidensintervall för andelen vid en liknande undersökning som tyckt det är okej med giftormar i sin lägenhet. Ange en lämplig punktskattning för den verkliga andelen av befolkningen som tycker det är okej. Motivera ditt svar! (2p)

3. Låt $Y \sim \mathrm{N}(4, 2)$ och definiera den stokastiska variabeln X enligt

$$X = \begin{cases} -1, & Y < 0, \\ 0, & 0 \leq Y < 3, \\ 1, & 3 \leq Y < 5, \\ 2, & 5 \leq Y < 6, \\ 3, & Y \geq 6. \end{cases}$$

   (a) Bestäm sannolikhetsfunktionen p_X för X. (4p)

   (b) Finn $E(X)$ och $V(X)$. (2p)

Vänd!

4. Låt A vara (den fyllda) triangeln med hörn i $(0,0)$, $(0,2)$ och $(3,0)$. Vi definierar en tvådimensionell täthetsfunktion enligt

$$f_{X,Y}(x,y) = \begin{cases} 2y(3-x)/9, & \text{punkten } (x,y) \text{ ligger i } A, \\ 0, & \text{för övrigt.} \end{cases}$$

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

(b) Beräkna $P(X < 1, Y > 4/3)$. (3p)

5. Hobbyherpetologen Hasse föder upp Gaboonhuggormar (*Bitis Gabonica*) i en stuga ute i Värmlandsskogen. Hasse tycker ormarna ser lite klena ut och ger dem därför ett tillväxthormon han köpt av en mindre noggräknad muskelbyggare. Nu är det oklart om hormonet verkar på ormarna så Hasse vill undersöka närmare. Han delar in sina 18 ormar i två grupper: grupp A med 10 ormar och grupp B med 8. Grupp A ges tillväxthormon. Hasse mäter vikten på alla sina ormar (innan matning) direkt innan försöket börjar, och sedan en månad senare (igen innan matning). Låt X_i och Y_i vara vikterna i grupp A och B innan försöket, och låt Z_i och W_i vara vikterna i grupp A och B efter en månad. Vi antar att variablerna är normalfördelade och att olika ormar är oberoende. Mätdata (i kg) med ormarna i samma ordning inom sina respektive grupper:

x_i	5.62	6.19	5.89	6.56	5.46	6.02	6.28	6.55	6.77	6.04
y_i	6.69	5.14	5.95	5.88	6.16	6.16	5.57	5.98	-	-
z_i	6.92	6.38	6.61	5.92	5.93	6.55	6.86	7.79	6.17	6.59
w_i	6.98	5.76	6.78	6.36	7.02	5.32	6.20	5.70	-	-

Följande siffror har Hasse sedan räknat ut:

$$\sum x_i = 61.38 \qquad \sum z_i = 65.72 \qquad \sum y_i = 47.53 \qquad \sum w_i = 50.12$$

$$\sum x_i^2 = 378.3336 \qquad \sum z_i^2 = 434.6374 \qquad \sum y_i^2 = 283.8291 \qquad \sum w_i^2 = 316.8288$$

$$\sum (x_i - z_i)^2 = 5.6924 \qquad \sum (y_i - w_i)^2 = 3.3083$$

(a) Ställ upp ett 90% konfidensintervall för skillnaden mellan väntevärderna för grupp A och grupp B efter behandlingen. (3p)

(b) Kan man säga att ormarna i grupp A blev tyngre efter behandlingen jämfört med innan? Använd ett 95% konfidensintervall för att besvara frågan. (3p)

6. Låt $X \sim \text{Exp}(\mu)$ och definiera $Y = [X]$ (Y är heltalsdelen av X, e.g, om $X = 3.7$ så är $Y = 3$). Finn sannolikhetsfunktionen för Y och beräkna $E(Y)$. (6p)

Lösningsskisser för matematisk statistik, 9MA241, 2013-05-28

1. (a) Eftersom $0.4 = P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$, så följer det att $P(A \cap B) = 0.12$.

   (b) Sannolikheten p att både A och B inträffar i ETT försök är $p = 0.12$ enligt föregående deluppgift. Vi upprepar 7 gånger och räknar antalet X gånger både A och B inträffar. Det följer (om vi antar oberoende) att $X \sim \text{Bin}(7, p = 0.12)$. Sålunda,

$$P(X \leq 3) = 0.88^7 + 7 \cdot 0.88^6 \cdot 0.12 + \binom{7}{2} 0.88^5 \cdot 0.12^2 + \binom{7}{3} 0.88^4 \cdot 0.12^3 = 0.9946$$

   (c) Med samma p som ovan undersöker vi nu variabeln Y som räknar antalet gånger vi upprepar försöket tills dess att både A och B inträffar för första gången. Alltså, $Y \sim \text{Ffg}(p = 0.12)$. Vi erhåller nu att

$$P(Y \leq 2) = 0.12 + 0.88 \cdot 0.12 = 0.2256$$

   Svar: (a) 0.12. (b) 0.995. (c) 0.2256.

