LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (92MA31, STN2)
2014-01-10 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 efter skrivningens slut. Lycka till!

---

1. (a) Låt $P((A \cup B)^c) = 0.4$ och $P(A) = P(B)$. Om A och B är oberoende, vad kan $P(A)$ anta för värden? (2p)

   (b) Om $P(A) = P(B)$, är A och B oberoende? Bevisa eller ange motexempel. (2p)

   (c) Låt $P((A \cup B)^c) = 0.4$, $P(A) = P(B)$ och $P(A \,|\, B) = 0.5$. Vad kan $P(A)$ anta för värden? (2p)

2. Vid en reaktion mäter man hur mycket av ett visst ämne som bildas beroende på hur lång tid man låter reaktionen fortgå.

   (a) Vid mätningar vid olika tidsvärden x_i (kända) fann man följande samband mellan det uppmätta innehållet y och x (i lämpliga enheter):

x	2	4	6	8	10	12
y	6.45	11.87	20.43	25.59	32.75	42.98

   Använd linjär regression med modellen $y_k = ax_k + b + \epsilon_k$, $k = 1, 2, 3, 4, 5, 6$, där $\epsilon_k \sim N(0, \sigma)$ är oberoende och $a, b \in \mathbf{R}$ är konstanter, för att bestämma en linje $y = ax + b$ som minimerar kvadratfelet. (2p)

   (b) Vad kan man säga om innehållet (y-värdet) efter 20 tidsenheter ($x = 20$)? (1p)

   (c) För ett okänt prov som reagerat klart gjorde man 15 upprepade mätningar och fann medelvärdet 37.1352 samt stickprovsvariansen $s^2 = 3.8416$. Antag att dessa mätningar är observationer från oberoende $N(\mu, \sigma)$-variabler. Ange ett 95% konfidensintervall för väntevärdet μ. (3p)

3. Timmy brukar sitta och jonglera en kniv när han känner sig rastlös. För en viss riktigt obalanserad Rambokniv är sannolikheten för en skada konstant lika med 10^{-3} per kast. Antag att olika kast är oberoende.

   (a) Om Timmy jonglerar genom att kasta kniven 25 gånger, vad är sannolikheten att han skadar sig själv högst en gång? (2p)

   (b) Timmy kör 500 stycken serier med 25 kast i varje, vad är approximativt sannolikheten att han blir skadad mer än åtta gånger? (2p)

   (c) Om Timmy kastar gång på gång tills dess att han till slut skadar sig, vad är sannolikheten att detta tar fler än 1000 kast? (2p)

Vänd!

4. Låt X och Y vara två oberoende stokastiska variabler. Låt X ha täthetsfunktionen $f_X(x) = c_1 x^2$ för $0 < x < 2$ och Y ha täthetsfunktionen $f_Y(y) = c_2$ för $-1 < y < 1$ (och noll för övriga x respektive y).

   (a) Bestäm c_1 och c_2 samt den simultana täthetsfunktionen för (X, Y). (2p)

   (b) Beräkna $P(X > 1 \text{ och } Y < 0)$. (1p)

   (c) Beräkna $P(X \leq 2Y)$. (3p)

5. Låt $X \sim \mathrm{N}(10, 2)$ där $\sigma = 2$ är standardavvikelsen.

   (a) Låt $Z = 0$ då $X < 8$, $Z = 1$ då $8 \leq X < 12$, $Z = 2$ då $X = 12$, och $Z = 3$ då $X > 12$. Bestäm sannolikhetsfunktionen för Z, $E(Z)$, $V(Z)$ samt $E(Z^2)$. (4p)

   (b) Bestäm $P(Z^2 = 2Z + 3)$ samt $P(|Z| > 2)$ med Z enligt ovan. (2p)

6. I en stor (oändlig) mängd förekommer en viss egenskap med sannolikhet p för ett slumpmässigt utvalt element. Om vi väljer ut n stycken element ur mängden brukar vi som bekant skatta p med hjälp av den stokastiska variabeln $\widehat{P} = \dfrac{X}{n}$ där $X \sim \mathrm{Bin}(n, p)$.

