Tentamen i TMA321 Matematisk Statistik, Chalmers Tekniska Högskola.
Tid: Lördagen den 18 April 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 medhörande tabeller.
Tentamen består av 8 frågor med sammanlagt 32 poäng.
Betygsgränser: För betyget "3" fordras minst 15 poäng, för betyget "4" minst 20 poäng och för betyget "5" minst 25 poäng .

---

1. Sannolikheten för att en kund som kommer till en viss kiosk i centrala Borås en typisk fredagskväll mellan **22.00** och **23.00** skall köpa cigaretter är ungefär **20%**. Erfarenhetsmässsigt kommer cigarettkunderna vid denna tid också att köpa en kvällstidning med den betingade sannolikheten **30%**. Denna andel är lite högre än genomsnittet kvällstidningsköpare **18%** bland samtliga kunder i samma tidsperioder.
   a. Vad är betingade sannolikheten att en slumpvald kund (i denna tidsperiod) som inte köper cigaretter handlar en kvällstidning? (2p)
   b. Vad är betingade sannolikheten för att en kund (i denna tidsperiod) som köper en kvällstidning också handlar cigaretter? (2p)

2. Logaritmen av en positiv stokastisk variabel **X** är normalfördelad med väntevärdesparameter $\boldsymbol{\mu}$ och standardavvikelseparamaeter $\boldsymbol{\sigma}$. Bestäm båda parametrarna $\boldsymbol{\mu}$ och $\boldsymbol{\sigma}$ (med god approximation) från vetskapen att sannolikheten **P(X<1)=0.7** och sannolikheten P**(X>2)=0.01** . (3p)

3. I en Poissonprocess med okänd intensitet **c** pulser/tidsenhet observeras antalet "pulser" **X** i ett tidsintervall av längden **15**. Då är $\hat{\mathbf{c}} = \mathbf{X/15}$ både maximum likelihood-skattning och en momentskattning av intensitetsparametern **c**.
   a. Är $\hat{\mathbf{c}}$ väntevärdesriktig (unbiassed)? Motivera svaret! (1p)
   b. Vilket standardfel (standard error) har $\hat{\mathbf{c}}$? (2p)
   c. Ange en lämplig observerad skattning av sannolikheten att du i ett visst tidsintervall, som är disjunkt från det observerade intervallet och som har längden **4** tidsenheter, inte observerar någon "puls" alls om du observerat att **x=7** ? (2p)

4. I ett normalfördelningsstickprov med **8** observationer har någon beräknat stickprovsmedelvärdet och stickprovsstandardavvikelsen till **2.45** och **0.16**.
   a. Du ombeds att förvandla informationen till ett uppåt begränsat konfidensintervall för det bakomliggande väntevärdet $\boldsymbol{\mu}$ för de enskilda observationerna med konfidensgraden **95%**. Vad blir resultatet? (2p)
   b. Du ombeds istället att pröva nollhypotesen $\mathbf{H_0}$: väntevärdet=**3.5** med signifikansnivån **1%** mot den alternativa hypotesen $\mathbf{H_1}$: väntevärdet $\neq$ **3.5** . Vad blir då din slutsats? (2p)

VÄND!

5. En tvådimensionell stokastisk variabel **(X,Y)** är likformigt fördelad på en triangel med de tre hörnen i **origo**, i punkten **(0,1)** och i punkten **(1,0).** D.v.s. **(X,Y)** har pdf **f(x,y)=** en konstant **c** för punkter **(x,y)** inne i triangeln och **f(x,y)=0** om **(x,y)** ligger utanför triangeln.
   a. Bestäm konstanten **c** . (1p)
   b. Bestäm väntevärdet **E[X]** och Variansen **Var[X]**. (2p)
   c. Bestäm den endimensionella fördelningsfunktionen för **X**. (1p)
   d. Bestäm den endimensionella sannolikhetstätheten för **X.** (1p)

6. Vikten på kex i en viss produktionsprocess antas variera lite grann på ett slumpmässigt sätt och oberoende för de individuella kexen. Väntevärdet av ett enskilt kex´s vikt antas vara **10.3 gram** och standardavvikelsen **0,81 gram**. Vad är (approximativt) sannolikheten för att **30** slumvalda kex väger minst **300 gram**. (3p)

7. I en enkel linjär regression med svarsvariabler **y** och inställningsvariabler **x** blev summan av de **n=11** beroende **y**-variablerna **15,3** och medelvärdet av **x**-variablerna var **0.32**. Skattningen av riktningskoeffecieten för regressionslinjen **β** blev **0.28**.

