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.

Onsdagen den 22 Maj 2012, 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.

Betygsgränser: För betyget ”3” fordras minst 15 poäng, för betyget ”4” minst 20 poäng och för betyget ”5” minst 25 poäng .

---

1. Du kastar två vanliga sex-sidiga tärningar. Vad är sannolikheterna att
   a. summan av poängen blir exakt **= 8**? (1p)
   b. summan av poängen blir minst **8**? (1p)
   c. Produkten av de två poängtalen blir exakt **=12**? (1p)

   Låt nu summan av poängtalen av de två kasten vara den stokastiska variabeln **Y**.
   d. Bestäm väntevärdet och variansen för **Y**. (2p)

2. Några oberoende-upgifter:
   a. Tre personer a, b och c ställer sig på måfå (i slumpmässig ordning) i en kö. Är händelserna **D** = ”a står före b” och **E** = ”a står före c” oberoende av varandra? Motivera. (1p)
   b. Om tre händelser **A**, **B** och **C** är oberoende av varandra så är också händelserna **A \ B** och **C** oberoende av varandra. Kan du motivera det? (2p)
   c. Om den stokastiska variabeln **X** och den stokastiska variabeln **Z=X+Y** är oberoende av varandra, så följer det typiskt att **Y** **inte** är oberoende av **X**. Undantaget är om **X** med sannolikheten **1** är en lika med en konstant:
   Antag att alla inblandade variablerna har ändliga varianser och försök motivera resultatet genom att studera kovarianser på lämpligt sätt. (2p)

3. Diametern på en viss axel och innerdiametern på ett tillhörande kullager tillverkas med mått så att de skall ”passa” i varandra. Du vet att båda är normalfördelade med väntevärden **15.000 mm**. respektive **14.995 mm**. och standardavvikelser **0.0012 mm**. respektive **0.0007 mm**. Vad är då sannolikheten att en slumpmässigt vald axel passar i kullagret med en diameterskillnad som är mellan **0.00 mm**. och **0.020** mm.? (3p)

4. Ett nytt studenthem för **100** studenter skall byggas och man planerar att förse intresserade studenter med varsin jordlott för odling av potatis, morötter och andra grönsaker. Man tänker sig tre olika storlekar på odlingslotter **100**, **150** och **200** kvadratmeter. Noggranna preliminära undersökningar tyder på att ungefär **15%** av

”typiska studenter” skulle vilja hyra en odlingslott (om de bestämde sig att flytta till det nya hemmet). Av de som vill hyra säger sig **73%** vilja ha den minsta storleken **100** kvadratmeter , **15%** mellanstorleken **150** kvadratmeter och **12%** den största storleken på **200** kvadratmeter.

a. Hur stor total odlingsyta bör reserveras för att man med **99%** sannolikhet skall kunna uppfylla alla hyresönskemålen för jordlotter för samtliga studenter som flyttar in? (Använd centrala gränsvärdessatsen.) (3p)

b. Hyser du några tveksamheter angående realismen i uppskattningen i a? Diskutera eventuella svagheter i analysen och förutsättningarna för centrala gränsvärdessatsen? (1p)

5. I en linjär regression med svarsvariabler **y** och inställningsvariabler **x** uppskattas blev medelvärdet av **y**-variablerna **15,3** och medelvärdet av **x**-variablerna var **0.32**. Skattningen av riktningskoeffecieten **β** blev **0.22**.

a. Beräkna en observerad punktskattning av interceptet **α.** (1p)
b. Beräkna en punktskattning av väntevärdet vid **x**-inställningen **3.5** . (1p)

6. 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 lite större än den borde (den skall vara **1/37** om han inte fuskar. Därför tänker du med stort tålamod hålla reda på hur många nollor som resulterat från **1500** Roulette-omgångar (hela semestern går nog tyvärr åt). Syftet är att försöka bevisa hans fusk med statistisk hypotesprövning.

a. Vad är den naturliga nollhypotesen för dig i denna situation och vilken fördelning har **X** = antalet nollor bland de **1500** 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 approximationsideerna i b. (och resultatet i c) för att bestämma förkastelseområdet vid signifikansnivån **α= 0.01** (2p)

7 . Stickprovsvariansen är som bekant en väntevärdesriktig punktskattning av den teoretiska variansen. Detta medför att längden i kvadrat på konfidensintervall beräknade med t-fördelningsmetoden enkelt kan jämföras med längden i kvadrat på motsvarande intervall när man känner den teoretiska variansen. Hur ser kvoten ut mellan väntevärdet av längden i kvadrat på ett enstickprovsbaserat symmetriskt t-intervall vid konfidensgrad **95%** vid stickprovsstorlek **10** och (den deterministiska) längdkvadraten för motsvarande normalfördelningsbaserade intervall när teoretiska variansen är känd? (3p)

