LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (92MA31, STN2) 2015-01-13 kl 8–12

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. Motivera svaren på följande frågor noggrant!

   (a) Sture drar på måfå tre kort ur en kortlek. Vad är sannolikheten att inget av korten är av färgen spader? (1p)

   (b) Svea ägnar sig istället åt tärningsspel och kastar två 6-sidiga tärningar. Vad är sannolikheten att summan överstiger 10? (1p)

   (c) Vilka av följande funktioner är täthetsfunktioner? (2p)

      i. $f_1(x) = 0.25\cos x,\ -\pi/2 \leq x \leq \pi/2,$

      ii. $f_2(x) = 0.5\sin x,\ -\pi/2 \leq x \leq \pi/2,$

      iii. $f_3(x) = 0.25\big(|\cos x| + |\sin x|\big),\ -\pi/2 \leq x \leq \pi/2.$

   (d) Visa att $P((A \cup B)^c) = P(A^c \cap B^c)$. (1p)

   (e) Låt $P(A) > 0$ och $P(B) > 0$. Om $P(A \,|\, B) = P(B \,|\, A)$, är A och B oberoende? (1p)

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

   (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. Låt $p(x, y) = cxy(1 + x)$ för $x = 0, 1, 2$ och $y = 0, 1, 2$.

   (a) Bestäm c så att $p(x, y)$ blir sannolikhetsfunktionen för en stokastisk variabel (X, Y). (1p)

   (b) Bestäm de marginella sannolikhetsfunktionerna $p_X(x)$ och $p_Y(y)$. Avgör också om X och Y är oberoende. (3p)

   (c) Beräkna sannolikheten $P(X + Y \leq 2)$. (2p)

Vänd!

4. Hembrännaren Herbert är slarvig och använder ganska dålig utrustning så sannolikheten för att en giftig mängd metanol finns med i en destillering är 5% (vid förtäring av en ”portion”).

   (a) Vid en stor fest serveras det ur 10 olika dunkar. Vad är sannolikheten att högst en av dessa dunkar är förgiftad? (2p)

   (b) Polisen tar Herbert och finner på hans gård 5000 fyllda dunkar. Åklagaren tycker inte att det räcker med olaglig alkoholtillverkning utan vill få Herbert fälld även för metanolinnehållet. Därför skickas 350 dunkar till SKL för analys. Där fann man att 21 dunkar innehöll en tillräckligt hög koncentration metanol för att klassas som giftig vid förtäring av en portion. Åklagaren vill gärna vid rättegången kunna säga att en av tio dunkar innehåller en giftig mängd metanol vid förtäring av en ”portion.” Hjälp åklagaren att (approximativt) hitta den minsta konfidensgrad som gör att detta påstående inte är orimligt. (4p)

5. Låt $f(x) = ax^2 + bx$ för $0 \leq x \leq 1$ och $f(x) = 0$ för övrigt, där $a, b \in \mathbf{R}$, vara en (eventuell) täthetsfunktion för en s.v. X sådan att $E(X) = 1$.

   (a) Bestäm a och b (om möjligt) så att f verkligen är en täthetsfunktion för X. (4p)

   (b) Hitta alla $m \in [0, 1]$ så att $P(X \leq m) = m$. (2p)

6. Låt $n > 20$ vara ett heltal och låt $X_i \sim \text{Bin}\left(n(i+1), \dfrac{1}{10(i+1)}\right)$, $i = 1, 2, \ldots, n$, vara oberoende stokastiska variabler. Definiera den stokastiska variabeln Y genom $Y = \sum_{i=1}^{n} X_i$. Beräkna approximativt sannolikheten $P\left(\dfrac{30Y}{n} > 10 + 3n\right)$. Motivera dina approximationer noggrant! (6p)

Lösningsskisser för matematisk statistik 2015-01-13

1. (a) Gynsamma dragningar innebär att vi ska välja 3 av 39 kort (alla icke-spader kort). Alltså, klassiska definitionen av sannolikhet ger att

$$P(\text{inget spaderkort}) = \frac{\binom{39}{3}}{\binom{52}{3}} = \frac{703}{1700} \approx 0.414.$$

(b) Om summan ska bli större än 10 finns bara möjligheterna $5+6$, $6+5$, samt $6+6$. Alltså tre gynsamma och 36 möjliga. Sannolikheten blir således $3/36 = 1/12 \approx 0.083$.

(c) Endast (iii) är en täthetsfunktion. I (i) så blir arean inte ett och i (ii) så är funktionen negativ (och arean är faktiskt noll).

(d) Faktum är att $(A \cup B)^c = A^c \cap B^c$ så sannolikheten behövs egentligen inte. Mängdlikheten kan ses ur ett Venn-diagram. Denna mängdlikhet är känd som en av deMorgans lagar.

