Tk Mattskatt

F1

sid...

~~5:-~~ 10:-

1 4 7 12 0.1 ______

2 5 8 11 0.5 ______

3 6 9 10 ______

/Per 3 år

Tentamensskrivning i Matematisk statistik F1
Onsdagen den 16/1, 2002.

Jour: Urban Hjorth, tel 772 5362, examinator Urban Hjorth.

Hjälpmedel: Tabeller som utdelats i kursen. Kalkylator med tomt textminne. Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 19 för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen.

Del 1:

1. Definiera varians och härled sedan variansen för $aX - bY + c$ om X och Y är oberoende med $\mu_x, \mu_y, \sigma_x, \sigma_y$. Ange var oberoendet utnyttjas.

2. Den tvådimensionella variabeln (X, Y) har frekvensfunktionen

$$f(x,y) = \frac{1 + x + y + cxy}{c+3} e^{-(x+y)}, x > 0, y > 0.$$

a) Beräkna den marginella frekvensfunktionen för X.

b) Bestäm väntevärdet för $X + Y$.

3. Ur en mängd med 7 blåa, 4 röda och 3 gröna föremål drar man slumpvis och utan återläggning 4 stycken. Bestäm sannolikheterna att få 1, 2 eller 3 färger.

4. Vid tillverkning av ett material skapas ett moln av fibrer i olika längder vilka blandas om helt slumpmässigt i en turbulent ström innan de sugs ner på en vira. För att studera fiberlängdsfördelningen indelas fibrerna i längdklasser. Studera ett litet delområde på viran. Antag att materialet från detta område skall detaljstuderas. Låt $X_i(t)$ beteckna antalet fibrer i längdklassen i som landat i delområdet fram till tiden t (fiberuppsamlingen startas vid tiden 0).

a) Föreslå en lämplig modell för $X_i(t)$. Motivera hur du resonerar, eventuella oberoendeförutsättningar etc.

b) Är det rimligt att anta att $X_i(t)$ är oberoende för olika i? Vilken fördelning får i så fall $\sum_{i=1}^{K} X_i(t)$ (det totala antalet fibrer i längdklasserna 1 till K i det detaljstuderade området).

Vänd!!

Del 2:

5. Variablerna X_i är oberoende $N(\mu_1, \sigma)$ och Y_i är oberoende $N(\mu_2, \sigma)$ (samma standardavvikelse). Alla variablerna är oberoende. Data:

x_i: 12.5 11.7 13.0 12.3 12.4

y_j: 10.8 12.2 9.9 10.5 11.3 11.3 10.6

Skatta parametrarna μ_1, μ_2, σ så bra som möjligt.

6. På 100 minuter registreras 22687 radioaktiva sönderfall med en apparatur som anses registrera samtliga sönderfall. Gör ett 99% konfidensintervall för sönderfallsintensiteten, d.v.s. det förväntade antalet sönderfall per minut.

7. För att visa att ett preparat har medicinsk verkan läter man 150 personer få preparatet och 150 andra får ett verkningslöst s.k. placebo med samma utseende och smak. Personerna utses slumpmässigt till de båda grupperna och patienterna och läkaren som bedömer utvecklingen är ovetande om vilken medicinering som getts så att inget av medlen gynnas. Data från försöket (antal patienter som blivit):

	Bättre	Oförändrat	Sämre
Preparat	94	33	23
Placebo	69	55	26

Gör ett hypotestest med signifikansnivån 0.01 (approximativt) och försök visa att medicinen är verksam. Observera att det bara är positiv effekt av medicinen som är intressant vilket man bör utnyttja.

8. I ett gammalt industriområde vill man studera halten av bly och blyföreningar och tar därför n jordprover, $x_1, \ldots, x_n$, i olika delar av området som sedan analyseras på ett lab. Som modell prövas frekvensfunktionen

$$f(x; \theta) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}},\ x > 0,$$

med en enda parameter θ och proverna bedöms som oberoende eftersom provtagningsplatserna slumpats ut. Finn maximum likelihoodskattningen av parametern och studera dess väntevärde.

Mat Stat F1 16/1 2002 . Lösningar.

1. Se kursmaterial

2. $f_X(x) = \int_0^\infty \frac{1+x+y+cxy}{c+3} e^{-(x+y)} dy$;

a) $\int_0^\infty e^{-y} dy = 1$, $\int_0^\infty y e^{-y} dy = 1$ ger $\frac{1+x+1+cx}{c+3} e^{-x} = f_X(x)$, $x > 0$.

b) $E[X] = \int_0^\infty x f_X(x) dx = \int \frac{x+x^2+x+cx^2}{c+3} e^{-x} dx$

$\int_0^\infty x^2 e^{-x} dx = \cancel{\left[-x^2 e^{-x}\right]_0^\infty} + \int_0^\infty 2x e^{-x} dx = 2 \Rightarrow E[X] = \frac{1+2+1+2c}{c+3} = \frac{4+2c}{c+3}$

