LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (92MA31, STN2)
2014-08-15 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 efter skrivningens slut. Lycka till!

---

1. Motivera svaren på följande frågor noggrant!

   (a) Låt $f(x) = 2(x+1)/3$ för $-2 \leq x < 1$ och $f(x) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (b) Låt $f(x) = 2(x+1)/3$ för $-1 \leq x \leq 1$ och $f(x) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (c) Låt $f(x,y) = 3x^2/2$ för $-1 < x < 1$ och $0 \leq y \leq 1$, och $f(x,y) = 0$ för övrigt. Är f en täthetsfunktion? (1p)

   (d) Låt $P(A) = P(B) = 0.3$. Beräkna $P(A \cap B)$ om A och B är oberoende. (1p)

   (e) Gäller formeln $P(A \mid B)P(B) + P(B \mid A)P(A) = 2P(A \cap B)$ för alla händelser A och B? (1p)

   (f) Låt $P(A) + P(B) = 1.2$. Motivera om händelserna A och B kan vara disjunkta. (1p)

2. Låt $X_1 \sim N(0,2)$ och $X_2 \sim N(1,2)$, där $V(X_1) = V(X_2) = 4$, vara oberoende.

   (a) Bestäm sannolikheten $P(X_2 < 1.5)$. (2p)

   (b) Bestäm sannolikheten $P(X_2 > 2X_1)$. (2p)

   (c) Bestäm konstanten a så att variansen av $aX_1 + (1-a)X_2$ blir så liten som möjligt. (2p)

3. (a) Vid en följd mätningar av oberoende stokastiska variabler X_i, $i = 1,2,3,4,5,6$, där varje $X_i \sim N(\mu, \sigma)$ och μ samt variansen σ^2 är okända, erhöll man följande mätdata:

   $$14.68 \quad 18.57 \quad 13.98 \quad 14.38 \quad 22.45 \quad 22.05$$

   Ange ett 95% konfidensintervall för väntevärdet μ. (3p)

   (b) Vid en undersökning av åsikten i en viss fråga där man vet att ca 40% har den åsikten, funderar man på hur många personer man måste fråga för att kunna få ett 95%-igt konfidensintervall för den okända andelen p av längd högst 0.10. Antag att den stora populationen är oändlig och bestäm approximativt det minsta antalet personer man måste fråga. (3p)

Vänd!

4. Svampsäsongen närmar sig och planerar man att plocka svamp bör man vara försiktig. Antag att det växer 14 olika sorters svampar i en skogsbacke. Alla är lika vanliga, men två av sorterna är giftiga och tre av de övriga icke-giftiga svamparna smakar så pass illa att man inte vill använda dessa som matsvamp.

   (a) Om man på måfå plockar 6 stycken svampar, vad är sannolikheten att minst två är giftiga? (3p)

   (b) LD50 (d v s dosen 50% av testpopulation dör av) för de giftiga sorterna ligger vid ca 10 stycken svampar. Om någon mot förmodan helt på måfå skulle käka 105 svampar, vad är sannolikheten att personen uppnår LD50-gränsen? Approximationer är OK om dessa motiveras. (3p)

5. Låt X och Y vara två oberoende exponentialfördelade stokastiska variabler med $E(X) = 5$ och $E(Y) = 7$.

   (a) Bestäm den simultana täthetsfunktionen $f_{X,Y}(x, y)$ för den 2-dimensionella s.v. (X, Y). Var noggran med vart funktionen är noll! (2p)

   (b) Beräkna $P(X > 1 \text{ och } Y < 1)$. (2p)

   (c) Beräkna $P(\max\{X, Y\} \leq 5)$. (2p)

6. Låt $X_1, X_2, \dots, X_n$ vara n stycken stokastiska variabler med $E(X_i) = \mu$ och $V(X_i) = \sigma^2$ för alla $i = 1, 2, \dots, n$. Visa att

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{och} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

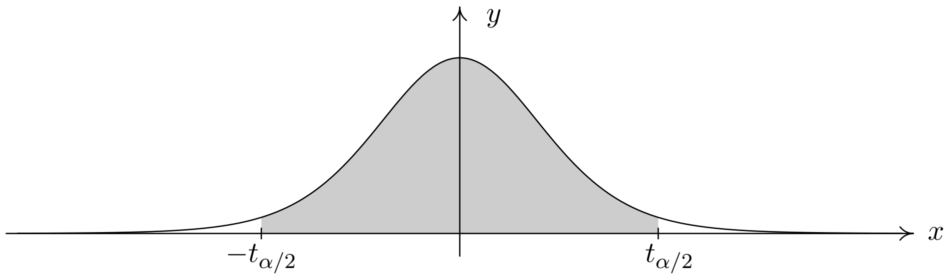
är väntevärdesriktiga skattningar av μ respektive σ^2. (6p)

