LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (9MA241/9MA341/LIMAB6, STN2) 2011-05-31 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. Nina har två katter, Franki och Lanki. Sannolikheten att Franki kräks på en dag är 0.3, och sannolikheten att Lanki kräks samma dag är 0.05.
   (a) Om händelserna att katterna kräks är oberoende, vad är sannolikheten att ingen av katterna kräks på en given dag? (2p)
   (b) Oberoendeantagandet ifrågasattes, och via empiriska studier fann man att den betingade sannolikheten att Franki kräks på en dag givet att Lanki kräks samma dag är 0.5. Vad är den betingade sannolikheten att Lanki kräks givet att Franki kräks samma dag? (2p)
   (c) Är två oförenliga händelser alltid oberoende? Bevisa påståendet eller ge motexempel. (2p)

2. Vid en mätning av en process fick man följande mätdata:

   6.04 4.96 4.93 3.40 7.04 4.73 3.57 7.70 4.55 3.82

   En hjälpsam examinator har räknat ut följande:

   $$\sum_{i=1}^{10} x_i = 50.74 \qquad \text{och} \qquad \sum_{i=1}^{10} x_i^2 = 276.2124.$$

   (a) Antag att mätningarna är ett stickprov på en normalfördelad variabel $X \sim N(\mu, 2)$ (man tycker sig veta så pass mycket om processen att standardavvikelsen anses vara känd). Beräkna ett 99% konfidensintervall för väntevärdet μ. (3p)
   (b) En som arbetar med processen håller inte med om att standardavvikelsen kan antas vara given, utan tycker att man måste skatta den utifrån datan. Hjälp personen i fråga med att ställa upp ett 99% konfidensintervall för väntevärdet μ då mätningarna är ett stickprov på en normalfördelad variabel $X \sim N(\mu, \sigma)$ och σ är okänd. (3p)

3. (a) Låt $X_1 \sim N(1, \sqrt{2})$, $X_2 \sim N(-1, \sqrt{2})$, och $X_3 \sim N(0, 1)$, vara oberoende stokastiska variabler. Beräkna $P(X_1 + X_2 - X_3 \leq \frac{2}{5})$. (4p)
   (b) Låt $X_i \sim N(1, 2)$, $i = 1, 2, \ldots$, vara en följd av (parvis) oberoende stokastiska variabler. Hur många måste man ta med i ett medelvärde för att få $V(\bar{X}) < \frac{1}{10}$? Med andra ord, hur stort måste n vara för att

   $$V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) < \frac{1}{10}?$$

4. Låt $X \sim \text{Exp}(1/2)$ och $Y \sim \text{Exp}(1/3)$ (parametrarna är väntevärderna) vara oberoende stokastiska variabler.

   (a) Vad blir den simultana täthetsfunktionen $f_{X,Y}(x, y)$? (1p)

   (b) Beräkna $P(X > 1/2,\ Y < 1/2)$. (2p)

   (c) Beräkna $E(X^2)$, variansen $V(X + Y)$, samt kovariansen $C(X, Y)$. (3p)

5. Dana Scully och Fox Mulder befinner sig i en liten stad i New Mexico, population 3200, för att utreda orsaken till en flygplanskrasch. Det rapporterats om många UFO observationer i området och Mulder tror att det rör sig om en utomjordisk inblandning. Dana är mer tveksam och tror att det handlar om en militär övning vid en närliggande "hemlig" bas som gått snett. Eftersom de inte kan komma överens om hur de skall börja, så övertalar Dana Mulder att de skall göra en statistik undersökning över vad befolkningen i staden tror. Man frågar 400 personer i staden om de tror att det rör sig om en utomjordisk inblandning, och 30 personer svarar ja.

   (a) Hjälp Mulder att ställa upp ett approximativt 95%, enkelsidigt (nedåt begränsat), konfidensintervall för andelen av befolkningen i staden som tror det rör sig om utomjordingar. (3p)

   (b) Dana tycker att man borde titta på skillnaden mellan hur många som tror det rör sig om utomjordingar och hur många som tror att det är militären som ligger bakom, istället. Hon frågar därför 300 personer i staden om de tror att militären ligger bakom. Av dessa svarade 20 stycken ja. Hjälp Scully ställa upp ett dubbelsidigt konfidensintervall, med approximativ konfidensgrad 95%, för skillnaden mellan andelen som tror att utomjordingar ligger bakom kraschen och andelen som tror att det är militärens fel. Kan man dra någon statistiskt signifikant slutsats? (3p)

