TMA321: Tentamen den 30 maj 2016 klockan 08:30-12:30

Tillåtna hjälpmedel: Egenhändigt handskriven formelsamling (4 A4-sidor på 2 blad) samt till skrivningen medföljande tabeller.

Tentamen består av 6 frågor om sammanlagt 24 poäng. (Upp till 5 bonuspoäng från tidigare inlämnade bonusuppgifter kan tillkomma.)

Betygsgränser: För godkänt (betyg 3) krävs minst 12 poäng, för betyg 4 krävs minst 16 poäng, och för betyg 5 krävs minst 20 poäng.

* * *

1. Antag att vi gör tre oberoende kast med en rättvis tärning, och låt S beteckna summan av utfallen i de tre kasten.

(a) Bestäm sannolikheten att minst ett av kasten resulterar i en sexa. (1p)

(b) Bestäm väntevärdet $\mathbf{E}(S)$. (1p)

(c) Bestäm variansen $\mathbf{Var}(S)$. (1p)

(d) Bestäm den betingade sannolikheten $\mathbf{P}(S \geq 17 \,|\, S \geq 16)$. (1p)

2. Ett par av kontinuerliga stokastiska variabler X och Y sägs ha en **centralsymmetrisk** gemensam fördelning om de har en gemensam täthetsfunktion $f_{X,Y}$ sådan att $f_{X,Y}(x,y)$ beror på x och y endast via $x^2 + y^2$ (det vill säga det skall existera en funktion g sådan att, för alla x och y, vi har $f_{X,Y}(x,y) = g(x^2 + y^2)$).

(a) Visa att om X och Y har en centralsymmetrisk gemensam fördelning så gäller att om kovariansen $\mathbf{Cov}(X,Y)$ existerar så är den lika med 0. (2p)

(b) Visa att om X och Y är oberoende och normalfördelade med väntevärde 0 och samma varians σ^2, så är deras gemensamma fördelning centralsymmetrisk. (1p)

(c) Visa, med hjälp av ett konkret motexempel, att två stokastiska variabler X och Y som har en centralsymmetrisk gemensam fördelning inte nödvändigtvis behöver vara oberoende. (1p)

3. Två punkter väljs i kuben $[0,1]^3$, enligt likformig fördelning och oberoende av varandra. Uppskatta sannolikheten p att punkterna hamnar inom euklidiskt avstånd 0.001 från varandra. Att beräkna p *exakt* är troligen för svårt, men ni förväntas härleda övre och undre gränser $p_{\max}$ och $p_{\min}$ som ringar in det verkliga värdet p, så att $p_{\min} \le p \le p_{\max}$. Ju lägre kvot $p_{\max}/p_{\min}$ desto bättre, och för full poäng på uppgiften krävs $p_{\max}/p_{\min} \le 1.01$. (4p)

4. Antag att $(X_1, X_2) = (0.7, 0.3)$ och $(Y_1, Y_2, Y_3) = (0.8, 0.9, 1.3)$ är oberoende stickprov på två normalfördelningar med okända väntevärden μ_X respektive μ_Y, och en och samma okända varians σ^2. Gör ett hypotestest med signifikansnivå 0.01 av nollhypotesen $\mu_X = \mu_Y$ mot alternativhypotesen $\mu_X \neq \mu_Y$. (4p)

5. Låt $X_1, \ldots, X_n$ vara ett stickprov på en stokastisk variabel med okänd fördelning, okänt väntevärde μ och okänd varians σ^2. Låt $\bar{X}$ beteckna stickprovsmedelvärdet $\frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Vad menas med att en skattning av σ^2 är väntevärdesriktig? (1p)

(b) Avgör om $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ är en väntevärdesriktig skattning av σ^2. (1p)

(c) Avgör om $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ är en väntevärdesriktig skattning av σ^2. (1p)

(d) Den som gjort rätt på uppgifterna (b) och (c) har funnit att exakt en av dessa skattningar av σ^2 är väntevärdesriktig. Är detta den *enda* väntevärdesriktiga skattningen av σ^2 baserad på stickprovet $X_1, \ldots, X_n$, eller kan det finnas andra? (1p)

6. Vi kan tänka oss att en linjär trend $y = \hat{\beta}_0 + \hat{\beta}_1 x$ med hjälp av linjär regression har anpassats till punkterna $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, där x_i-värdena anger tidpunkt, och är storleksordnade så att $x_1 < x_2 < \ldots x_n$. Denna uppgift går ut på att visa att det är fullt möjligt att erhålla en positiv skattad trend $\hat{\beta}_1 > 0$, trots att tidsintervallet $[x_1, x_n]$ kan sönderdelas i överlappande delintervall sådana att samtliga enskilda delintervall ger negativ trend.