Lösningsskisser för matematisk statistik 2014-08-15

1. (a) Nej, $f < 0$ då $-2 < x < -1$.
   (b) För $-1 \leq x \leq 1$ så är $f \geq 0$, men
   $$\int_{-1}^{1} f(x)\,dx = \left[\frac{2}{3}\frac{(x+1)^2}{2}\right]_{-1}^{1} = \frac{4}{3} \neq 1,$$
   så detta är inte heller en täthetsfunktion.
   (c) Funktionen är ≥ 0 i området så det återstår att kontrollera integralvillkoret:
   $$\int_0^1 \int_{-1}^1 \frac{3x^2}{2}\,dxdy = \frac{3}{2}\left[\frac{x^3}{3}\right]_{-1}^{1} = 1.$$
   Alltså en täthetsfunktion!
   (d) Eftersom $P(A \cap B) = P(A)P(B)$ då händelserna är oberoende följer det att $P(A \cap B) = 0.09$.
   (e) Svar ja, eftersom
   $$P(A \mid B)P(B) + P(B \mid A)P(A) = \frac{P(A \cap B)}{P(B)}P(B) + \frac{P(B \cap A)}{P(A)}P(A) = 2P(A \cap B)$$
   för alla händelser där $P(A) > 0$ och $P(B) > 0$. Vad händer om någon händelse har sannolikhet noll?
   (f) Händelserna kan inte vara disjunkta, ty då skulle $P(A \cup B) = P(A) + P(B) = 1.2 > 1$.

2. (a) $P(X_2 < 1.5) = P\left(\frac{X_2 - 1}{2} \leq 0.25\right) = \Phi(0.25) = 0.60$
   (b) Låt $Y = X_2 - 2X_1$. Då är $Y \sim \mathrm{N}(1, \sqrt{20})$ (där $V(Y) = 20$) och
   $$P(X_2 > 2X_1) = P(Y > 0) = 1 - P(Y \leq 0) = 1 - \Phi(-1/\sqrt{20}) = \Phi(1/\sqrt{20}) = 0.5885$$
   (c) Vi låter $v(a) = V(aX_1 + (1-a)X_2)$ och vill alltså minimera funktionen v. Vi börjar med att förenkla. Eftersom variablerna är oberoende följer det att
   $$v(a) = 4a^2 + 4(1-a)^2 = 4(2a^2 + 1 - 2a).$$
   Vi gör en funktionsundersökning; $v'(a) = 0$ sker endast då $a = 1/2$ och $v''(1/2) > 0$ så detta är det minimum vi är ute efter. Svaret är alltså $a = 1/2$.

3. (a) Lämplig testvariabel är
   $$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(14)$$
   och ser att
   $$P(-t_{\alpha/2}(5) \leq T \leq t_{\alpha/2}(5)) = 1 - \alpha.$$
   Vi löser ut μ ur intervallet i sannolikhetsmåttet och erhåller att
   $$\overline{Y} - \frac{S}{\sqrt{n}}t_{\alpha/2}(5) \leq \mu \leq \overline{Y} + \frac{S}{\sqrt{n}}t_{\alpha/2}(5).$$
   Vi ersätter $\overline{Y}$ med den observerade punktskattningen $\widehat{y} = 17.685$ och stickprovsvariansen S^2 med $s^2 = 3.9044^2$. Då erhåller vi ett konfidensintervall med konfidensgrad 95%:
   $$I_\mu = [13.59,\ 21.78],$$
   där vi använt $\alpha = 0.05$ och $t_{0.025}(5) = 2.57$.



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

(b) En naturlig skattning på den verkliga andelen ges av

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

där $X \sim \text{Bin}(n, p)$ är antalet som har åsikten, n är antalet vi frågar och p är den okända andelen. Vi antar att n är tillräckligt stor för att vi ska kunna göra en normalapproximation, i.e.,

$$np(1-p) \approx n \cdot 0.40 \cdot 0.60 \geq 10,$$

eller $n \geq 42$ (förutsätter att p ligger nära 0.40). 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

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} \approx \sqrt{0.24/n}.$$

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

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

Intervallängden ges då av

$$l = 2\lambda_{0.025} d \leq 0.10,$$

vilket ger att $n \geq \frac{(2\lambda_{0.025})^2 \cdot 0.24}{0.10^2}$ eller $n \geq 369$. Detta uppfyller kravet för normalapproximation.