8. 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 $\mathbf{(1-p)^2}$, $\mathbf{2p(1-p)}$ 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.

a. Vad är likelihoodfuntionen **L**, för hela försöket? (1p)

b. Hur ser en ML-skattning av **p** ut? (2p)

c. Hur ser en observerad ML-skattning ut om **107** av avkommorna har **0** kopior, **63** har 1 kopia och **30** har **2** kopior? (1p)

d. Är skattningen av **p** i b-uppgiften väntevärdesriktig? (1p)

22/5 2012 TMA 321 F-KF

LÖSNINGSFÖRSLAG AV OLLE NERMAN MV, CHALMERS

---

1 a) SUMMAN 8 FÅS FÖR UTFALLEN $\{(2,6),(3,5),(4,4),(5,3),(6,2)\} \Rightarrow$. ANTALET GYNNSAMMA UTFALL = 5, TOTALA ANTALET UTFALL = 36 SÅ DÄRFÖR ÄR SÖKT SANNOLIKHET = 5/36

b) MINST SUMMAN 8 FÅS PÅ 15 OLIKA SÄTT $\left((5+4+3+2+1)=15\right) \Rightarrow$ SÖKT SANNOLIKHET = 15/36

c) PRODUKTEN 12 FÅS PÅ 4 OLIKA SÄTT ((2,6),(6,2),(3,4),(4,3)) $\Rightarrow$ SÖKT SANNOLIKHET = 4/36 = 1/9

d) VÄNTEVÄRDE VID KAST AV EN TÄRNING = 3,5 $\left(= \frac{1}{6}(1+2+3+4+5+6)\right)$
VARIANS VID KAST AV EN TÄRNING =
$\frac{1}{6}(1^2+2^2+\ldots+6^2) - 3{,}5^2 = \frac{35}{12}$ $\Rightarrow$ TILLSAMMANS MED OBEROENDET ATT E["SUMMAN"] = 3,5 + 3,5 = 7 OCH VARIANSEN FÖR SUMMAN = $\frac{35}{12}+\frac{35}{12} = \frac{35}{6}$

---

2 a) P.G.A SYMMETRI $P(D)=P(E)=1/2$, I TVÅ FALL AV 6 INTRÄFFAR $D \cap E$ SÅ ATT $P(D\cap E) = \frac{1}{3} \neq \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = P(D)P(E)$
$\therefore$ D OCH E ÄR INTE OBEROENDE

---

b) $P(A\setminus B) = P(A) - P(A\cap B) \underset{\text{OBER.}}{=} P(A) - P(A)P(B)$ OCH

$P((A\setminus B)\cap C) = P((A\cap C)\setminus(B\cap C)) = P(A\cap C) - P(A\cap B\cap C)$
$\underset{\text{OBER}}{=} P(A)\cdot P(C) - P(A)P(B)P(C) = (P(A)-P(A)P(B))\cdot P(C) = P(A\setminus B)\cdot P(C)$

c) $KOV[X,Y] = KOV[X, Z-X] = KOV[X,Z] - KOV[X,X] =$
$0 - VAR[X]$ EFTERSOM $KOV[X,Z]=0$, PÅ X OCH Z OBER.
$\Rightarrow$ ATT OM $VAR[X]>0$ SÅ ÄR X OCH Y INTE OBER.

---

3 X = AXELDIAM. $\sim N(15, 0{,}0012^2)$ OCH $Y \sim N(14{,}995, 0{,}0007^2)$
(DÄR Y = KULLAGERDIAM.) OBEROENDE $\Rightarrow$ $Y - X \sim N(-0{,}005, 0{,}0007^2+0{,}0012^2)$
$P(0 \le Y-X \le 0{,}020) = \Phi\left(\frac{0{,}020+0{,}005}{\sqrt{0{,}0012^2+0{,}0007^2}}\right) - \Phi\left(\frac{0+0{,}005}{\sqrt{0{,}0012^2+0{,}0007^2}}\right)$

$\approx 1-\Phi(3{,}59)$ UTANFÖR TABELL! $\approx$ ca. 0,1 ‰ DVS. PRAKTISKT TAGET = 0. (VID VÄNDNING AV DIAMETER-VÄNTEVÄRDENA SÅ FÅS $\approx 1-\Phi(-3{,}59) \approx 0{,}9999$)

