Tentamen i TMA321 Matematisk Statistik, Chalmers Tekniska Högskola.

Tid: Onsdagen den 26 Augusti 2015, 14.00-18.00.
Examinator och Jour: Olle Nerman, tel. 7723565, rum 3056, MV, Chalmers.
Hjälpmedel: Valfri räknare, egenhändigt handskriven formelsamling (4 A4-sidor på 2 blad) och till skrivningen medföljande tabeller.

Tentamen består av 8 frågor med sammanlagt 32 poäng.
(Eventuella bonuspoäng från två inlämningsuppgifter tillkommer med maximalt 2+2 = 4 poäng)
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 **6**-sidiga tärningar och beräknar **X=**summan av tärningskasten.
   - **a.** Vad är betingade sannolikheten för **{X=7}** givet att de två tärningarna fått utfall som **inte** är lika med varandra? (1p)
   - **b.** Vad är betingade sannolikheten för **A=** händelsen att minst en av tärningarna visar en etta, givet att **{X=7}**? (1p)
   - **c.** Är händelserna **A (**definierad som i b-delen) och **{X=7}** oberoende av varandra? (motivera svaret) (2p)

2. För en viss exponentialfördelad stokastisk variabel **X** är sannolikheten att få ett utfall i intervallet **[0,2]** lika med **0.7**. Bestäm median, väntevärde och varians för **X**. (3p)
3. Persikor som slumpmässigt väljs från ett stort parti väger i genomsnitt (väntevärdet) **135** gram, och vikten varierar enligt en fördelning med en varians som är **150** (gramkvadrat).
   a. Du antar (approximerar med) att successivt valda persikor har en vikt som är oberoende av varandra och följer en fördelning med dessa parametrar. Vad är (approximativt) sannolikheten att de **16** först valda persikorna tillsammans väger mindre än **2** Kg? (2p)
   b. Hur många persikor behöver du minst ta för att sannolikheten för att de tillsammans skall väga minst **5** Kg, skall vara större än **0.9 ?** (approximera på lämpligt sätt) (2p)
4. Du observerar **Y=**antalet händelsetidpunkter i en Poissonprocess med intesiteten c /minut under en halvtimme.
   - **a.** Vilken fördelning har **Y**? (1p)
   - **b.** Vad är **p(10)=**sannolikheten uttryckt i **c** för att det **inte** finns någon händelsetidpunkt i ett kommande **10**-minutersintervall? (1p)
   - **c.** Punktskatta **p(10),** definierad som i b-delen, på lämpligt sätt med hjälp av observationen att **y=5**. (2p)
5. Den empiriska fördelningsfunktionen för ett stickprov av **n=400** kontinuerligt likafördelade oberoende stokastiska variabler observerades i punkten **x=7,3** till värdet **0,34**.
   - **a.** Vad betyder utsagan ovan (d.v.s. vad har man gjort för att räkna ut siffran **0,34**)? (1p)
   - **b.** Beräkna på lämpligt sätt ett uppått begränsat observerat konfidensintervall med approximativ konfidensgrad **99%** för fördelningsfunktionens värde i punkten **x=7,3**. (2p)

6. 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**. Varianserna för samtliga **Y**-variablerna antas alla vara **4**.

   a. Beräkna maximum likelihoodskattingen av riktningskoeffecienten **b** om du har observerat de fyra **y**-observationerna: **1.5, 3.8, 6.1** och **10.0**, vid **x**-inställningarna **-1**, **2**, **5** respektive **8**. (2p)

   b. Beräkna ett symmetriskt prediktionsintervall för en ny oberoende observation med konfidensgraden 5% vid inställningen $\mathbf{x_0=12}$. (2p)

7. En forskare du känner har två gånger genomfört en datainsamling av **n=7** respektive **m=15** observationer som alla kan anses vara observationer av normalfördelade stokastiska variabler med samma okända varianser men med olika okända väntevärden i båda stickproven. Han har sedan räknat ut ett t-intervall för väntevärdet μ_1 i det första stickprovet som blev:

   $$\mu_1 \in [10.3 , 14.7] \quad (99\%) .$$

   Han hade också räknat ut ett t-intervall för väntevärdet μ_2 i det andra stickprovet och resultatet blev:

   $$\mu_2 \in [8.4 , 11.6] \quad (99\%) .$$

   a. Nu skulle han vilja se hur ett observerat konfidensintervall för differensen $\Delta=\mu_2 - \mu_1$ mellan de båda väntevärdena med konfidensgraden **95%** baserat på alla **22** observationerna ser ut? Men han har tappat bort de båda stickprovens observationsvärden och inte heller sparat deras aritmetiska medelvärden och stickprovs-standardavvikelser. Han frågar därför dig (som har rykte om dig att kunna räkna bra) om du kan hjälpa honom på något sätt? Kan du det? (3p)

   b. Eftersom du var så duktig på att hjälpa honom i a-uppgiften så undrar han också om du kan hjälpa honom att rekonstruera hur slutsatsen i testet skulle blivit om du använt ett tvåsidigt t-test, med signifikansnivån **5%,** för att testa nollhypotesen att de båda väntevärdena är lika, dvs $\Delta=0$. Med hjälp av enkel titt på resultatet i a svarar du snabbt…..? Ja vad blir svaret och vad var det för regel du använde? (2p)

8. I ett viss enkelt genetiskt korsningsförsök med **100** 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 **100** ”avkommor” kan antas ha oberoende antal kopior av koden ifråga. Du skall genomföra ett försök där du för var och en av de 100 avkommorna observerar kopieantalet av den nämnda genetiska koden.

