Tentamen i TMA321 Matematisk Statistik, Chalmers Tekniska Högskola.

Hjälpmedel: Valfri räknare, egenhändigt handskriven formelsamling (4 A4-sidor på 2 blad) och till skrivningen medhörande tabeller.
Måndagen den 26 Maj 2014, förmiddag 8.30-12.30.
Examinator och Jour: Olle Nerman, tel. 7723565, rum 3056, MV, Chalmers.
Tentamen består av 8 frågor med sammanlagt 32 poäng+ eventuella bonuspoäng från dina två inlämningsuppgifter (max 4 poäng totalt).
Betygsgränser: För betyget "3" fordras minst 15 poäng, för betyget "4" minst 20 poäng och för betyget "5" minst 25 poäng .

---

1. En stokastisk variabel **X** är normalfördelad med väntevärdesparameter **μ** och standardavvikelseparamaeter $\boldsymbol{\sigma}$. Bestäm båda parametrarna **μ** och $\boldsymbol{\sigma}$ (med god approximation) från vetskapen att sannolikheten **P(X>5)=0.5** och sannolikheten P**(X>10)=0.01** . (3p)

2. En tvådimensionell stokastisk variabel **(X,Y)** har en sannolikhetstäthet **f(x,y)** som är proportionell mot **g** definierad av att:
g(x,y)=xy, för **x** ∈ **[0,1] och y** ∈ **[0,2]**
(g(x,y)= **0** för övrigt.)

   Bestäm:
   a. Konstanten **c** som gör att **f(x,y)=cg(x,y)** sannolikhetstäthet. (1p)
   b. Korrelationen mellan **X** och **Y**. (2p)
   c. Kovariansen mellan **Z=XY** och **X.** (2p)

3. a. Beräkna momentgenererande funktionen **M(t)** för en stokastisk variabel **X** som har utfallsrummet **Ω**x = **{-1,0,2}** och sannolikhetsfunktionen **p(x)=1/3** för alla **x** ∈ **Ω**x. (2p)
   b. Använd resultatet i a-delen för att ange momentgenererande funktionen för aritmetiska medelvärdet av två oberoende stokastiska variabler med samma fördelning som **X** . (2p)

4. Betrakta en Poissonprocess N(t)med intensiteten **3** pulser per tidsenhet och definiera händelserna **A**=att det inte kommer någon puls i intervallet **[0,5]** respektive **B=** att det inte kommer någon puls i intervallet **[0,7]**:
   a. Uttryck **A** och **B** i termer av processen **N(t).** (1p)
   b. Vad är betingade sannolikheten för händelsen **A** givet **B**, **P(A|B)** ? (1p)
   c. Vad är betingade sannolikheten för B givet **A**, **P(B|A)**? (2p)

5. Du misstänker att en viss "Croupier" (han som snurrar roulette-hjulet och sätter fart på kulan) är skicklig på att kasta så att sannolikheten för **0** är <u>lite mindre än den borde</u> (den skall vara **1/37** om han inte fuskar). Därför tänker du med stort tålamod hålla reda på hur många nollor som resulterar från **2000** Roulette-omgångar (hela semestern går nog tyvärr åt). Syftet är att försöka bevisa hans fusk med statistisk hypotesprövning.

VÄND!

a. Vad är den naturliga nollhypotesen för dig i denna situation och vilken fördelning har **X** = antalet nollor bland de **2000** omgångarna under antagande av denna för dig naturliga nollhypotesen? (1p)
b. Hur kan du approximera fördelningen för **X** under nollhypotesen? (1p)
c. Vilka utfall på **X**, stora, små (eller både stora och små), bör du välja som förkastelseområde vid ett formellt hypotestest? (Diskutera gärna) (1p)
d. Använd approximationsidéerna i b. (och resultatet i c) för att bestämma förkastelseområdet vid approximativa signifikansnivån $\boldsymbol{\alpha}$= **0.05** (2p)