22/5 2012 TMA 321 I-KT
FÖRBÄTTRING LÖSNINGSFÖRSLAG AV O.N. / MV CHALMERS

4. LÅT X_i VARA EFTERFRÅGAD ODLINGSAREA AV STUDENT i.
DÅ BLIR

$$P(X_i=0)=0{,}85$$
$$P(X_i=100)=0{,}15\cdot 0{,}73=0{,}1095$$
$$P(X_i=150)=0{,}15\cdot 0{,}15=0{,}0225$$
$$P(X_i=200)=0{,}15\cdot 0{,}12=0{,}018$$

HÄRAV FÖLJER ATT

$$E[X_i]=0\cdot 0{,}85+100\cdot 0{,}1095+150\cdot 0{,}0225+200\cdot 0{,}018=17{,}925$$

$$VAR(X_i)=E[X_i^2]-E[X_i]^2\approx 44{,}7^2, \text{ TY } E[X_i^2]=100^2\cdot 0{,}1095+150^2\cdot 0{,}0225+200^2\cdot 0{,}018\approx$$

a) $\sum_{i=1}^{100} X_i \approx N(1792{,}5\,;\,447^2)$ PÅ GRUND AV C.GR.V. SATSEN.
VI SKALL VÄLJA AREAL a SÅ ATT

$$P\left(\sum_{i=1}^{100} X_i\le a\right)\approx 0{,}99 \Rightarrow \Phi\left(\frac{a-1792{,}5}{447}\right)=0{,}99$$
$$\Rightarrow a=1792{,}5+2{,}33\cdot 447\approx 2834 \quad (2850)$$

b) 1) VI ÄR I EN SVANS PÅ FÖRDELNINGEN OCH DEN ÄR SKEV FRÅN BÖRJAN...
2) OBEROENDEANTAGANDET KAN VARA OPTIMISTISKT EFTERSOM STUDENTERNA KANSKE BESTÄMMER SIG TILLSAMMANS...
3) DET KAN VARA OSÄKERHET I SANNOLIKHETSSKATTNINGARNA SOM VI UTGÅR FRÅN SOM VI INTE BRYR OSS OM.

5. a) $\hat{\alpha}=\bar{y}-\bar{x}\cdot\hat{\beta}=15{,}3-0{,}32\cdot 0{,}22=15{,}2206$
b) $\hat{\alpha}+\hat{\beta}\cdot 3{,}5=15{,}2206+0{,}22\cdot 3{,}5=15{,}9906$

6. a) $X\sim$ BINOMIALF.$(1500, p)$, DÄR $H_0: p=\frac{1}{37}$ NATURLIG NOLLHYPOTES (EV.: $p\le\frac{1}{37}$ KAN ANVÄNDAS SOM H_0)
b) $X\approx N\left(\frac{1500}{37}, \frac{1500}{37}\cdot\frac{36}{37}\right)$ UNDER H_0
c) MAN BÖR FÖRKASTA H_0 FÖR STÖRRE VÄRDEN PÅ X EFTERSOM MISSTANKTA "FUSKET" SVARAR MOT SÅDANA.
d) $P(X\ge c\,|\,H_0)\approx 0{,}01 \Rightarrow c=\frac{1500}{37}+2{,}33\sqrt{\frac{1500}{37}\cdot\frac{36}{37}}\approx \cancel{40{,}54}$ 55
ALLTSÅ SKALL VI FÖRKASTA FÖR $X\ge \cancel{41}$, 55

7. FÖRVÄNTAD LÄNGDKVADRAT PÅ t-INTERVALLET $=\frac{2^2\cdot 6^2\cdot 2{,}262^2}{10}$
LÄNGD PÅ NORMALF. BASERAT INTERVALL
$=\frac{2^2\cdot 6^2\cdot 1{,}96^2}{10}$; SVAR $\frac{2{,}262^2}{1{,}96^2}\approx 1{,}332$

8. OM X_i BETECKNAR UTFALLET 0, 1 ELLER 2 OCH Y_0 = ANTALET X_i MED UTFALLET 0, Y_1 ANTALET $X_i=1$ OCH Y_2 = ANTALET $X_i=2$ SÅ
a) $L(p)=((1-p)^2)^{y_0}(2p(1-p))^{y_1}(p^2)^{y_2}$
b) $\frac{d\ln L(p)}{dp}=0 \Rightarrow \hat{p}=\frac{y_1+2y_2}{400}$
c) $\frac{63+2\cdot 36}{400}\approx 0{,}375$
d) JA, TY $E[\hat{P}]=(200\cdot 2\cdot p(1-p)+200\cdot 2p^2)/400=p$