$EY = EX$ (symmetri). $E[X+Y] = E[X] + E[Y] = \frac{8+4c}{c+3}$

---

3. X = antal färger

$P(X=1) = P(A) + P(B)$; $A = \{4 \text{ blåa}\}$, $B = \{4 \text{ röda}\}$ är disjunkta

$= \frac{\binom{7}{4}}{\binom{14}{4}} + \frac{\binom{4}{4}}{\binom{14}{4}} = \frac{7 \cdot 6 \cdot 5 \cdot 4 + 4 \cdot 3 \cdot 2 \cdot 1}{14 \cdot 13 \cdot 12 \cdot 11} = \underline{\underline{0{,}036}}$

$P(X=3) = P(2 \text{ blå}, 1 \text{ röd}, 1 \text{ grön}) + P(\text{en blå}, 2 \text{ röda}, \text{en grön})$
$+ P(1 \text{ blå}, 1 \text{ röd}, 2 \text{ gröna}) = \frac{\binom{7}{2} \cdot 4 \cdot 3 + 7\binom{4}{2} 3 + 7 \cdot 4 \cdot \binom{3}{2}}{\binom{14}{4}} =$

$= \frac{21 \cdot 4 \cdot 3 + 7 \cdot 6 \cdot 3 + 7 \cdot 4 \cdot 3}{1001} = \underline{0{,}4615}$

$P(X=2) = 1 - 0{,}036 - 0{,}4615 = \underline{\underline{0{,}5025}}$

---

4. a) $X_i(t)$ bör vara en Poissonprocess med intensitet c_i. Det är ständigt samma intensitet att en fiber av längdklass i träffar området.

b) Oberoendet verkar rimligt eftersom fibrer av annan längd kan landa där oberoende av $X_i(t)$.

$\sum_1^k X_i(t)$ blir en summa av ober. Poissonvariabler och alltså Poissonfördelad $Po\left(\sum_1^k c_i t\right)$.

5. $\hat{\mu}_1 = \bar{x} = 12.38 \quad \hat{\mu}_2 = \bar{y} = 10.94 \; ; \; s_x^2 = \frac{1}{n_x - 1}\sum(x_i - \bar{x})^2$

$$s^2 = \frac{4s_x^2 + 6s_y^2}{4+6} = 0.412 \; ; \quad \hat{\sigma} = s = \sqrt{0,412} = 0,642$$

---

6. X = antalet sönderfall på 100 minuter.

$X \in Po(100c) \approx N(100c, \sigma = \sqrt{100c})$

$\frac{X - 100c}{\sqrt{100c}}$ är $N(0,1)$; $\quad P\left(-2,58 < \frac{X-100c}{\sqrt{100c}} < 2,58\right) = 0,99$

Om detta gäller för vårt observerade x så är

$$-2,58 < \frac{22687 - 100c}{\sqrt{100c}} < 2,58$$

Exakt lösn: $22687 - 385 < 100c < 22687 + 392$

$$223.0 < c < 230.8$$

Enkel approx $100c = 22687 \pm \underbrace{2.58\sqrt{22687}}_{388.6}$; $223.0 < c < 230.8$

---

7. Jämför antalet som blivit bättre.

X = antal med prep. som är bättre; $\quad X \in Bi(150, p_1)$

Y = placebo $\quad Y \in Bi(150, p_2)$

$H_0: p_1 = p_2 \; ; \quad H_1: p_1 > p_2$.

$X - Y$ är approx $N(150p_1 - 150p_2, \sqrt{150p_1(1-p_1) + 150p_2(1-p_2)})$

$\frac{X - Y - 150(p_1 - p_2)}{\sqrt{\dots}}$ är $N(0,1)$

H_0 ger att $T = \frac{X - Y - 0}{\sqrt{150\hat{p}_1(1-\hat{p}_1) + 150\hat{p}_2(1-\hat{p}_2)}}$ approx $N(0,1)$

$\hat{p}_i$ kan antingen skattas individuellt $\frac{94}{150}$ resp $\frac{69}{150}$ eller lika $\frac{163}{300}$

Individuella $\hat{p}_i$ ger $\quad T = \frac{94 - 69}{\sqrt{72,35}} = 2.94$

H_1: ensidig $\Rightarrow$ 1% risk i högra kanten $\Rightarrow$ Kritiskt område $\underline{\underline{T > 2,33}}$

$\underline{H_0 \text{ förkastas}}$ på signifikansnivån 1%.

---

8. $\ln L = \sum \ln x_i - n\ln\theta - \frac{\sum x_i^2}{2\theta}$;

$\frac{\partial \ln L}{\partial \theta} = \frac{-n}{\theta} + \frac{\sum x_i^2}{2\theta^2} = 0$ ger $\hat{\theta} = \frac{\sum x_i^2}{2n}$ (maximum).