6. Låt $X \sim \text{Exp}(1)$ vara en stokastisk variabel och låt $F_X(a) = P(X \leq a)$ som vanligt. Vidare så låter vi $Y \sim \text{Ffg}(F_X(a))$ vara en annan stokastisk variabel (det vill säga, Y är för-första-gången fördelad med sannolikheten $p = F_X(a)$). Bestäm det största talet a så att

$$P(Y \leq 2) \leq \frac{1}{10}.$$

Lösningsskisser för tentamen i matematisk statistik, 9MA241, 2011-05-31

1. Låt A vara händelsen att Franki kräks och B händelsen att Lanki kräks.

   (a) Då ges händelsen att ingen katt kräks på en given dag av $(A \cup B)^*$. Vi får

   $$\begin{aligned} P((A \cup B)^*) &= 1 - P(A \cup B) = 1 - \big(P(A) + P(B) - P(A \cap B)\big) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= 1 - 0.3 - 0.05 + 0.015 = 0.665, \end{aligned}$$

   där vi utnyttjat att A och B antas vara oberoende på näst sista raden.

   (b) Ifrån uppgiften vet vi att $P(A \,|\, B) = 0.5$. Detta ger att

   $$P(A \cap B) = P(A \,|\, B)P(B) = 0.5 \cdot 0.05 = 0.025$$

   enligt definitionen av betingad sannolikhet. Följaktligen blir

   $$P(B \,|\, A) = \frac{P(A \cap B)}{P(A)} = \frac{0.025}{0.3} = 0.0833$$

   efter ännu en tillämpning av definitionen på betingad sannolikhet.

   (c) Tag, t ex, händelserna A och A^* (alltså händelserna att Franki kräks och att Franki inte kräks). Uppenbarligen så är dessa händelser oförenliga (kan ej inträffa samtidigt). Men eftersom $P(A \cap A^*) = P(\emptyset) = 0$ och $P(A)P(B) = 0.3 \cdot 0.7 \neq 0$ så är dessa händelser ej oberoende.

2. **Modell I**: vi betraktar siffrorna som ett stickprov från oberoende s. v. $X_j \sim N(\mu, 2)$. Vi punktskattar väntevärdet μ med

   $$\mu^* = \bar{X} = \frac{1}{10}\sum_{j=1}^{10} X_j \sim N(\mu, 2/\sqrt{10}).$$

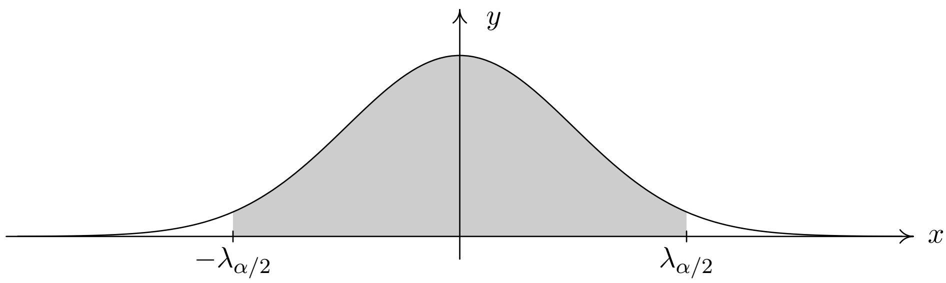
   Vi skapar testvariabeln

   $$Z = \frac{\mu^* - \mu}{\sigma/\sqrt{10}} \sim N(0,1).$$

   Det följer då att

   $$P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha, \tag{0.1}$$

   och då vi söker ett 99% konfidensintervall så är $\alpha = 0.01$ och $\lambda_{0.005} \approx 2.575$ (det sista ur tabell).



Figur 1: Det skuggade området är $(1 - \alpha) \cdot 100\%$ av sannolikhetsmassan.

Vi löser ut μ ur olikheten i sannolikhetsmåttet i ekvation (0.1) ovan och erhåller att

$$\mu^* - \frac{2.575 \cdot 2}{\sqrt{10}} < \mu < \mu^* + \frac{2.575 \cdot 2}{\sqrt{10}}.$$

Om vi ersätter μ^* med den observerade punktskattningen $\mu^*_{\text{obs}} = 5.074$ (medelvärdet av observationerna) så får vi ett konfidensintervall $I_\mu = [3.45, 6.70]$ med konfidensgrad 99%.