6. I ett viss enkelt genetiskt korsningsförsök med **200** upprepningar, så säger teorin att varje enskild avkomma i korsningsförsöket skall ha **0**, **1** eller **2** kopior av en viss genetisk kod, med sannolikheter som är $\mathbf{(1\text{-}p)^2}$, **2p(1-p)** respektive $\mathbf{p^2}$. Här är **p** en okänd sannolikhet som vi vill bestämma. Dessutom gäller att alla försökets **200** ”avkommor” kan antas ha oberoende antal kopior av koden ifråga. Om du använder ett stickprov med antalet kopior $\mathbf{X_i}$ för alla försöken, i=1,2…,200. Vad blir då
   a. Momentskattningen av parametern **p**? (2p)
   b. Är skattninen i a-delen väntevärdesriktig? (motivera) (1p)
7. Du vet att visst slags mynt har vikter som följer en fördelning med väntevärdet 10.55 gram och standardavvikelse 0.2 gram. Du vill packa mynten i påsar om tusen i varje och får den strålande iden att först väga påsen och sedan slänga i mynt ett och ett, utan att räkna dem tills totalvikten överstiger påsens vikt + 10.5 kilo, då du direkt slutar.
   a. Vad är chansen att du får fler mynt än 1000 i påsen (approximativt)? (2p)
   b. Vad är chansen att du får fler mynt än 999 i påsen (approximativt)? (1p)
   c. Använd svaren i a och b delen för att approximera chansen att det blir exakt 1000 mynt i påsen. (1p)
8. I en linjär regressionsmodell av oberoende normalfördelade stokastiska variabler (**Y**) med avseende på inställningsvariabeln (**x**) antas väntevärdena på **Y**-variablerna följa det linjära sambandet **a+bx** för två okända reella parametrar **a** och **b**. Varianserna för Y-variablerna är kända och alla antas vara **=0.25** . Du observerar de tre **y**-variabler till **6.10**, **1.70**, respektive **3.85** vid **x**-inställningarna **10.0** , **4.0**, respektive **7.0** .

   a. Beräkna en observerad punktskattning av regressionslinjens värde vid inställningen **x=5.5** . (1p)
   b. Beräkna ett symmetriskt konfidensintervall för riktningskoeffecienten **b** med konfidensgraden **95 %**. (2p)
   c. Testa på signifikansnivån **5 %** nollhypotesen att **b=0** mot den tvåsidiga alternativa hypotesen att **b≠0.** Vad blir din slutsats? (motivera). (1p)

Lycka Till!

LÖSNINGSFÖRSLAG PÅ MATEMATISK STATISTIK
TMA321, IFysik21, 26 MAJ 2014.
AV OLLE NERMAN

---

1. $P(X>5)=0{,}5$, $X \sim N(\mu,\sigma^2) \Rightarrow 1-\Phi\left(\frac{5-\mu}{\sigma}\right)=\frac{1}{2}$

$\Rightarrow \frac{5-\mu}{\sigma}=0 \Rightarrow \mu=5$,

$P(X>10)=0{,}01 \Rightarrow \Phi\left(\frac{10-5}{\sigma}\right)=0{,}99 \Rightarrow \frac{5}{\sigma}=2{,}33$

$\Rightarrow \sigma=\frac{5}{2{,}33}\approx 2{,}15$

---

2. a. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} c\cdot g(x,y)\,dx\,dy=1 \Rightarrow \int_0^1\left(\int_0^2 c\,x\,y\,dy\right)dx=1$

$\Leftrightarrow 1=c\cdot\left[\frac{x^2}{2}\right]_0^1\cdot\left[\frac{y^2}{2}\right]_0^2=c\cdot\frac{1}{2}\cdot 2=c \quad \therefore \quad c=1$

b. $E[X]=\int_0^1\left(\int_0^2 x\cdot xy\,dy\right)dx=\frac{2}{3}$ $\quad E[Y]=\int_0^1\left(\int_0^2 y\,xy\,dy\right)dx=\frac{4}{3}$

$E[XY]=\int_0^1\left(\int_0^2 xy\cdot xy\,dy\right)dx=\frac{8}{9} \Rightarrow KOV[X,Y]=0 \Rightarrow \rho_{X,Y}=0.$

c. $E[ZX]=E[X^2Y]=\int_0^1\left(\int_0^2 x^2y\cdot xy\,dy\right)dx=\left[\frac{x^4}{4}\right]_0^1\cdot\left[\frac{y^3}{3}\right]_0^2=\frac{8}{12}$

$KOV[Z,X]=E(ZX)-E[Z]\cdot E[X]=\frac{8}{12}-\frac{8}{9}\cdot\frac{2}{3}=\frac{2}{27}$

---

3 a. $M(t)=E[e^{tX}]=\frac{1}{3}(e^{-t}+1+e^{2t})$

b. $M(t)=E\left[e^{t\left(\frac{X_1+X_2}{2}\right)}\right]\underset{\uparrow\,\text{OBER}}{=}E\left[e^{t\frac{X_1}{2}}\right]\cdot E\left[e^{t\frac{X_2}{2}}\right]=\left(M\left(\frac{t}{2}\right)\right)^2$

$=\frac{1}{3}(e^{-\frac{1}{2}t}+1+e^{t})^2$

---

4. a. $\{N(5)=0\}$ RESP $\{N(7)=0\}$ ÄR A RESP B.