$E[\theta^*] = E\left[\frac{1}{2n}\sum X_i^2\right] = \frac{1}{2}E[X^2] = \frac{1}{2}\int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}dx = \left| \begin{matrix} x = y\sqrt{\theta} \\ dx = dy \cdot \sqrt{\theta} \end{matrix} \right|$

$= \frac{\theta}{2}\int_0^\infty y^3 e^{-y^2/2}dy = \frac{\theta}{2}\left\{\left[-y^2 e^{-y^2/2}\right]_0^{\infty} + \int_0^\infty 2y e^{-y^2/2}dy\right\} = \theta\left[-e^{-y^2/2}\right]_0^\infty = \underline{\underline{\theta}}$

Tentamensskrivning i Matematisk statistik F1
Torsdagen den 11/1, 2001.

Jour: Mikael Knutsson, tel 7725380, examinator Urban Hjorth.
Hjälpmedel: Tabeller och fördelningstablå som utdelas vid skrivningen. Kalkylator med tomt textminne.
Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 19 för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen.

Del 1:

1. Till en brandstation inkommer i genomsnitt 2 larm i veckan. Vi antar att denna siffra gäller alla årets veckor (dvs vi bortser från den extra stora risken vid storhelger som jul och nyår).

a) Vilken sannolikhetsfördelning bör man ansätta på antalet larm under en vecka?

b) Vad är sannolikheten för 5 eller fler larm under en vecka?

c) Vad är sannolikheten att någon av årets 52 veckor ger 5 eller fler larm?

2. Bestäm väntevärde och standardavvikelse för summan av 10 oberoende variabler som alla kan anta värdena 1, 2 eller 3 med sannolikheterna 0.6, 0.3 och 0.1 respektive.

3. Livslängden X för en fordonskomponent (del av bromssystemet) har modellerats som en stokastisk variabel med frekvensfunktionen

$$f(x) = abx^{b-1}e^{-ax^b}, \; x > 0.$$

(Enhet 10000 km körsträcka.) Parametrarna har skattats till $a = 0.0024$, $b = 2.2$. För att bedöma sannolikheten att komponenter som körts felfritt sträckan t håller i ytterligare ett serviceintervall (1.5 enheter) så beräknas den betingade sannolikheten

$$P(X > t + 1.5 | X > t).$$

Finn denna sannolikhet för $t = 3$ och $t = 9$.

4. a) Låt X och Y vara oberoende med momentgenererande funktioner $m_X(t)$ och $m_Y(t)$. Härled momentgenererande funktionen för $aX + bY + c$ uttryckt i m_X och m_Y.

b) Låt nu X_i, $i = 1, \ldots, n$ vara oberoende $N(\mu, \sigma)$. Tillämpa a-delen (generaliserad till flera variabler) och härled momentgenererande funktionen för $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ där $\bar{X} = \sum_1^n X_i/n$. Obs. det är härledningen och inte slutresultatet som efterfrågas.

Vänd!!

Del 2:

5. Andelen θ av alla tvillingfödslar är enäggstvillingar ocg dessa får säkert samma kön. Resten är tvååggstvillingar och får manligt eller kvinnligt kön oberoende av varandra. (Sannolikheten för pojke är ca 0.514 men vi räknar här med lika sannolikheter 0.5 för flicka och pojk.) Av 1074 tvillingfödslar under ett år var 667 likkönade medan resten, dvs 407 stycken, gav olika kön. Skatta parametern θ.

6. För att jämföra fiskars tillväxt i olika miljöer så väger och märker man ett antal ungefär lika stora fiskar som sedan fördelas slumpvis och får leva i nätburar i respektive vatten. Efter en tid mäter man fiskarnas viktökning. Data (gram viktökning)

Miljö 1: 72, 35, 54, 68, 61, 49, 66, 70

Miljö 2: 39, 54, 50, 51, 46, 43, 33

Antag att alla fiskar tillväxer oberoende av varandra. Pröva med ett lämpligt test på signifikansnivån $\alpha = 0.10$ om väntevärdet av tillväxten kan vara lika i de båda miljöerna. (Lika varians får förutsättas).

7. Genom en ytbehandling ändras sannolikheten för defekter från p till θp där både p och θ är okända konstanter. Ställ upp likelihooden och finn maximum likelihoodskattningen av θ och p med hjälp av följande data från oberoende försök

utan ytbehandling: 1076 enheter gav 287 defekter;

med ytbehandling: 812 enheter gav 113 defekter.

8. Vi har gjort två regressionsanalyser med samma typ av modell $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ men på helt oberoende datamaterial. Standardavvikelsen σ för ε_i är samma i båda datamaterialen trots en rätt stor skillnad på de skattade standardavvikelserna. Gör ett konfidensintervall för att studera om lutningen β_1 är olika i de båda datamaterialen. Lämplig statistisk felrisk 0.01. Från en datorkörning får vi följande resultat:

Datamaterial 1: $n_1 = 20$ data. $\hat{\beta}_0 = 4.6761$, $\hat{\beta}_1 = 0.6249$, $s = 3.3926$, $X'X = \begin{pmatrix} 20 & 540 \\ 540 & 17538 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 0.2965 & -0.0091 \\ -0.0091 & 0.00034 \end{pmatrix}$.