Konkret är uppgiften att hitta på ett exempel med fyra datapunkter $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$, med $x_1 < x_2 < x_3 < x_4$, sådana att följande tre påståenden blir sanna:

- Linjär regression baserad på $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ ger $\hat{\beta}_1 > 0$,
- linjär regression baserad på $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ ger $\hat{\beta}_1 < 0$,
- linjär regression baserad på $\{(x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ ger $\hat{\beta}_1 < 0$.

(4p)

TMA321: Lösningar till tentamen 30 maj 2016

(1.) X_1, X_2, X_3 oberoende och likf $\{1, 2, \ldots, 6\}$. $S = X_1 + X_2 + X_3$

(a) $P(\text{minst en sexa}) = 1 - P(\text{ingen sexa}) = 1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}$

(b) $E[X_i] = \frac{1}{6}(1+2+\ldots+6) = 3.5$, så att $E[S] = \sum_{i=1}^{3} E[X_i] = 3 \cdot 3.5 = 10.5$

(c) $Var[X_i] = \frac{1}{6}\sum_{i=1}^{n}(i-3{,}5)^2 = \frac{35}{12}$

(d) $S \geq 17$ om $(X_1, X_2, X_3) \in \{\underbrace{(6,6,6), (6,6,5), (6,5,6), (5,6,6)}_{4 \text{ utfall av } 6^3 = 216 \text{ möjliga}}\}$

så att $P(S \geq 17) = \frac{4}{216}$.

$S = 16$ om $(X_1, X_2, X_3) \in \{\underbrace{(6,6,4), (6,4,6), (4,6,6), (6,5,5), (5,6,5), (5,5,6)}_{6 \text{ utfall av } 6^3 \text{ möjliga}}\}$

så att $P(S=16) = \frac{6}{216}$ och $P(S \geq 16) = P(S=16) + P(S \geq 17) = \frac{10}{216}$.

$$P(S \geq 17 \mid S \geq 16) = \frac{P(S \geq 17, S \geq 16)}{P(S \geq 16)} = \frac{P(S \geq 17)}{P(S \geq 16)} = \frac{2}{5}.$$

---

(2) (a) $x^2 + y^2 = x^2 + (-y)^2$, så centralsymmetri $\Rightarrow f_{X,Y} = f_{X,-Y}$

$\Rightarrow Cov[X, Y] = Cov[X, -Y] = -Cov[X, Y]$ så att $Cov[X, Y] = 0$.

(b) $$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} = \underbrace{\frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}}_{\text{beror endast på } x^2+y^2}$$

(c) Låt t.ex. (X, Y) vara likformigt fördelat över enhetscirkeln, så att

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{om } x^2 + y^2 \leq 1 \\ 0 & \text{annars} \end{cases}$$

Om X och Y vore oberoende skulle t.ex. $P(X > 0.8, Y > 0.8) = P(X > 0.8) P(Y > 0.8)$

Men $P(X > 0.8, Y > 0.8) = 0$ trots att $P(X > 0.8) > 0$ och $P(Y > 0.8) > 0$

så oberoende gäller ej.

(3) (X_1, Y_1, Z_1) och (X_2, Y_2, Z_2) oberoende och $\sim$ likf$[0,1]^3$.

$p = P(A)$ där A är händelsen att avståndet mellan (X_1,Y_1,Z_1) och (X_2,Y_2,Z_2) är högst 0.001.

Givet $(X_1, Y_1, Z_1) = (x_1, y_1, z_1)$ inträffar A med sannolikhet

$$\frac{\text{volym}\left([0,1]^3 \cap \text{klot med } r=0.001 \text{ runt } (x_1,y_1,z_1)\right)}{\text{volym}\left([0,1]^3\right)} \leq$$

$$\leq \frac{\text{volym}\left(\text{klot med } r=0.001 \text{ runt } (x_1,y_1,z_1)\right)}{\text{volym}\left([0,1]^3\right)} = \frac{4\pi}{3} 0.001^3$$

så att $p \leq \frac{4\pi}{3} \cdot 10^{-9}$.

Låt B vara händelsen att (X_1, Y_1, Z_1) inte hamnar inom avstånd 0.001 från någon av kubens 6 väggar. Om B inträffar blir olikheten $(*)$ till en likhet, så att

$$P(A) \geq P(A,B) = P(B)P(A|B) = \frac{4\pi}{3} 10^{-9} \cdot P(B) \geq \frac{4\pi}{3} 10^{-9}(1 - 6 \cdot 0.001) = 0.994 \cdot \frac{4\pi}{3} \cdot 10^{-9}$$