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

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

c. Är skattningen i b väntevärdesriktig? (2p)

Lycka till!

LÖSNINGSFÖRSLAG TMA 321, MATEMATISK STATISTIK, CHALMERS
26 AUGUSTI 2015. OLLE NERMAN

1 a. UTFALLSRUMMET $\Omega = \{(1,1), \ldots, (6,6)\}$ HAR $6 \cdot 6 = 36$ ELEMENT.
OM B = HÄNDELSEN ATT TÄRNINGARNA ÄR OLIKA SÅ HAR B 30 ELEMENT. KLASSISK SANNOLIKHETSDEFINITION GER $P(B) = \frac{30}{36} = \frac{5}{6}$. HÄNDELSEN $\{X=7\} \cap A = \{X=7\}$ HAR 6 ELEMENT $((1,6),(2,5), \ldots (6,1))$. DETTA GER

$$P(\{X=7\}|A) = \frac{6/36}{30/36} = \frac{1}{5}$$

b. $A \cap \{X=7\} = \{(1,6),(6,1)\}$ SÅ $P(A \cap \{X=7\}) = \frac{2}{36}$
OCH $P(X=7) = \frac{6}{36} = \frac{1}{6} \Rightarrow P(A|\{X=7\}) = \frac{2/36}{1/6} = \frac{1}{3}$

c. $A = \{(1,1),(1,2),(1,3), \ldots (2,1),(3,1) \ldots (6,1)\}$, SÅ ATT ANTALET ELEMENT I A ÄR 11 OCH $P(A) = \frac{11}{36}$. EFTERSOM $P(A|\{X=7\}) = \frac{1}{3} \neq \frac{11}{36} = P(A)$ SÅ ÄR A OCH $\{X=7\}$ BEROENDE HÄNDELSER

2. $P(X \in [0,2]) = P(X \leq 2) = 1 - e^{-\lambda 2} = 0{,}7 \Rightarrow \lambda = \frac{-\ln 0{,}3}{2} \approx 0{,}602$
$E[X] = \frac{1}{\lambda} = \frac{1}{0{,}602} \approx 1{,}66$ $\quad VAR[X] = \frac{1}{\lambda^2} \approx 2{,}76$ OCH
$1 - e^{-\lambda m} = \frac{1}{2}$ GER MEDIANEN $m = \frac{\ln 2}{\lambda} \approx 1{,}15$

3. a. T = TOTALVIKTEN $\approx$ NORMALFÖRDELAD $(16 \cdot 135, 16 \cdot 150)$
$= N(2160, 2400)$ [SKALA GRAM] $\Rightarrow$
$$P(T \leq 2000) \approx \Phi\left(\frac{2000-2160}{\sqrt{2400}}\right) = \Phi(-3{,}26) = 1 - \Phi(3{,}26) \approx 0{,}0006$$
(VI ANVÄNDE C. GR.V. SATSEN)

b. OM T_n ÄR TOTALVIKTEN OM VI VÄLJER n PERSIKOR OCH VI ANVÄNDER CENTRALA GRÄNSVÄRDESSATSEN FÅS VILLKORET (APPR.)
$$P(T_n \geq 5000) \approx 1 - \Phi\left(\frac{5000 - n \cdot 135}{\sqrt{n \cdot 150}}\right) \geq 0{,}9$$
SOM ÄR UPPFYLLT OM
$$\frac{5000 - n \cdot 135}{\sqrt{n \cdot 150}} \leq -1{,}65$$
MED ANDRAGRADSEKVATION FÖR $\sqrt{n}$ FÅS EN APPROXIMATIV LÖSNING MELLAN $n=37$ OCH 38
$\therefore$ $n = 38$ RÄCKER.

4. a. Y ÄR POISSONFÖRDELAD ($\lambda = 30c$)

b. SOM DEFINIERAD, ÄR $p(10) = \frac{(10c)^0 e^{-10c}}{0!} = e^{-10c}$

c. LÄMPLIG SKATTNING AV c ÄR $\hat{c} = \frac{Y}{30}$, OBSERV. $\hat{c} = \frac{5}{30}$.
DÄRFÖR ÄR LÄMPLIG SKATTNING AV $p(10)$ $e^{-10\hat{c}} =$
$= e^{-\frac{5}{3}} \approx 0,19$

5. a. 0,34 ÄR ANTALET OBSERVATIONER BLAND DE 400 SOM ÄR $\leq 7,3$ DIVIDERAT MED 400.