Datamaterial 2: $n_2 = 30$ data. $\hat{\beta}_0 = 2.4258$, $\hat{\beta}_1 = 0.8819$, $s = 5.0514$, $X'X = \begin{pmatrix} 30 & 764 \\ 764 & 27672 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 0.1123 & -0.0031 \\ -0.0031 & 0.00012 \end{pmatrix}$.

Vi erinrar om att vid normalfördelning har vektorn $\hat{\beta}$ en $N(\beta, \sigma^2 (X'X)^{-1})$-fördelning.

Tenta i Matematisk statistik för F1, 2001-01-11

UPPGIFT 1: (a) Poisson(2)

(b) $P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} P(X=x) =$

$$= 1 - \sum_{x=0}^{4} \frac{e^{-2} 2^x}{x!} \approx 0{,}0527$$

(c) $1 - (1-0{,}0527)^{52} \approx 0{,}94$

UPPGIFT 2: $Y = \sum_{i=1}^{10} X_i$, där $X_1, X_2, \ldots, X_{10}$ ober.

$E[X_i] = 1 \cdot 0{,}6 + 2 \cdot 0{,}3 + 3 \cdot 0{,}1 = 1{,}5$

$E[X_i^2] = 1^2 \cdot 0{,}6 + 2^2 \cdot 0{,}3 + 3^2 \cdot 0{,}1 = 2{,}7$

$\Rightarrow Var(X_i) = 2{,}7 - 1{,}5^2 = 0{,}45$

$\Rightarrow E[Y] = \sum_{i=1}^{10} E[X_i] = 10 \cdot 1{,}5 = 15$

$Var(Y) = \sum Var(Y_i) = 10 \cdot 0{,}45 = 4{,}5$,

dvs standardavvikelse $= \sqrt{4{,}5} \approx 2{,}12$

UPPGIFT 3:

$$P(X > t) = \int_t^{\infty} abx^{b-1} e^{-ax^b} dx = \left[-e^{-ax^b}\right]_t^{\infty} =$$

$$= e^{-at^b}$$

$$\Rightarrow P(X > t+1{,}5 \mid X > t) = \frac{P(X > t+1{,}5)}{P(X > t)} = \frac{e^{-a(t+1{,}5)^b}}{e^{-at^b}} =$$

$$= \begin{cases} 0{,}9620, & \text{då } t = 3 \\ 0{,}8853, & \text{då } t = 9 \end{cases}$$

UPPGIFT 4: (a) $m(t) = E[e^{t(aX+bY+c)}] =$

$$= E[e^{atX}]E[e^{btY}]e^{ct} = e^{ct} m_X(at) m_Y(bt)$$

(b) $E[e^{t[(\bar{X}-\mu)/(\sigma/\sqrt{n})]}] = e^{-\mu t/(\sigma/\sqrt{n})} \prod_{i=1}^{n} E[e^{t \cdot \frac{1}{n} \cdot \frac{X_i}{\sigma/\sqrt{n}}}]$

$$= e^{-\mu t/(\sigma/\sqrt{n})} \prod_{i=1}^{n} E[e^{tX_i/(\sigma\sqrt{n})}] = e^{-\mu t/(\sigma/\sqrt{n})} \prod_{i=1}^{n} m_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right) =$$

$$= e^{-\frac{\mu t}{\sigma/\sqrt{n}}} \prod_{i=1}^{n} e^{\frac{\mu t}{\sigma\sqrt{n}} + t^2/2n} = e^{t^2/2}$$

UPPGIFT 5: $P(\text{samma kön}) = \theta + (1-\theta)\cdot 0.5 = 0.5(1+\theta)$

$P(\text{olika kön}) = (1-\theta)\cdot 0.5$

Alternativ 1: Skatta θ med det θ som maximerar

$L(\theta) = [0.5(1+\theta)]^{667}[0.5(1-\theta)]^{407}$

$\Rightarrow \hat{\theta} = 260/1074 \approx 0.2421$

Alternativ 2: Skatta $0.5(1+\theta)$ med $667/1074$

$\Rightarrow$ skatta θ med $\hat{\theta} = 260/1074 \approx 0.2421$

UPPGIFT 6: Miljö 1: $n_1 = 8$, $\bar{x}_1 = 59.375$, $s_1^2 \approx 160.554$

Miljö 2: $n_2 = 7$, $\bar{x}_2 = 45.143$, $s_2^2 = 54.476$

$$S_p^2 = \frac{7s_1^2 + 6s_2^2}{13} \approx 111.595$$

Testa $H_0: \mu_1 = \mu_2$ mot $H_1: \mu_1 \neq \mu_2$

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \approx 2.6 > 1.771 = t_{13,0.05}$$

$\Rightarrow$ Förkasta H_0 på $\alpha = 0.10$-nivå.