Svar: $0.994 \frac{4\pi}{3} 10^{-9} \leq p \leq \frac{4\pi}{3} 10^{-9}$

---

(4) $\bar{X} = \frac{1}{2}\sum_{i=1}^{2} X_i = 0.5 \qquad \bar{Y} = \frac{1}{3}\sum_{i=1}^{3} Y_i = 1.0$

Sammanvägda stickprovsvariansen $S_p^2 = \frac{1}{(2-1)+(3-1)}\left(\sum_{i=1}^{3}(X_i - \bar{X})^2 + \sum_{i=1}^{3}(Y_i - \bar{Y})^2\right) = \frac{0.22}{3}$

Under nollhypotesen ($H_0: \mu_X = \mu_Y$) är (enligt Rice, s. 422)

$$\frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{2}+\frac{1}{3}}} \quad \text{t-fördelad med 3 frihetsgrader.}$$

Vi förkastar H_0 om $\left|\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{2}+\frac{1}{3}}}\right| > \underbrace{t_3(\overbrace{0.005}^{\text{Halva signifikansnivån p.g.a. symmetriskt test.}})}_{= 5.841 \text{ ur t-fördelningstabell}}$

$$= \frac{0.5}{\sqrt{\frac{0.22}{3}}\sqrt{\frac{5}{6}}} \approx 2 \quad \text{(klarar vi utan miniräknare...)}$$

Alltså förkastar vi inte H_0.

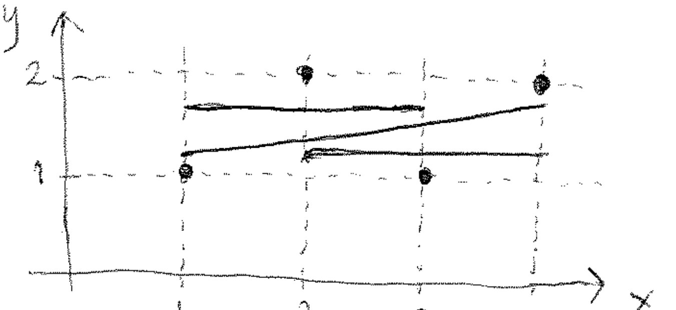
(5.) (a) En skattning A av σ^2 är väntevärdesriktig om $E[A]=\sigma^2$.

(b) $$E[\hat{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n} E[(X_i-\bar{X})^2] = E[(X_1-\bar{X})^2] = Var[X_1-\bar{X}] =$$
$$= Var\left[X_1 - \frac{1}{n}(X_1+\ldots+X_n)\right] = Var\left[\frac{n-1}{n}X_1 + \frac{1}{n}(X_2+\ldots+X_n)\right] =$$
$$= \left(\frac{n-1}{n}\right)^2\sigma^2 + \frac{n-1}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$ så $\hat{\sigma}^2$ <u>ej</u> väntevärdesriktig.

(c) $$E[s^2] = E\left[\frac{n}{n-1}\hat{\sigma}^2\right] = \frac{n}{n-1}E[\hat{\sigma}^2] = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2,$$ så s^2 väntevärdesriktig.

(d) Det finns en mängd väntevärdesriktiga (men mer eller mindre knasiga) skattningar av σ^2. T.ex. följer av (b) att $E[(X_i-\bar{X})^2] = \frac{n-1}{n}\sigma^2$, så om vi fixerar konstanter $C_1,\ldots,C_n$ sådana att $\sum_{i=1}^{n} C_i = \frac{n}{n-1}$ och skattar σ^2 med $A = \sum_{i=1}^{n} C_i(X_i-\bar{X})^2$, så blir
$$E[A] = \sum_{i=1}^{n} C_i E[(X_i-\bar{X})^2] = \frac{(n-1)\sigma^2}{n}\sum_{i=1}^{n} C_i = \sigma^2.$$

(6.)



$(x_1,y_1) = (1,1)$
$(x_2,y_2) = (2,2)$
$(x_3,y_3) = (3,1)$
$(x_4,y_4) = (2,4)$

I föreläsningen den 23 maj såg vi att

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \begin{cases} 0.2 & \text{baserat på alla fyra punkter} \\ 0 & \text{—"— de tre första} \\ 0 & \text{—"— de tre sista} \end{cases}$$

Detta duger inte riktigt, ty vi vill ha <u>strikt</u> negativ lutning i de två sista fallen. Om vi ersätter $y_1=1$ med $y_1=1+\varepsilon$ och $y_2=2$ med $y_2=2+\varepsilon$ för tillräckligt litet $\varepsilon>0$ borde vi få det utan att rubba den första trendlinjen alltför mycket. Det visade sig att $\varepsilon=0.1$ ger

$$\hat{\beta}_1 = \begin{cases} 0.08 & \text{baserat på alla fyra punkter} \\ -0.05 & \text{—"— de tre första} \\ -0.05 & \text{—"— de tre sista} \end{cases}$$