b. OM $\hat{F}(x)$ ÄR EMPIRISKA FÖRDELNINGSFUNKTIONEN I x SÅ ÄR $400 \cdot \hat{F}(x) \sim$ BINOMIALFÖRDELAD $(400, F(x))$
VÄLJ $x = 7,3$ OCH ANVÄND BINOMIALINTERVALL BASERAT PÅ CENTRALA GRÄNSVÄRDESSATSEN:

$$F(7,3) \leq \hat{F}(7,3) + 2,33\sqrt{\frac{\hat{F}(7,3)(1-\hat{F}(7,3))}{400}} \quad (\approx 99\%)$$

OBSERVERAT

$$F(7,3) \leq 0,34 + 0,05 = 0,39 \quad (\approx 99\%)$$

6. a. $\hat{b} = \frac{\sum_{i=1}^{4}(x_i - \bar{x})y_i}{\sum_{i=1}^{4}(x_i-\bar{x})^2} = \frac{41,9}{45} \approx 0,931$

b. (BLEV LITE RÄKNEKRÅNGLIG) Y_0 NY OBSERVATION VID INSTÄLLNING $x_0 = 12$. DÅ ÄR

$$Y_0 - (\hat{a} + \hat{b} \cdot 12) \sim N\left(0, 4 \cdot \left(1 + \frac{1}{4} + \frac{(12-3,5)^2}{45}\right)\right)$$
$$= N(0, 11,42)$$

GER

$$Y_0 \leq \hat{a} + 12\hat{b} + 2,33\sqrt{11,42} = \hat{a} + 12\hat{b} + 7,87 \quad (99\%)$$

OBSERVERAT $\hat{a} + 12\hat{b} = \bar{y} + 8,5 \cdot 0,931 \approx 13,26$

SLUTRESULTATET ÄR ETT OBSERVERAT PREDIKTIONSINTERVALL

$$Y_0 \leq 21,13 \quad (99\%)$$

7. a. VI BEHÖVER FORMLERNA FÖR ENSTICKPROVS OCH TVÅSTICKPROVS t-INTERVALL (FÖR VÄNTEVÄRDEN RESP. VÄNTEVÄRDESDIFFERENSER.

$$\mu = \bar{x} \pm c \cdot \frac{s_x}{\sqrt{n}} \quad (99\%)$$

DÄR $F_{t(n-1)}(c) = 0{,}995$

OCH

$$\mu_Y - \mu_X = \bar{y} - \bar{x} \pm c' \cdot s_p \sqrt{\frac{1}{m} + \frac{1}{n}} \quad (95\%)$$

DÄR $F_{t(m+n-2)}(c') = 0{,}975$ OCH $s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(m+n-2)}$

FÖRST OM $n=7 \Rightarrow c = 3{,}707$ GER ATT

$\bar{x} = 12{,}5$ OCH $s_x = \frac{\sqrt{7}\, 2{,}2}{3{,}707} \approx 1{,}57$.

PÅ SAMMA SÄTT FÅS NÄR STICKPROVSSTORLEKEN 15 ATT $F_{t(14)}(c) = 0{,}995 \Rightarrow c = 2{,}977$ VILKET GER ATT $\bar{y} = 10$ OCH $s_y = \frac{\sqrt{15} \cdot 1{,}6}{2{,}977} \approx 4{,}33$

UR DETTA FÅS $\bar{y} - \bar{x} = -2{,}5$ OCH $s_p \approx 1{,}94$ VILKET GER OBSERVERADE INTERVALLET

$$\mu_Y - \mu_X = -2{,}5 \pm \underbrace{2{,}086 \cdot 1{,}94 \cdot 0{,}477}_{\approx 1{,}85} \quad (95\%)$$

b. VI ANVÄNDER KORRESPONDENSMETODEN MELLAN TEST OCH KONFIDENSINTERVALL. H_0 $\mu_Y - \mu_X = 0$ FÖRKASTAS ($\alpha = 0{,}05$) OM KONFIDENSINTERVALLET MED KONFIDENSGRAD 95% INTE INNEHÅLLER 0

$\therefore$ VI FÖRKASTAR H_0

8 a. OM ANTALET INDIVIDER MED i KOPIOR BETECKNAS x_i SÅ ÄR LIKELIHOODFUNKTIONEN

$$L(p) = (1-p)^{2x_0} \cdot (2p(1-p))^{x_1} \cdot p^{2x_2}$$

(OM VI HÅLLER ORDNING PÅ ANTALET KOPIOR FÖR VARJE ENSKILD INDIVID.

b. $\frac{d \ln L(p)}{dp} = \frac{-2x_0 + x_1}{1-p} + \frac{x_1 + 2x_2}{p} = 0 \Rightarrow \hat{p} = \frac{x_1 + 2x_2}{2(x_0 + x_1 + x_2)} = \frac{x_1 + 2x_2}{200}$

c. $E[X_1] = 100\, 2p(1-p)$ OCH $E[X_2] = 100p^2 \Rightarrow$

$E\left[\frac{X_1 + 2X_2}{200}\right] = 200(p - p^2) + 200p^2 = p$ SVAR: JA!