UPPGIFT 7: Skatta p med det p som maximerar

$L_1(p) = p^{287}(1-p)^{789} \Rightarrow \hat{p} = \frac{287}{1076} \approx 0.2667$

Skatta θp med det värde som maximerar

$L_2(\theta p) = (\theta p)^{113}(1-\theta p)^{699} \Rightarrow (\widehat{\theta p}) = \frac{113}{812}$

$\Rightarrow$ Skatta θ med $\hat{\theta} = (113/812)/(287/1076) \approx 0.5217$

UPPGIFT 8:

99% k.i. för $\beta_{11} - \beta_{12}$:

$$\hat{\beta}_{11} - \hat{\beta}_{12} \pm t^{(0.005)}_{(20-2)+(30-2)} \cdot S_p\sqrt{0.00034 + 0.00012} =$$

$$= 0.6249 - 0.8819 \pm 2.687 \cdot \sqrt{\frac{18 \cdot 3.3926^2 + 28 \cdot 5.0514^2}{46}} \cdot \sqrt{0.00046} =$$

$$= -0.257 \pm 0.258$$

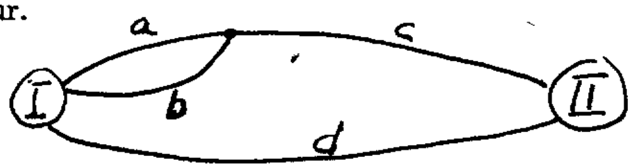
Tentamensskrivning i Matematisk statistik F1
Tisdagen den 15/8, 2000.

Jour: Mikael Knutsson, tel 7725380, examinator Urban Hjorth.
Hjälpmedel: Tabeller och fördelningstablå som utdelas vid skrivningen. Kalkylator med tomt textminne.
Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 20 (19 utan bonus) för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen. Bonus från kontrollskrivningen adderas till poängen på del 1.

Del 1:

1. Ur en mängd med N olika stora objekt drar man slumpvis k stycken, $2 \leq k \leq N$, (utan återläggning). Låt A vara händelsen att det största objektet blir draget och B händelsen att det minsta objektet blir draget. Bestäm sannolikheten för $A \cap B$.

2. Ett elnät består av matningsnoden I, belastningsnoden II och länkarna a, b, c, d enligt figur.



Låt A, B, C, D vara händelserna att rspektive länk är felfri och p_a, p_b, p_c, p_d motsvarande sannolikhete.

a. Uttryck händelsen att I och II har kontakt med hjälp av A, B, C, D.

b. Finn sannolikheten för samma händelse om länkarna är oberoende och $p_a = p_b = 0.8$, $p_c = p_d = 0.9$.

3. Vid arbete i radioaktiv miljö bär varje person en mätare av personlig stråldos. Resultatet på denna kan med någorlunda noggrannhet omräknas till antalet registrerade radioaktiva impulser. 20 arbetare tillåts arbeta i vardera 2 timmar och har då fått i genomsnitt $\overline{x}$ registrerade impulser. Längre in i systemet tillåts 8 arbetare vistas i 15 minuter vardera och har då i genomsnitt fått $\overline{y}$ registrerade impulser. För att jämföra de båda miljöernas strålning per timme bildar man differensen

$$z = \frac{\overline{x}}{2} - 4\overline{y}.$$

Modellera detta genom att införa stokastiska variabler för varje arbetares dos och bestäm väntevärde och varians för z uttryckt i dessa fördelningars parametrar. (Vi förutsätter att radiaktivitetens slumpvariationer är den enda variationskällan, vilket bör gälla om strålningen är homogen i respektive utrymme och mätarna är identiska men inte annars.)

4. Formulera och visa centrala gränsvärdessatsen.

Vänd!!

Del 2:

5. Värdena 15.7 17.3 17.8 13.5 18.2 20.5 16.8 är observationer av oberoende $N(\mu, \sigma)$-fördelade variabler. Gör 95% konfidensintervall för μ och för σ och ange tydligt vilka fördelningar som uppstår i beräkningarna.

6. Rayleighfördelningen kan härledas ur normalfördelingen och är användbar bl.a. för att beskriva vågstorlek. Fördelningen har frekvensfunktionen

$$f(x) = \frac{2x}{\gamma} e^{-x^2/\gamma}, \; 0 < x < \infty, \; \gamma > 0.$$

Från $n = 16$ oberoende observationer $x_1, \ldots, x_n$ skattas parametern γ med maximum likelihoodmetoden.

a. Finn ett uttryck för skattningen.

b. Finn skattningens väntevärde (som funktion av γ).

7. Ett företag med kontinuerlig drift har utsläpp i en älv och påstås att deras återvinningsmetoder är så bra att inga kvicksilverföreningar släpps ut. För att kontrollera detta görs under 10 veckor först ett slumpvis val av tidpunkt då man tar ett vattenprov uppströms om utsläppet och omedelbart efteråt tar man sedan ett nytt prov nedströms om utsläppet. Proverna analyseras och värdena ges nedan. Man vet att det finns ytterligare föroreningskällor längre upp utefter älven.

a. Inför beteckningar och ange vilka oberoende/beroendeantaganden som bör göras med hänsyn till föroreningskällor längre upp i älven.

b. Bedöm med lämpligt test eller konfidensintervall om företaget släpper ut kvicksilver. Arbeta med 1% statistisk felrisk.