4. Vi låter X vara antal giftiga svampar bland n plockade på måfå. Då är $X \sim \text{Bin}(n, 1/7)$ eftersom varje svamp har en sannolikhet på 2/14 att vara giftig.

(a) I denna deluppgift är $n = 6$ och vi är intresserade av sannolikheten att $X \geq 2$. Vi får

$$P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - 0.7931 = 0.2069.$$

Alltså är sannolikheten 20.7%.

(b) I det här fallet är $n = 105$, så vi är inte hjälpt av varken tabell eller direkta kalkyler (alldeles för stora siffror). Men, eftersom $np(1-p) = 12.86 \geq 10$, så försöker vi med en normalapproximation. Med andra ord, X är approximativt $N(15, \sqrt{12.86})$-fördelad. Alltså,

$$P(X \geq 25) = 1 - P(X \leq 10) \approx 1 - \Phi\left(\frac{10-15}{\sqrt{12.86}}\right) \approx 0.92.$$

Sannolikheten är alltså ca 92%.

5. (a) Vi har två oberoende s.v. X och Y som är exponentialfördelade. Alltså är $f_X(x) = \exp(-x/5)/5$ för $x \geq 0$, $f_Y(y) = \exp(y/7)/7$ för $y \geq 0$, och sålunda $f_{X,Y}(x,y) = \exp(-x/5 - y/7)/35$ i första kvadranten (dvs $x \geq 0$ och $y \geq 0$). För övrigt är funktionen lika med noll.

(b) Eftersom X och Y är oberoende så är

$$\begin{aligned} P(X > 1, Y < 1) &= P(X > 1)P(Y < 1) \\ &= \left(\int_1^{\infty} \frac{e^{-x/5}}{5}\,dx\right)\left(\int_0^1 \frac{e^{-y/7}}{7}\,dy\right) \\ &= e^{-1/5}(1 - e^{-1/7}) \approx 0.11. \end{aligned}$$

(c) Om maximum av två tal är ≤ 5 så måste båda talen uppfylla detta villkor. Alltså, på samma sätt som förra deluppgiften,

$$\begin{aligned} P(\max\{X,Y\} \leq 5) &= P(X \leq 5, Y \leq 5) = P(X \leq 5)P(Y \leq 5) \\ &= \left(\int_0^5 \frac{e^{-x/5}}{5}\,dx\right)\left(\int_0^5 \frac{e^{-y/7}}{7}\,dy\right) \\ &= (1 - e^{-1})(1 - e^{-5/7}) \approx 0.32. \end{aligned}$$

6. Vi ska visa att $E(\bar{X}) = \mu$ och att $E(S^2) = \sigma^2$. Vi börja med den första likheten:

$$E(S^2) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu.$$

Skattningen $\bar{X}$ är alltså väntevärdesriktigt. Att beräkna $E(S^2)$ är lite bökigare, men inte omöjligt:

$$\begin{aligned} E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) &= \frac{1}{n-1}E\left(\sum_{i=1}^{n} X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\big(E(X_i^2) - 2E(X_i\bar{X}) + E(\bar{X}^2)\big). \end{aligned}$$

Vi vet att $E(\bar{X}) = \mu$ och att $V(\bar{X}) = \sigma^2/n$. Steiners formel säger att $E(Y^2) = V(Y) + E(Y)^2$ för en stokastisk variabel Y, vilket vi kan utnyttja för att skriva $E(X_i^2) = V(X_i) + E(X_i)^2 = \sigma^2 + \mu^2$ samt $E(\bar{X}^2) = \sigma^2/n + \mu^2$. Vidare så ser vi att

$$E(X_i\bar{X}) = E\left(X_i \frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_iX_k)$$

och eftersom $E(X_iX_k) = E(X_i)E(X_k) = \mu^2$ om $i \neq k$ (eftersom dessa variabler är oberoende) och $E(X_i^2) = \sigma^2 + \mu^2$ (då $i = k$) kan vi skriva $E(X_i\bar{X}) = ((n-1)\mu^2 + \sigma^2 + \mu^2)/n = \mu^2 + \sigma^2/n$. Vi återgår det sökta väntevärdet:

$$E(S^2) = \frac{1}{n-1}\sum_{i=1}^{n}\big(\sigma^2 + \mu^2 - 2(\mu^2 + \sigma^2/n) + \sigma^2/n + \mu^2\big) = \frac{n\sigma^2 - n\sigma^2/n}{n-1} = \sigma^2.$$

Alltså är även S^2 en väntevärdesriktig skattning (av σ^2).