   a. Beräkna en observerad punktskattning av regressionslinjens intercept **α.** (1p)
   b. Antag att felvarianserna i den linjära regressionsmodellen för **Y-**variablerna i regressionsmodellen är känd och lika med **9,** att residualerna (=felen) är oberoende och normalfördelade, och att den (på vanligt enstickprovs-vis) beräknade "stickprovsvariansen" för de deterministiska **x-**variablerna är **8.3.** Vilken fördelning har då den teoretiska punktskattningen av **β?** (2p)
   c. Använd resultatet i c (och förutsättningarna i **c**) för att beräkna ett observerat konfidensintervall för riktningskoefficienten **β** med konfidensgraden **99%.** (1p)

8. Du skall observera **7** oberoende stokastiska variabler $X_1, \ldots, X_7$ med samma sannolikhetstäthet som beror på en okänd positiv parameter θ och har formen

$$p(x;\theta) = \theta \exp(-2\theta |x|), \; x \in R .$$

   a. Härled Maximum Likelihood-skattningen av θ. (2p)
   b. Ange den observerade punktskattningen baserad på observation av stickprovet: **1.2** , **1.8** , **-3.3** , **5.2** , **4.9** , **-2.0** , **0.6** . (1p)
   c. Är skattingen i **a-**delen väntevärdesriktig? (1p)

LÖSNINGSFÖRSLAG MAT. STAT F2, TMA321
LÖRDAGEN 18 APRIL 2015. OLLE NERMAN

1\. a. SÖKT SANNOLIKHET = X ⇒ TOTALA SANNOLIKHETSLAGEN GER EKVATIONSSYSTEMET $0{,}2 \cdot 0{,}3 + 0{,}8 \cdot x = 0{,}18$ OCH SVARET $x = 15\%$

b. BAYES FORMEL GER MED A = "GRAFITTIKÖP" OCH B = "TIDNINGSKÖP"

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} = \frac{0{,}2 \cdot 0{,}3}{0{,}18} = \frac{1}{3}$$

2\. $P(X<1) = 0{,}7 \Rightarrow P(\ln X < 0) = 0{,}7 \Rightarrow \Phi\left(\frac{0-\mu}{\sigma}\right) = 0{,}7$ (TABELL) GER $-\mu/\sigma \approx 0{,}76$

PÅ LIKNANDE VIS $P(X>2) = 0{,}01 \Rightarrow \Phi\left(\frac{\ln 2 - \mu}{\sigma}\right) = 0{,}99$ (TABELL) GER $\frac{\ln 2 - \mu}{\sigma} \approx 2{,}33$. FRÅN DESSA TVÅ SYSTEM FÅS

$\frac{\ln 2}{\sigma} \approx 2{,}33 - 0{,}76$ D.V.S. $\sigma \approx \frac{\ln 2}{2{,}33 - 0{,}76} \approx 0{,}441$

OCH $\mu \approx -0{,}76 \cdot \frac{\ln 2}{2{,}33 - 0{,}76} \approx -0{,}336$

3\. a. JA, TY $E[\hat{c}] = E\left[\frac{X}{15}\right] = \frac{1}{15} E[X] = \frac{1}{15} \cdot 15c = c$

b. $VAR[\hat{c}] = VAR\left[\frac{X}{15}\right] = \frac{VAR[X]}{15^2} = \frac{15c}{15^2} = \frac{c}{15} \Rightarrow$

STANDARDFELET $= \sqrt{VAR[\hat{c}]} = \sqrt{\frac{c}{15}}$

c. SÖKT PARAMETER ÄR $P(\text{"}Po(4c)\text{"} = 0) = e^{-4c}$. SKATTAS LÄMPLIGEN AV $e^{-4\hat{c}}$. OBSERVERAT $\hat{c} = 7/15$ GER OBSERVERAD SKATTNING $e^{-4\hat{c}} = e^{-(28/15)} \approx 0{,}155$

4\. a. $\sqrt{8}\,\frac{\bar{X} - \mu}{S} \sim$ t-FÖRDELAD(7). $F_{t(7)}(-1{,}895) = 0{,}05$ GER

$P\left(\sqrt{8}\,\frac{\bar{X} - \mu}{S} \geq -1{,}895\right) = 0{,}95$. GER TEORETISKT INTERVALL:

$\mu \leq \bar{X} + 1{,}895 \cdot \frac{S}{\sqrt{8}}$ (95%). OBSERVERAT

KONFIDENSINTERVALL

$\mu \leq 2{,}45 + 1{,}895 \cdot \frac{0{,}16}{\sqrt{8}}$ (95%)

b. TESTVARIABEL $T = \sqrt{8}\,\frac{\bar{X} - 3{,}5}{S} \underset{H_0}{\sim}$ t-FÖRDELAD (7)

EFTERSOM $F_{t(7)}(3{,}5) = 0{,}995$ SÅ FÖRKASTAR DET TVÅSIDIGA TESTET OM $T \leq -3{,}5$ ELLER $T \geq 3{,}5$.