Vecka	1	2	3	4	5	6	7	8	9	10
Uppströms	1.0	2.2	2.1	1.2	2.8	4.0	3.7	3.0	5.6	3.3
Nedströms	1.9	2.0	2.5	1.9	3.4	4.2	3.7	2.9	5.9	3.8

8. Med en given mätmetod registreras den radioaktiva bakgrundsstrålningen under 30 minuter varvid 1372 impulser avläses. Efter en misstänkt radioaktiv läcka görs en ny mätning där man utnyttjar exakt samma mätmetod.

a. Kan man bevisa någon ökad strålning om man under 30 minuter får 1503 radioaktiva impulser?

b. Hur ändras analysen om man istället mäter i bara 10 minuter efter den misstänkta läckan och får 501 impulser, men har samma 30 minuters avläsning av bakgrundsstrålningen.

Använd 95% konfidensgrad eller 5% signifikansnivå i analysen. OBS: Det är bara ökningar av radioaktiviteten som är av intresse.

Tenta i Matematisk statistik för F1, 2000-08-15.

UPPGIFT 1: $P(A\cap B) = \dfrac{\binom{N-2}{k-2}}{\binom{N}{k}} = \dfrac{\frac{(N-2)!}{(k-2)!(N-k)!}}{\frac{N!}{k!(N-k)!}} = \dfrac{k(k-1)}{N(N-1)}$

UPPGIFT 2: (a) $((A\cup B)\cap C)\cup D$

(b) $P(((A\cup B)\cap C)\cup D) = P((A\cup B)\cap C) + P(D) - P((A\cup B)\cap C\cap D)$

$= P(A\cap C) + P(B\cap C) - P(A\cap B\cap C) + P(D) - P(A\cap C\cap D) - P(B\cap C\cap D) + P(A\cap B\cap C\cap D)$

$= p_A p_C + p_B p_C - p_A p_B p_C + p_D - p_A p_C p_D - p_B p_C p_D + p_A p_B p_C p_D = 0.9864$

UPPGIFT 3: Antag att impulser i början av systemet inträffar enligt Poissonprocess med intensitet λ_1 / timme, och längre in i systemet med λ_2 / timme.

Arbetare i början: $X_1, X_2, \ldots, X_{20}$; $X_i \sim \text{Poisson}(2\lambda_1)$

Arbetare längre in: $Y_1, Y_2, \ldots, Y_8$; $Y_i \sim \text{Poisson}(\lambda_2/4)$

$E[X_i] = Var(X_i) = 2\lambda_1 \Rightarrow E[\bar{X}] = 2\lambda_1,\ Var(\bar{X}) = \frac{2\lambda_1}{20} = \frac{\lambda_1}{10}$

$E[Y_i] = Var(Y_i) = \frac{\lambda_2}{4} \Rightarrow E[\bar{Y}] = \frac{\lambda_2}{4},\ Var(\bar{Y}) = \frac{\lambda_2}{4\cdot 8} = \frac{\lambda_2}{32}$

$E[Z] = \frac{1}{2}E[\bar{X}] - 4E[\bar{Y}] = \frac{1}{2}\cdot 2\lambda_1 - 4\cdot\frac{1}{4}\lambda_2 = \lambda_1 - \lambda_2$

$Var(Z) = \frac{1}{4}Var(\bar{X}) + 16\,Var(\bar{Y}) = \frac{\lambda_1}{40} + \frac{\lambda_2}{2}$

UPPGIFT 5: $n = 7$, $\bar{x} = 17.11$, $s \approx 2.17$

$\dfrac{\bar{X}-\mu}{S/\sqrt{7}} \sim t_6 \Rightarrow \bar{X} \pm t_{0.025,6}\cdot\frac{S}{\sqrt{7}}$ är ett 95% konf. int. för μ. Vi får

$17.11 \pm 2.447\cdot\frac{s}{\sqrt{7}} = 17.11 \pm 2.01 \quad (15.10, 19.12)$

$\dfrac{6S^2}{\sigma^2} \sim \chi^2_6 \Rightarrow \left(\sqrt{\dfrac{6S^2}{\chi^2_{0.025,6}}}, \sqrt{\dfrac{6S^2}{\chi^2_{0.975,6}}}\right)$ är ett 95% konf. int. för σ. Vi får

$\left(\sqrt{\dfrac{6s^2}{14.46}}, \sqrt{\dfrac{6s^2}{1.24}}\right) = (1.39, 4.78)$

UPPGIFT 6: (a) $L(\gamma) = \prod_{i=1}^{16} \frac{2X_i}{\gamma} e^{-X_i^2/\gamma} = \frac{1}{\gamma^{16}} \left(\prod_{i=1}^{16} X_i\right) e^{-\frac{1}{\gamma}\sum_{i=1}^{16} X_i^2} \cdot 2^{16}$