(e) Nej, A och B kan vara beroende. Vi kan se detta genom att undersöka likheten lite närmare:

$$P(A \mid B) = P(B \mid A) \quad \Leftrightarrow \quad \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)} \quad \Leftrightarrow \quad P(A \cap B)\frac{P(A) - P(B)}{P(A)P(B)} = 0$$

så det räcker att $P(A) = P(B)$. Det är alltså inte *nödvändigt* att $P(A \cap B) = P(A)P(B)$.

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

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

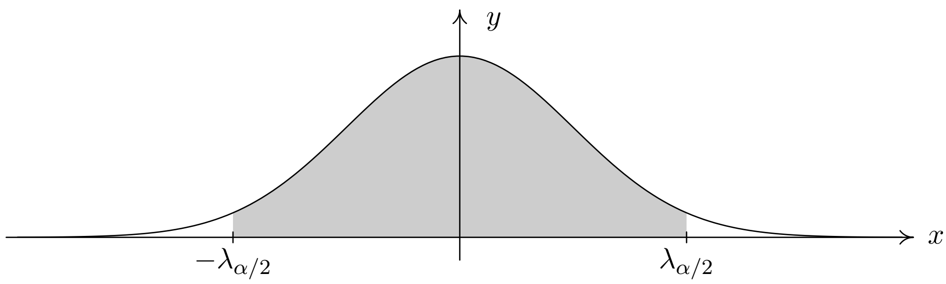
Vi skapar testvariabeln

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

Det följer då att

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

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 (*) ovan och erhåller att

$$\widehat{M} - \frac{2.575 \cdot 2}{\sqrt{10}} < \mu < \widehat{M} + \frac{2.575 \cdot 2}{\sqrt{10}}.$$

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

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

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

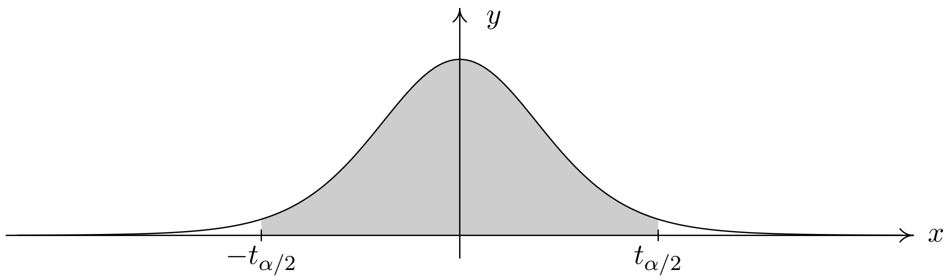
är stickprovsvariansen. Vi skapar testvariabeln

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

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

$$\widehat{M} - \frac{3.25 \cdot 1.444}{\sqrt{10}} < \mu < \widehat{M} + \frac{3.25 \cdot 1.444}{\sqrt{10}}.$$

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

3. (a) Vi kan direkt räkna ut summan av alla $p(x, y)$ då $x = 0, 1, 2$ och $y = 0, 1, 2$:

$$\sum_{x=0}^{2}\sum_{y=0}^{2} p(x, y) = p(0, 0) + p(0, 1) + \cdots + p(2, 2) = 24.$$

Då måste $c = 1/24$.

(b) Vi kan direkt se från definitionen av $p(x, y)$ att det verkar gå att faktorisera som en funktion av bara x och en funktion av bara y så det verkar som X och Y är oberoende. Mer precist, vi räknar ut de marginella sannolikhetsfunktionerna:

$$p_X(x) = 24^{-1}\sum_{y=0}^{2} xy(1 + x) = \frac{x(1 + x)}{8}, \quad x = 0, 1, 2,$$

och

$$p_Y(y) = 24^{-1}\sum_{y=0}^{2} xy(1 + x) = \frac{y}{3}, \quad y = 0, 1, 2.$$

Här ser vi att $p(x, y) = p_X(x)p_Y(y)$ för alla x och y. Alltså är X och Y oberoende.

(c) $P(X+Y\le 2) = 1 - P(X+Y>2) = 1 - p(1,2) - p(2,1) - p(2,2) = 1 - \dfrac{4+6+12}{24} = \dfrac{1}{12}$.

4. (a) 10 dunkar, var och en med sannolikheten 5% att innehålla tillräckligt med metanol. Låt X vara antalet av dessa 10 med tillräcklig metanolhalt. Då är $X \sim \mathrm{Bin}(10, 0.05)$. Alltså blir $P(X \le 1) = P(X=0) + P(X=1) = 0.9139$.