Modell II: vi betraktar siffrorna som ett stickprov från oberoende s. v. $X_j \sim N(\mu, \sigma)$. Vi punktskattar med $\mu^* = \bar{X} \sim N(\mu, \sigma/\sqrt{10})$ som tidigare och skattar σ med s, där

$$s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2 \approx 2.0842$$

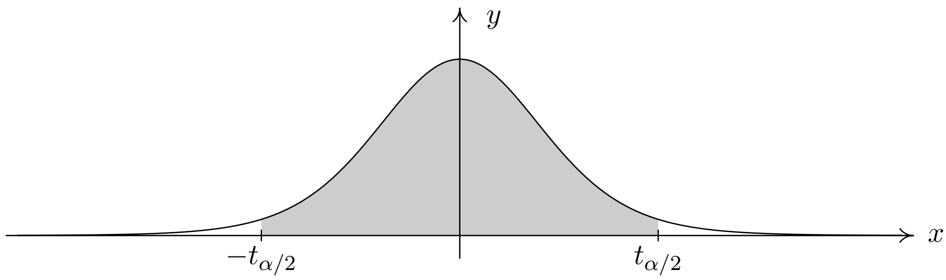
är stickprovsvariansen. Vi skapar testvariabeln

$$T = \frac{\mu^* - \mu}{s/\sqrt{10}} \sim t(9).$$

Som i förra deluppgiften följer det att

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

där $t_\alpha(9)$ är kvantilerna till $t(9)$-fördelningen.



Figur 2: Samma situation som tidigare fast med $t(9)$-fördelningen istället.

Ur tabell finner vi $t_{0.005}(9) = 3.25$. Genom att lösa ut μ ur olikheten i sannolikhetsmåttet får vi

$$\mu^* - \frac{3.25 \cdot 1.444}{\sqrt{10}} < \mu < \mu^* + \frac{3.25 \cdot 1.444}{\sqrt{10}}.$$

Om vi ersätter μ^* med den observerade punktskattningen $\mu^*_{\text{obs}} = 5.074$ (medelvärdet av observationerna) så får vi ett konfidensintervall $I_\mu = [3.59, 6.56]$ med konfidensgrad 99%.

Svar: a) $I_\mu = [3.45, 6.70]$ b) $I_\mu = [3.59, 6.56]$.

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

$$Z = X_1 + X_2 - X_3 \sim N\left(1 + (-1) + 0,\ \sqrt{(\sqrt{2})^2 + (\sqrt{2})^2 + 1}\right),$$

Vi beräknar:

$$P(Z \le 2/5) = \Phi\left(\frac{(2/5 - 0)}{\sqrt{5}}\right) = \Phi(2/(5\sqrt{5})) \approx \Phi(0.179) = 0.571.$$

(b) Vi vet att

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\big(0, \frac{2}{\sqrt{n}}\big),$$

så $V(\bar{X}) = 4/n$. Det följer att $V(\bar{X}) < 1/10$ om $n > 40$. Det duger alltså med $n = 41$.

4. (a) Eftersom X och Y är oberoende gäller att $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, så

$$f_{X,Y}(x,y) = \begin{cases} 6e^{-(2x+3y)}, & x,\ y > 0, \\ 0, & \text{för övrigt.} \end{cases}$$

(b) Återigen, eftersom variablerna är oberoende erhåller vi

$$\begin{aligned} P(X > 1/2,\ Y < 1/3) &= P(X > 1/2)P(Y < 1/2) \\ &= \int_{1/2}^{\infty} 2e^{-2x}\,dx \cdot \int_0^{1/2} 3e^{-3y}\,dy \\ &= e^{-1} \cdot (1 - e^{-3/2}) = e^{-1} - e^{-5/2}. \end{aligned}$$

(c) Eftersom X och Y är oberoende så gäller att

$$V(X+Y) = V(X) + V(Y) = \frac{1}{4} + \frac{1}{9} = \frac{13}{36}.$$

Vidare så är $E(X^2) = V(X) + E(X)^2 = 1/4 - (1/2)^2 = 1/2$, och eftersom X och Y är oberoende så måste $C(X,Y) = 0$.

5. Som testvariabler använder vi $\widehat{P}_1 = X_1/400$ och $\widehat{P}_2 = X_2/300$, där X_1 är antalet som tror på UFO-förklaringen, och X_2 är antalet som tror att det är militärens fel. Det följer att båda är hypergeometriskt fördelade: $X_1 \sim \text{Hyp}(3200, 400, p_1)$ och $X_2 \sim \text{Hyp}(3200, 300, p_2)$, där p_1 och p_2 är de okända andelarna av befolkningen som vi vill undersöka. Vi ser att $\widehat{P}_1$ och $\widehat{P}_2$ är (bra) punktskattningar för dessa andelar. Vidare finner vi att

$$\begin{aligned} E(\widehat{P}_1) &= \frac{1}{400}E(X_1) = \frac{400p_1}{400} = p_1, \\ E(\widehat{P}_2) &= \cdots = p_2, \\ V(\widehat{P}_1) &= \frac{1}{400^2}\left(\frac{3200-400}{3199}400p_1(1-p_1)\right) = 2.188 \cdot 10^{-3} \cdot p_1(1-p_1), \\ V(\widehat{P}_2) &= \cdots = 3.022 \cdot 10^{-3} \cdot p_2(1-p_2). \end{aligned}$$