   (a) Visa att $\widehat{P}$ är en väntevärdesriktig skattning av p. (1p)

   (b) Visa att $\dfrac{X^2}{n^2}$ är en skattning av p^2 och undersök sedan hurvida den är väntevärdesriktig. Motivera dina svar ordentligt! (3p)

   (c) Undersök om $\dfrac{X^2}{n^2}$ är en asymptotiskt väntevärdesriktig skattning av p^2, dvs undersök om

$$\lim_{n \to \infty} E(X^2/n^2) = p^2.$$

(2p)

Lösningsskisser för matematisk statistik 2014-01-10

1. (a) Låt $P(A) = x$. Eftersom $P(A \cap B) = P(A)P(B) = x^2$ erhåller vi

$$0.6 = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2x - x^2.$$

Vi löser $x^2 - 2x + 0.6 = 0$ och finner att $x = 1 \pm \sqrt{0.4} = 1 \pm 0.63$. Endast $x = 0.37$ är möjlig (varför?).

(b) Nej. Samma sannolikhet har inget med oberoende att göra. Singla en slant och låt A vara krona och B klave. Ett rättvist mynt har $P(A) = P(B) = 1/2$, men uppenbarligen är $P(A \cap B) = 0$ så händelserna är beroende (vilket även verkar vettigt intuitivt).

(c) Eftersom $P((A \cup B)^c) = 0.4$ så är $P(A \cup B) = 0.6$. Låt $P(A) = x$. Vi vet att $P(A \mid B) = 0.5$ och att

$$0.6 = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2x - P(A \cap B),$$

så

$$0.5 = P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{2x - 0.6}{x}.$$

Alltså måste $0.5x = 2x - 0.6$, eller $x = 0.6/1.5 = 0.4$.

Svar: (a) 0.37 (b) Se ovan. (c) 0.40.

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

$$b_0 = -1.70 \qquad \text{och} \qquad b_1 = 3.58.$$

Vår regressionslinje blir alltså $y = b_0 + b_1 x = -1.70 + 3.58x$.

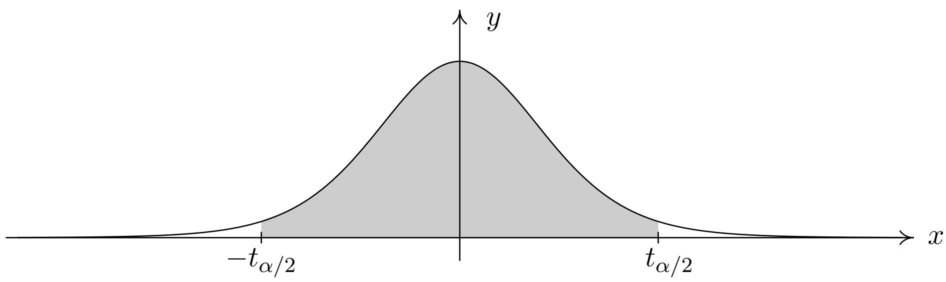
(b) Inget kan sägas då $x = 20$ ligger långt utanför det intervall vi har mätningar inom. Sådan extrapolation är livsfarlig! Till exempel skulle ämnena som reagerar kunna ta slut efter säg $x = 15$ tidsenheter så inget händer därefter.

(c) Vi har $n = 15$ och bildar testvariabeln

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t(14)$$

och ser att

$$P(-t_{\alpha/2}(14) \le T \le t_{\alpha/2}(14)) = 1 - \alpha.$$



Figur 1: Den markerade arean innehåller 95% av sannolikheten för $t(14)$-fördelningen.