2. (a) Vi söker konfidensintervall för andelen som svarat att det är OK. Totalt sett $n = 360$ stycken personer, och fått $X = 18$ stycken OK. En naturlig skattning på den verkliga andelen ges av

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

   och vårt observerade värde är $\widehat{p} = 18/360 = 0.05$. Fördelningen för X ges av $X \sim \text{Bin}(n, p)$. Vidare ser vi att

$$np(1-p) \approx 360 \cdot 0.05 \cdot 0.95 = 17.1,$$

   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

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{17.1/360^2} = 0.0115.$$

   Vår testvariabel blir alltså

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

   Vi stänger in Z:

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

   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 $I_p = [0.0275, 0.0725]$ om vi ersätter $\widehat{P}$ med det observerade värdet $\widehat{p} = 0.05$.

   (b) Egentligen är vilken siffra vi än hittar på (som är en sannolikhet) en punktskattning, men om dessa är lämpliga är mer tveksamt. Konfidensintervall brukar vara symmetriska (om de är dubbelsidiga, vilket vi antar här), så en vettig punkskattning på andelen p är talet i mitten av intervallet. Vi skattar alltså med $\widehat{p} = (0.039 + 0.062)/2 = 0.0505$.

   Svar: (a) $I_p = [0.0275, 0.0725]$ (b) T.ex. $\widehat{p} = 0.0505$.

3. (a) Vi behöver sannolikheterna att hamna i de olika fallen:

$$
\begin{aligned}
P(X=-1) &= P(Y<0) = \Phi((0-4)/2) = 1-\Phi(2) = 0.0228 \\
P(X=0) &= P(0 \le Y < 3) = \Phi((3-4)/2) - \Phi((0-4)/2) = 0.2858 \\
P(X=1) &= P(3 \le Y < 5) = \Phi((5-4)/2) - \Phi((3-4)/2) = 0.3829 \\
P(X=2) &= P(5 \le Y < 6) = \Phi((6-4)/2) - \Phi((5-4)/2) = 0.1499 \\
P(X=3) &= P(Y \ge 6) = 1-\Phi((6-4)/2) = 0.1586
\end{aligned}
$$

Kontrollera att sannolikheterna summerar till ett! Eftersom $p_X(k) = P(X=k)$ är vi klara med denna deluppgift.

(b) Vi använder definitionen och finner att

$$E(X) = \sum_{k=-1}^{3} k p_X(k) = 1.1357.$$

Vidare kan vi räkna ut att

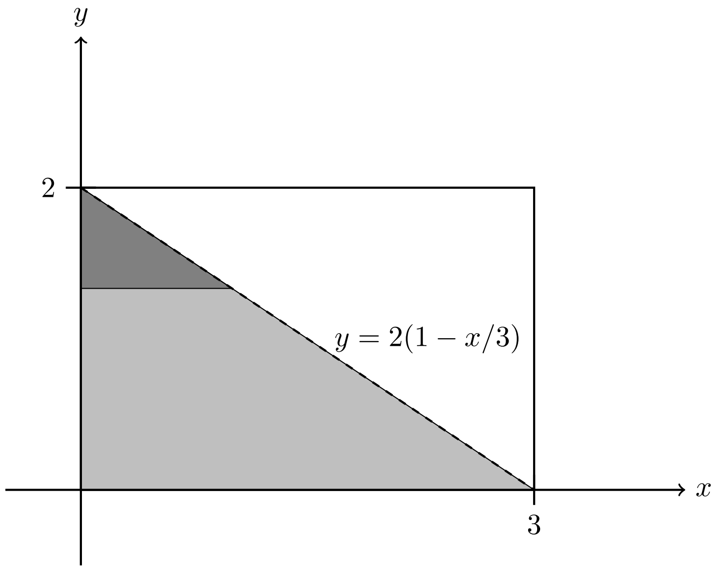
$$E(X^2) = \sum_{k=-1}^{3} k^2 p_X(k) = 2.4327.$$

Steiners sats implicerar nu att

$$V(X) = E(X^2) - E(X)^2 = 2.4327 - 1.2898 = 1.1429$$

Svar: (a) Se ovan. (b) $E(X) = 1.1357$ och $V(X) = 1.1429$.

4. Vi skisserar de intressanta områderna:



Figur 1: Den ljust skuggade triangeln är området där $f_{X,Y}(x,y) \neq 0$, och den mörkt skuggade triangeln är delmängden där $X < 1$ och $Y > 4/3$. Den inritade rektangeln är området där $f_X(x)f_Y(y) \neq 0$.

(a) Vi räknar ut de marginella täthetsfunktionerna:

$$f_X(x) = \int_0^{2(1-x/3)} \frac{2y(3-x)}{9}\,dy = \cdots = \frac{4(3-x)^3}{81}, \quad 0 \le x \le 3,$$

och

$$f_Y(y) = \int_0^{3(2-y)/2} \frac{2y(3-x)}{9}\,dx = \cdots = y(1-y^2/4), \quad 0 \le y \le 2.$$

Vi ser direkt att $f_X(x)f_Y(y) \neq f_{X,Y}(x,y)$ för åtminstone många punkter utanför triangeln där $f_{X,Y}(x,y) \neq 0$, så X och Y är beroende.