OBSERVERAD TESTSTATISTIKA ÄR

$t = \sqrt{8}\,\frac{2{,}45 - 3{,}5}{0{,}16}$ ÄR $< -3{,}5$ ALLTSÅ

FÖRKASTAS H_0.

TMA 321 18 APRIL 2015
FORTS. OLLE NERMAN.

5a. $\iint_{R^2} f(x,y)\,dx\,dy = 1 = \int_0^1\left(\int_0^{1-x} c\,dy\right)dx = \frac{c}{2} = 1 \Rightarrow c = 2$

b. $E[X] = \iint_{R^2} x\, f(x,y)\,dx\,dy = \int_0^1\left(\int_0^{1-x} x\, 2\,dy\right)dx = \frac{1}{3}$

$E[X^2] = \iint_{R^2} x^2 f(x,y)\,dx\,dy = \int_0^1\left(\int_0^{1-x} x^2\, 2\,dy\right)dx = \frac{1}{6} \Rightarrow$

$VAR[X] = E[X^2] - E[X]^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$

c. $F_X(x) = \begin{cases} 0 & \text{OM } x \le 0 \\ \int_0^x\left(\int_0^{1-u} 2\,dy\right)du & \text{OM } x \in [0,1] \\ 1 & \text{OM } x \ge 1 \end{cases}$ $= 2x - x^2$ OM $x \in [0,1]$

d. $f_X(x) = \frac{d}{dx}F_X(x) = \begin{cases} 2(1-x) & x \in [0,1] \\ 0 & \text{FÖ.} \end{cases}$

6. OM T TOTALVIKTEN I GRAM AV DE 30 KEXEN SÅ GER C.GR.V. SATSEN $T \approx N(309, \sigma^2 = 30 \cdot 0{,}81)$
$= N(309, 24.3)$
$P(T > 300) = 1 - P(T \le 300) \approx 1 - \Phi\left(\frac{300-309}{\sqrt{24{,}3}}\right) = \Phi\left(\frac{9}{\sqrt{24{,}3}}\right) \approx 0{,}965$

7 a. $\hat\alpha = \bar y - \hat\beta \bar x = \frac{15{,}3}{11} - 0{,}28 \cdot 0{,}32 \approx$

b. $\hat\beta \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{11}(x_i - \bar x)^2}\right) = N\left(\beta, \frac{9{,}0}{10 \cdot 8{,}13}\right) \approx N(\beta, 0{,}1084)$

c. ANVÄND NORMALFÖRDELNINGSINTERVALL:

$\beta = 0{,}28 \pm 2{,}58 \cdot 0{,}329$ (99%) TY $\Phi(2{,}58) = 0{,}995$
$\beta = 0{,}28 \pm 0{,}85$ (99%)

8 a. $L(\theta) = \theta^7 e^{-\theta(\sum_{i=1}^7 |x_i|)} \Rightarrow \frac{d \ln L(\theta)}{d\theta} = 0$ GER
$\frac{7}{\theta} - \sum_{i=1}^7 |x_i| = 0$ SÅ ATT $\hat\theta = \frac{7}{\sum_{i=1}^7 |x_i|}$

b. $\sum_{i=1}^7 |x_i| = 19$ GER $\hat\theta = \frac{7}{19} \approx 0{,}368$

c. BLEV ALLDELES FÖR SVÅR! MAN KAN VISA ATT
$Y = \sum_{i=1}^7 |x_i| \sim \Gamma$-FÖRDELAD$(7, \theta)$ OCH ATT $E\left[\frac{1}{Y}\right] = \frac{\theta}{6}$
SÅ ATT $E\left[\frac{7}{\sum_{i=1}^7 |x_i|}\right] = \frac{7}{6} \cdot \theta \ne \theta$ ALLTSÅ ÄR ML-SKATTN.
INTE VÄNTEVÄRDESRIKTIG!