$\ln L(\gamma) = -16 \ln\gamma + \sum_{i=1}^{16} \ln X_i - \frac{1}{\gamma}\sum_{i=1}^{16} X_i^2 + 16\ln 2$

$\frac{\partial \ln L(\gamma)}{\partial \gamma} = 0 \Rightarrow -16/\gamma + 1/\gamma^2 \sum_{i=1}^{16} X_i^2 = 0,$

vilket ger ML-skattningen $\hat{\gamma} = \frac{1}{16}\sum_{i=1}^{16} X_i^2$

(b) $E[\hat{\gamma}] = E[\frac{1}{16}\sum_{i=1}^{16} X_i^2] = \frac{1}{16}\sum_{i=1}^{16} E[X_i^2] = E[X^2] =$

$= \int_0^\infty x^2 \frac{2x}{\gamma} e^{-x^2/\gamma} dx = \int_0^\infty \frac{2x^3}{\gamma} e^{-x^2/\gamma} dx = \left[\begin{matrix} t = x^2, \ x = \sqrt{t} \\ dx = dt/2\sqrt{t} \end{matrix}\right]$

$= \int_0^\infty \frac{t}{\gamma} e^{-t/\gamma} dt = \gamma$ (väntevärde för exp. fördelning)

UPPGIFT 7: (a) Låt $X_1, X_2, \ldots, X_{10}$ beteckna uppströms-observationerna och $Y_1, Y_2, \ldots, Y_{10}$ motsvarande (X_i, Y_i beroende) nedströms. Låt $D_1, D_2, \ldots, D_{10}$ vara skillnaderna (oberoende) $D_i = Y_i - X_i$ och antag att $D_i \overset{iid}{\sim} N(\mu_D, \sigma_D^2)$. Antag att föroreningar som ej beror på företag är lika stora uppströms som nedströms vid de olika mättillfällena (samma väntevärde). Om företaget inte släpper ut kvicksilver är $\mu_D = 0$.

(b) Testa $H_0: \mu_D = 0$ mot $H_1: \mu_D > 0$

Om H_0 sann, så är $T = \frac{\bar{d}}{s_d/\sqrt{10}} \sim t_9$,

Så förk. på 1%-nivå om $T > 2.821$

d-stickprov: 0,9, −0,2, 0,4, 0,7, 0,6, 0,2, 0, −0,1, 0,3, 0,5

$\bar{d} = 0{,}33$, $s^2 = 0{,}129 \Rightarrow t \approx 2.905$

$\Rightarrow$ Förk. H_0 på 1%-nivå, dvs finns signifikanta bevis på utsläpp.

UPPGIFT 8: Före läcka: Impulser enligt Poissonprocess med intensitet λ_1/minut

Efter läcka: Impulser enligt Poissonprocess med intensitet λ_2/minut

Testa $H_0: \lambda_1 = \lambda_2$ mot $H_1: \lambda_1 < \lambda_2$

Låt $X_1, X_2, \ldots, X_{n_1}$ vara impulser under n_1 minuter före läckan och $Y_1, Y_2, \ldots, Y_{n_2}$ efter läckan.

$X_i \sim Poisson(\lambda_1) \Rightarrow E[X_i] = Var(X_i) = \lambda_1$, $E[\bar{X}] = \lambda_1$, $Var(\bar{X}) = \frac{\lambda_1}{n_1}$

$Y_i \sim Poisson(\lambda_2) \Rightarrow E[Y_i] = Var(Y_i) = \lambda_2$, $E[\bar{Y}] = \lambda_2$, $Var(\bar{Y}) = \frac{\lambda_2}{n_2}$

$\bar{Y} - \bar{X} \overset{appr}{\sim} N(\mu = \lambda_2 - \lambda_1, \sigma^2 = \frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2})$

$$\Rightarrow \frac{\bar{Y} - \bar{X} - (\lambda_2 - \lambda_1)}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}} \overset{appr.}{\sim} N(0,1)$$

Om H_0 sann så är $Z = \frac{\bar{Y} - \bar{X}}{\sqrt{\hat{\lambda}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \overset{appr}{\sim} N(0,1)$,

där $\hat{\lambda} = (n_1 \bar{X}_1 + n_2 \bar{X}_2)/(n_1 + n_2)$,

så förkasta H_0 på 5%-nivå om $Z > 1{,}645$.

(a)

$$Z = \frac{\frac{1503}{30} - \frac{1372}{30}}{\sqrt{\frac{1372+1503}{60}\left(\frac{1}{30} + \frac{1}{30}\right)}} \approx 2{,}44$$

$\Rightarrow$ Förkasta H_0 på 5%-nivå.

(b)

$$Z = \frac{\frac{501}{10} - \frac{1372}{30}}{\sqrt{\frac{1372+501}{40}\left(\frac{1}{30} + \frac{1}{10}\right)}} \approx 1{,}75$$

$\Rightarrow$ Förkasta H_0 på 5%-nivå.