(b) Sannolikheten räknar vi direkt ut ur definitionen:

$$P(X < 1,\ Y > 4/3) = \int_0^1 \int_{4/3}^{2(1-x/3)} \frac{2y(3-x)}{9}\,dydx = \cdots \frac{25}{81}.$$

Eftersom punkten $(1, 4/3)$ ligger precis på linjen som avgränsar området behöver vi inte dela upp i fall.

Svar: (a) Se ovan. Variablerna är beroende. (b) 25/81.

5. (a) Vi saknar uppgift om varianser, så dessa måste skattas. Vi vill jämföra två stickprov (mätningarna efter behandlingen av grupp A) av storlek $n_1 = 10$ och $n_2 = 8$. Vi punktskattar $\mu = \mu_1 - \mu_2$ med $\mu^* = \bar{Z} - \bar{W}$. Lämplig testvariabel blir

$$T = \frac{\mu^* - (\mu_1 - \mu_2)}{S\sqrt{1/10 + 1/8}},$$

där S^2 skattas med

$$s^2 = \frac{9s_1^2 + 7s_2^2}{16}$$

som är den sammanvägda skattningen av σ^2. Det följer att $T \sim t(16)$. Vi räknar ut s_1 och s_2 med hjälp av siffrorna vi fått:

$$s_1 = 0.5503 \quad \text{och} \quad s_2 = 0.6355.$$

Nu fullföljer vi som vanligt, med $s^2 = 0.3470$, $\mu^*_{\text{obs}} = 0.3070$, $s\sqrt{1/10 + 1/8} = 0.2794$, och $t_{0.05}(16) = 1.746$, och får konfidensintervallet $I_{\mu_1-\mu_2} = [-0.18, 0.80]$. Eftersom nollan ingår i intervallet kan vi inte med säkerhet säga att det är någon skillnad mellan grupperna efter en månad.

(b) Modellen är stickprov i par. Vi betraktar skillnaden som ett enda stickprov och skattar väntevärdet med

$$\bar{\Delta} = \sum_{i=1}^{8} x_i - z_i = -0.4340$$

och variansen med

$$s = \left(\frac{1}{10-1} \sum_{i=1}^{10} (x_i - z_i - \bar{\Delta})^2 \right)^{1/2} = 0.6505.$$

Detta kan räknas ut med formeln

$$\sum_{i=1}^{n} (u_i - \overline{u})^2 = \sum_{i=1}^{n} u_i^2 - 2n\overline{u}^2 + n\overline{u}^2 = \sum_{i=1}^{n} u_i^2 - n\overline{u}^2,$$

där $u_i = x_i - z_i$ och $n = 10$. Vi bildar testvariabeln

$$T = \frac{\Delta^* - \Delta}{S/\sqrt{n}} \sim t(9)$$

där vi skattar S med s, så $s/\sqrt{n} = 0.2057$ och

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

Vi löser ut Δ ur intervallet i sannolikhetsmåttet och får att

$$\Delta^* - 2057t_{\alpha/2}(9) < \Delta < \Delta^* + 0.2057t_{\alpha/2}(9).$$

Vi ersätter Δ^* med den observerade punktskattningen $\Delta^*_{\text{obs}} = -0.4340$. Vi erhåller ett konfidensintervall av konfidensgrad 95%:

$$I_\Delta = [-0.90, 0.03],$$

där vi använt $\alpha = 0.05$ och $t_{0.025}(9) = 2.262$. Eftersom $0 \in I_\Delta$ kan vi inte förkasta hypotesen att inget hänt.

Svar: (a) $[-0.18, 0.80]$
(b) K.I. för stickprov i par: $[-0.90, 0.03]$, kan ej styrka skillnad före och efter.

6. Det är klart att $P(Y = k) = 0$ för $k < 0$ (varför?). Antag att $k \geq 0$ (icke-negativt heltal). Vi ser att $Y = k$ precis då $[X] = k$, vilket innebär att $k \leq X < k + 1$ (vi klipper alltså bara bort alla decimaler på talet X). Så vad är sannolikheten att $k \leq X < k + 1$? Den kan vi räkna ut då vi vet att $f_X(x) = \mu^{-1}\exp(-x/\mu)$ för $x \geq 0$. Alltså,

$$\begin{aligned} P(Y = k) &= \int_k^{k+1} \mu^{-1} e^{-x/\mu}\, dx = \left[-e^{-x/\mu}\right]_k^{k+1} \\ &= e^{-k/\mu} - e^{-(k+1)/\mu} = e^{-k/\mu}(1 - e^{-1/\mu}) \\ &= p^k(1 - p), \end{aligned}$$

där $p = e^{-\mu^{-1}}$. Vi ser alltså att $Y \sim \text{Geo}(p)$ (Y är geometrisk fördelad med parametern p).

Svar: $X \sim \text{Geo}(e^{-\mu^{-1}})$.