   (b) Eftersom 5000 är så pass stor kan vi approximativt tänka oss att den mängden är oändlig. En naturlig skattning på den verkliga andelen ges av

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

där $X \sim \mathrm{Bin}(n,p)$ är antalet dunkar med giftig mängd metanol, n är antalet dunkar vi undersöker och p är den okända andelen. Vi vet att $n = 350$ och skattar p med $\widehat{p} = 21/350 = 0.06$, så för att vi ska kunna göra en normalapproximation behöver vi

$$np(1-p) \approx 350 \cdot 0.06 \cdot (1-0.06) = 19.74 \ge 10.$$

Här förutsätter vi att p ligger nära skattningen. Vi borde alltså 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

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} \approx 0.0127.$$

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 löser ut p ur olikheten inuti sannolikhetsmåttet, och erhåller då

$$\widehat{P} - \lambda_{\alpha/2} d \le p \le \widehat{P} + \lambda_{\alpha/2} d.$$

Vi vill välja $\lambda_{\alpha/2}$ så att $\widehat{P} + \lambda_{\alpha/2} d > 0.10$ för att kunna säga att ”en på tio” inte är orimligt. Alltså måste $\lambda_{\alpha/2} > (0.10 - 0.06)/d = 3.151$. Eftersom $P(Z \le \lambda_{\alpha/2}) = \Phi(3.151) = 0.9992$ följer det att $\alpha = 2 \cdot (1 - 0.9992) = 0.016$. Den minsta konfidensgraden är alltså 98.4% (approximativt).

5. (a) Vi har två krav: både arean och väntevärdet måste bli ett. Alltså,

$$1 = \int_0^1 (ax^2 + bx)\,dx = \frac{a}{3} + \frac{b}{2} \quad \Leftrightarrow \quad 2a + 3b = 6$$

och

$$1 = \int_0^1 x(ax^2 + bx)\,dx = \frac{a}{4} + \frac{b}{3} \quad \Leftrightarrow \quad 3a + 4b = 12.$$

Vi kan lösa ut $a = 4(1 - b/3)$ ur den andra ekvationen och sätta in det i den första, vilket ger $b = 6$. Därmed är även a bestämd till $a = -4$. En eventuell täthetsfunktion måste alltså ha formen $f(x) = -4x^2 + 6x$ för $0 \le x \le 1$. Det som återstår att undersöka är om $f(x) \ge 0$. Enklast gör vi detta genom faktoriseringen

$$f(x) = 2x(3 - 2x) \ge 0$$

eftersom $3 - 2x \ge 1$ för $0 \le x \le 1$. Detta är alltså en täthetsfunktion för en variabel X med $E(X) = 1$.

(b) Vi försöker lösa ekvationen $P(X \leq m) = m$. Vi ställer upp vänsterledet:

$$P(X \leq m) = \int_0^m (ax^2 + bx)\,dx = \frac{am^3}{3} + \frac{bm^2}{2} = \frac{-4}{3}m^3 + 3m^2.$$

Således måste

$$-4m^3 + 9m^2 = 3m \quad \Leftrightarrow \quad m(4m^2 - 9m + 3) = 0 \quad \Leftrightarrow \quad m = 0 \text{ eller } m = \frac{9 \pm \sqrt{33}}{8}.$$

Endast $m = 0$ och $m = 9/8 - \sqrt{33}/8$ uppfyller kravet att $m \in [0, 1]$ (eftersom $5 < \sqrt{33} < 6$).

6. Vi kan inte direkt normalapproximera X_i (för alla i) eftersom

$$n(i+1)\frac{1}{10(i+1)}\left(1 - \frac{1}{10(i+1)}\right) = \frac{n}{10}\left(1 - \frac{1}{10(i+1)}\right)$$

är ganska liten för små n. Däremot går Poissonapproximation bra eftersom $n(i+1) > 10$ och $1/(10(i+1)) < 0.1$. Alltså är $X_i \overset{\text{appr.}}{\sim} \text{Po}(n/10)$. Eftersom variablerna dessutom är oberoende så ser vi att

$$Y = \sum_{i=1}^{n} X_i \overset{\text{appr.}}{\sim} \text{Po}(n^2/10)$$

eftersom summan av oberoende Poissonfördelade variabler är Poissonfördelad. Eftersom $n^2/10 > 15$ kan vi nu normalapproximera variabeln Y enligt $Y \overset{\text{appr.}}{\sim} \text{N}(n^2/10, n/\sqrt{10})$. Den sökta sannolikheten kan (approximativt) beräknas enligt

$$\begin{aligned} P\left(\frac{30Y}{n} > 10 + 3n\right) &= P(Y > n/3 + n^2/10) = 1 - P(Y \leq n/3 + n^2/10) \\ &= 1 - P\left(\frac{Y - n^2/10}{n/\sqrt{10}} \leq \frac{n/3}{n/\sqrt{10}}\right) = 1 - \Phi(\sqrt{10}/3) = 1 - \Phi(1.05) = 0.15. \end{aligned}$$