b. $B\subseteq A \Rightarrow P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$

c. $B\subseteq A \Rightarrow P(B|A)=\frac{P(B\cap A)}{P(A)}=\frac{P(B)}{P(A)}=\frac{e^{-3\cdot 7}}{e^{-3\cdot 5}}=e^{-6}$

---

5. a. H_0: X = ANTALET NOLLOR $\sim Bin(2000,\frac{1}{37})$

b. C.GR.V.S. $\Rightarrow X \approx N\left(\frac{2000}{37}, 2000\cdot\frac{1}{37}\cdot\frac{36}{37}\right)$

c. FÖRKASTA FÖR SMÅ VÄRDEN PÅ X

d. FÖRKASTA FÖR $X\le C$ DÄR $P_{H_0}(X\le C)\approx 0{,}05$

MED HJÄLP AV b FÅS $C\approx \frac{2000}{37}-1{,}65\sqrt{2000\,\frac{36}{37^2}}$

$\approx 42.$

LÖSNINGSFÖRSLAG PÅ TMA 321, 26 MAJ 2014, FORTS.
OLLE NERMAN

---

6. a. $E[X_1] = 0\cdot(1-p)^2 + 1\cdot 2p(1-p) + 2p^2 = 2p$

EKV. FÖR MOMENTSK. $\hat{p}$: $\quad 2\hat{p} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)$

$$\Rightarrow \hat{p} = \left(\frac{\sum_{i=1}^{200} X_i}{400}\right)$$

b. $$E[\hat{p}] = \frac{\sum_{i=1}^{200} E[X_i]}{400} = \frac{200\cdot 2p}{400} = p \Rightarrow \text{SVARET:}$$

$\hat{p}$ ÄR VÄNTEVÄRDESRIKTIG.

---

7 a. OM $\sum_{i=1}^{1000} X_i < 10500$ SÅ FLER ÄN 1000 MYNT I PÅSEN. MEN $\sum_{i=1}^{1000} X_i \approx N(10550, 1000\cdot 0{,}2^2)$ ENLIGT CENTR. GRÄNSVÄRDESSATSEN $\Rightarrow$

$$P(\text{FLER ÄN 1000 MYNT I PÅSEN}) \approx \Phi\left(\frac{-50}{\sqrt{1000}\cdot 0{,}2}\right) =$$
$$\Phi(-7{,}9) = 1-\Phi(7{,}9) \quad \text{(NÄRA NOLL)}$$

b. PÅ S.S. FÖR 999 MYNT

$$P(\text{FLER ÄN 99 MYNT I PÅSEN}) \approx \Phi\left(\frac{-39{,}45}{\sqrt{999}\cdot 0{,}2}\right)$$
$$= \Phi(-6{,}24) = 1-\Phi(6{,}24) \quad \text{(NÄRA NOLL)}$$

c. $P(\text{EXAKT } 1000) = P(\text{FLER ÄN } 999) - P(\text{FLER ÄN } 1000) \approx$
$\Phi(7{,}9) - \Phi(6{,}24) =$ NÄRA 0.

OBS! SIFFRORNA FELVALDA AV MIG! MEDVETET PÅ SKRIVNINGEN!

---

8. a. $\bar{x} = 7$, $\bar{y} = 3{,}55$, $\hat{b} = \frac{\sum_{i=1}^{3}(x_i-\bar{x})\,y_i}{\sum_{i=1}^{3}(x_i-\bar{x})^2} = \frac{3\cdot 6{,}1 - 3\cdot 1{,}7}{18} \approx$

$\approx 0{,}733 \qquad \hat{a} = \bar{y} - \hat{b}\bar{x} \approx -1{,}25$

$a+b\,5{,}5$ SKATTAS AV $\hat{a}+\hat{b}\,5{,}5 = \bar{y} - 1{,}5\,\hat{b}$

$= 3{,}55 - 1{,}5\cdot 0{,}733 \approx 3{,}45$

b. $\hat{b} \sim N\left(b;\, 0{,}25\cdot\frac{1}{\sum_i (x_i-\bar{x})^2}\right) = N\left(b;\, \frac{1}{72}\right)$ GER

$b = \hat{b} \pm 1{,}96\cdot\sqrt{\frac{1}{72}}$ (95%) OBSERVERAT:

$b = 0{,}733 \pm 0{,}231$ (95%)

c. KORRESPONDENS MELLAN TEST OCH KONF. INTERVALL $\Rightarrow$
$0 \notin 0{,}733 \pm 0{,}231$ (95%) $\Rightarrow H_0: b=0$ FÖRKASTAS PÅ SIGNIFIKANSNIVÅN 5%