Om vi ersätter p_1 och p_2 med respektive observerade punktskattningar

$$\widehat{p}_1 = \frac{30}{400} \qquad \text{och} \qquad \widehat{p}_2 = \frac{20}{300},$$

finner vi att $V(\widehat{P}_1) \approx 1.518 \cdot 10^{-4}$ och $V(\widehat{P}_2) \approx 1.880 \cdot 10^{-4}$. Vi ser även att $V(X_1) \approx 24.3$ och att $V(X_2) \approx 16.9$ (t ex följer det från identiteterna $X_1 = 400\widehat{P}_1$ och $X_2 = 300\widehat{P}_2$), så normalapproximation borde fungera adekvat. Vi bör inte approximera till Binomialfördelning först då kvoterna 400/3200 och 300/3200 ligger lite väl nära gränsen 0.1. För att summera:

$$\widehat{P}_1 \overset{\text{appr.}}{\sim} N(p_1, \sqrt{1.518 \cdot 10^{-4}}) \qquad \text{och} \qquad \widehat{P}_2 \overset{\text{appr.}}{\sim} N(p_2, \sqrt{1.880 \cdot 10^{-4}}).$$

(a) Eftersom vi söker ett nedåt begränsat konfidensintevall söker vi $\lambda_{0.05}$ så att

$$P(Z_1 < \lambda_{0.05}) = 0.95,$$

där Z_1 är hjälpvariabeln

$$Z_1 = \frac{\widehat{P}_1 - p_1}{\sqrt{1.518 \cdot 10^{-4}}} \overset{\text{appr.}}{\sim} N(0,1).$$

I tabell hittar vi $\lambda_{0.05} \approx 1.645$. Vi löser ut p_1 ur olikheten $Z_1 < 1.645$ och erhåller intervallet

$$I_{p_1} = [0.05,\ 1].$$

(b) Tvåsidigt intervall för andelsskillnad med approximativt 95% konfidensgrad. Vi använder hjälpvariabeln

$$Z = \frac{\widehat{P}_1 - \widehat{P}_2 - (p_1 - p_2)}{\sqrt{(1.518 + 1.880) \cdot 10^{-4}}}.$$

Eftersom $\widehat{P}_1$ och $\widehat{P}_2$ är approximativt normalfördelade så gäller att

$$Z \overset{\text{appr.}}{\sim} N(0,1).$$

Vi behöver hitta $\lambda_{0.025}$ så att

$$P(-\lambda_{0.025} < Z < \lambda_{0.025}) = 0.95.$$

Ur tabell finner vi att $\lambda_{0.025} = 1.96$. Löser vi ut $p_1 - p_2$ ur olikheten får vi konfidensintervallet

$$I_{p_1 - p_2} = [-0.028, 0.044].$$

Det går inte att dra några statistiskt säkra slutsatser från detta. Det är fullt rimligt att det inte är någon skillnad mellan andelarna (d v s att båda åsikterna är lika vanliga).

6. Eftersom $X \sim \text{Exp}(1)$ så får vi

$$p = F_X(a) = \int_0^a e^{-x}\, dx = 1 - e^{-a}, \qquad a > 0.$$

Då $Y \sim \text{ffg}(p)$ blir

$$\begin{aligned} P(Y \le 2) &= p + (1-p)p = (1 - e^{-a}) + (1 - (1 - e^{-a}))(1 - e^{-a}) \\ &= 1 - e^{-2a}. \end{aligned}$$

Om $P(Y \le 2) \le 1/10$ så måste alltså $1 - e^{-2a} \le 1/10$. Ekvivalent,

$$\frac{9}{10} \le e^{-2a} \quad \Leftrightarrow \quad \ln\left(\frac{9}{10}\right) \le -2a \quad \Leftrightarrow \quad a \le -\frac{1}{2}\ln\left(\frac{9}{10}\right) = \frac{1}{2}\ln\left(\frac{10}{9}\right).$$

Vi väljer alltså $a = (1/2)\ln(10/9)$.