Vi löser ut μ ur intervallet i sannolikhetsmåttet och erhåller att

$$\overline{Y} - \frac{S}{\sqrt{n}} t_{\alpha/2}(14) \le \mu \le \overline{Y} + \frac{S}{\sqrt{n}} t_{\alpha/2}(14).$$

Vi ersätter $\overline{Y}$ med den observerade punktskattningen $\widehat{y} = 10.59$ och stickprovsvariansen S med $s = \sqrt{3.84} = 1.96$. Då erhåller vi ett konfidensintervall med konfidensgrad 95%:

$$I_\mu = [36.0,\ 38.2],$$

där vi använt $\alpha = 0.05$ och $t_{0.025}(14) = 2.145$.

Svar: (a) $y = -1.70 + 3.58x$. (b) Inget. (b) $I_\mu = [36.0,\ 38.2]$.

3. (a) Låt X vara antalet gånger Timmy skadar sig vid 25 kast. Då är $X \sim \mathrm{Bin}(25, 1/1000)$, och således

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \binom{25}{0} p^0(1-p)^{25} + \binom{25}{1} p(1-p)^{24} = 0.9997,$$

där $p = 1/1000$.

(b) Låt $X_i \sim \mathrm{Bin}(25, 1/1000)$ för $i = 1, 2, \ldots, 500$ vara oberoende och definiera $Y = \sum_{i=1}^{500} X_i$. Då är $E(Y) = 500 \cdot 25/1000 = 12.5$ och $V(Y) = 500 \cdot 12.5 \cdot (1 - 1/1000) = 3.534^2$. Eftersom det är många termer i summan kommer den att bli approximativt normalfördelad, så $Y \overset{\text{appr.}}{\sim} \mathrm{N}(12.5, 3.534)$. Från detta följer att

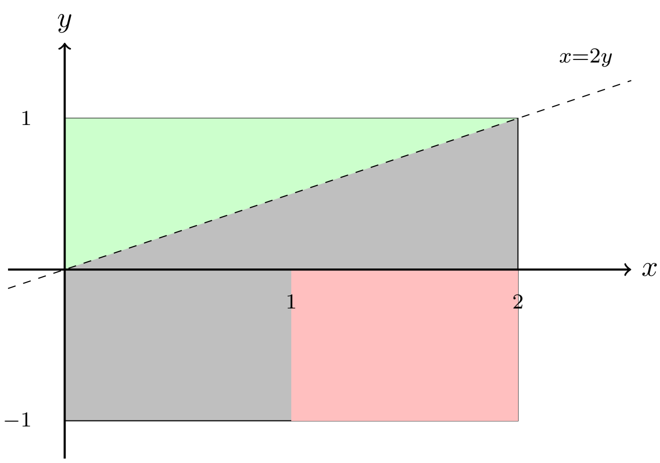
$$P(Y > 8) = 1 - P(Y \leq 8) \approx 1 - \Phi((8 - 12.5)/3.534) \approx 0.90.$$

(c) Låt $Z \sim \mathrm{Ffg}(1/1000)$. Vi söker sannolikheten att $Z > 1000$:

$$P(Z > 1000) = \sum_{k=1001}^{\infty} p(1-p)^k = \frac{1}{1000} \sum_{k=1001}^{\infty} (999/1000)^k = [\text{ geom. serie }] = 0.367.$$

Svar: (a) Ca 99.97%. (b) 90%. (c) 36.7%.

4. Vi skisserar de intressanta områdena i Figur 2.



Figur 2: Den skuggade rektangeln är området där $f_{X,Y}$ inte är noll. Den övre triangeln (grön) är det område där $2y \geq x$ och den undre rektangeln (rosa) är det område där $x > 1$ och $y < 0$.

(a) Konstanten c_1 fås ur sambandet

$$c_1^{-1} = \int_0^2 x^2\, dx = \frac{8}{3},$$

så $c_1 = 3/8$. Den s.v. Y är likformigt fördelad på $[-1, 1]$ så $c_2 = 1/2$. Eftersom variablerna är oberoende får vi den simultana täthetsfunktionen i form av produkten av de marginella:

$$f_{X,Y}(x, y) = \begin{cases} \frac{3x^2}{16}, & \text{för } -1 < y < 1 \text{ och } 0 < x < 2, \\ 0, & \text{för övrigt.} \end{cases}$$

(b) Eftersom X och Y är oberoende så är

$$P(X > 1, Y < 0) = P(X > 1)P(Y < 0) = \left(\int_1^2 \frac{3x^2}{8}\, dx \right) \frac{1}{2} = \frac{7}{16}.$$

(c) Här blir det lite bökigare då gränserna för X och Y beror på varandra, men en skiss (se Figur 2) leder till att

$$P(X \leq 2Y) = \frac{3}{16} \int_0^1 \int_0^{2y} x^2\, dxdy = \cdots = \frac{1}{8}.$$

Svar: (a) $c_1 = 3/8$ och $c_2 = 1/2$. (b) 7/16. (c) 1/8.

5. (a) De möjliga värdena Z kan anta är $0, 1, 2, 3$. Vi ställer upp sannolikhetsfunktionen för Z:

$$\begin{aligned}
P(Z = 0) &= P(X < 8) = \Phi((8 - 10)/2) = \Phi(-1) = 0.1587, \\
P(Z = 1) &= P(8 \leq X < 12) = \Phi((12 - 10)/2) - \Phi((8 - 10)/2) = 0.6827, \\
P(Z = 2) &= P(X = 12) = 0, \\
P(z = 3) &= P(X > 12) = 1 - \Phi((12 - 10)/2) = 0.1587.
\end{aligned}$$

Från detta kan vi sedan direkt bestämma att

$$E(Z) = \sum_{k=0}^{3} kP(Z = k) = 1.1588 \quad \text{och} \quad E(Z^2) = \sum_{k=0}^{3} k^2 P(Z = k) = 2.1110.$$

Vidare, $V(Z) = E(Z^2) - E(Z)^2 = 0.7682$.

(b) Eftersom $Z^2 - 2Z - 3 = 0 \quad \Leftrightarrow \quad (Z - 3)(Z + 1) = 0$ så är

$$P(Z^2 - 2Z - 3 = 0) = P(Z = 3) + P(Z = -1) = 0.1587 + 0.$$

Vidare, det enda Z med positiv sannolikhet där $|Z| > 2$ är $Z = 3$, så

$$P(|Z| > 2) = P(Z = 3) = 0.1587.$$

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

6. (a) Eftersom $X \sim \text{Bin}(n, p)$ så är $E(\widehat{P}) = \dfrac{1}{n}E(X) = \dfrac{np}{n} = p$. Således väntevärdesriktig!

(b) Allt (ja, varenda en du kan komma på!) är skattningar av p^2. Men de flesta är så klart väldigt dåliga. Är $\widehat{P}^2$ väntevärdesriktig? Vi undersöker:

$$\begin{aligned}
E(\widehat{P}^2) &= \frac{E(X^2)}{n^2} = \frac{V(X) + E(X)^2}{n^2} = \frac{np(1 - p) + (np)^2}{n^2} \\
&= p^2 \left(1 - \frac{1}{n}\right) + \frac{p}{n}.
\end{aligned}$$

Alltså kommer inte väntevärdet för skattningsvariabeln att bli p^2. Detta är alltså inte en väntevärdesriktig skattning! Tycker du den verkar vettig ändå?

(c) Från uttrycket i föregående deluppgift kan vi enkelt se att

$$\lim_{n \to \infty} E(\widehat{P}^2) = p^2\,(1-0) + 0 = p^2,$$

så enligt definitionen i uppgiften är detta en s.k. asymptotiskt väntevärdesriktig skattning av p^2.

Svar: Se ovan.