Tentamensskrivning i Matematisk statistik F1
Lördagen den 20/5, 2000.

Jour: Mikael Knutsson, tel 7725380, examinator Urban Hjorth.
Hjälpmedel: Tabeller och fördelningstablå som utdelas vid skrivningen. Kalkylator med tomt textminne.
Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 20 (19 utan bonus) för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen. Bonus från kontrollskrivningen adderas till poängen på del 1.

Del 1:

1. Vid tiden noll frigörs en partikel i en punkt som är placerad på en talaxel enligt en normalfördelning med $\mu = 1$, $\sigma = 2$. Därifrån rör den sig slumpmässigt på talaxeln så den oberoende av startpunkt vid tiden ett har förflyttat sig X från utgångspunkten, där X är normalfördelad med väntev'ärde 2 och standardavvikelse 3. Finn sannolikheten att partikeln ligger på den negativa sidan vid tiden 1.

2. En detalj inhandlas från två olika leverantörer. Leverantör I står för 2/3 av enheterna och leverantör II står för 1/3 och enheterna blir slumpmässigt omblandade innan de säljs till kunder. En detalj från I har 90% chans att hålla i minst 5 år och en detalj från II har 60% sannolikhet att hålla så länge. Den enhet vi handlat går sönder före fem år. Hur stor är den betingade sannolikheten att den är från I.

3. Ange förutsättningarna för att en variabel skall bli binomialfördelad. Visa hur frekvensfunktionen följer och härled på enklast möjliga sätt väntevärde och varians för en sådan $Bi(n,p)$-fördelad variabel.

4. Ett radioaktivt preparat har halveringstiden 27 timmar och vid tiden 0 har det intensiteten 8000 sönderfall per sekund.

a) Ange frekvensfunktionen för en radioaktiv atoms livslängd.

b) Hur många radioaktiva atomer finns det i preparatet vid tiden 0?

c) Ange fördelningen för antalet radioaktiva atomer som ännu inte sönderfallit efter en månad (30 dagar). Finn speciellt ett numeriskt värde på sannolikheten att ingen atom är kvar.

Del 2:

5. En variabel X kan anta värdena 0, 1, 2, 3, 4, 5. Symmetri och variabelns konstruktion ger att dess frekvensfunktion uppfyller villkoren $f(0) = f(5) = p$, $f(1) = f(4) = p^2$, $f(2) = f(3) = \frac{1}{2}(1 - 2p - 2p^2)$. Från 200 oberoende observationer får man följande resultat:

x	antal observationer
0	53
1	24
2	17
3	26
4	19
5	61

a) Skatta parametern p med maximum likelihoodmetoden.

b) Beräkna $\mu = \mu(p)$ och motivera varför momentmetoden inte kan användas på sin ursprungliga form.

6. Två nya material har framställts och för att jämföra deras dragstyrka så fogar man ihop dem så att fogen inte brister och sedan drar man tills det ena materialet går sönder. Vid 30 sådana provningar vann material A över material B i 21 fall och förlorade i 9 fall. Räcker detta för att visa att material A är starkare? Gör en modell och analysera med ett lämpligt konfidensintervall med 95% konfidensgrad.

7. I ett utmattningsförsök för metallstavar studerar man effekten av att polera materialets yta. Tanken är att poleringen eventuellt minskar risken för brott genom att sprickbildning inte så lätt startar vid ytan. Antalet belastningscykler till materialbrott avläses och har en mycket skev sannolikhetsfördelning. Därför studerar man istället logaritmerna av antalet belastningscykler som har ett betydligt mer normalfördelningsliknande utseende. Data (log antal cykler):

Utan polering: 9.00, 6.16, 10.28, 10.66, 7.36, 12.73, 12.73, 9.91, 10.75, 10.40, 9.57, 11.66, 8.64, 15.02, 9.68, 10.26, 12.45, 10.13, 9.78, 8.08, 10.67, 6.92, 11.64, 13.73, 8.40. Medelv. 10.26, skattad standardavvikelse 2.10.

Med polering: 14.35, 15.30, 8.47, 8.84, 13.67, 11.34, 13.95, 14.25, 14.00, 15.39, 13.90, 15.15, 9.41, 12.25, 11.92, 8.45, 12.91, 9.76, 15.69, 10.36. Medelv. 12.47, skattad standardavvikelse 2.48.

a) Vi skall testa hypotesen att poleringen inte har någon effekt mot den ensidiga hypotesen att polering ökar livslängden. Visa teorikunskaper genom skriva upp modellen, formulera hypoteserna i modellens termer och skriva ut fördelningen för alla led i räkningarna och ange det kritiska området C. Lika varians kan förutsättas. b) Genomför testet på signifikansnivån $\alpha = 0.01$. (a och b kan göras tillsammans).

8. Man har gjort regressionsanalys med modellen $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

a) Ange modellens förutsättningar om ε_i. Ge sedan modellen på vektorform och härled kovariansmatrisen för $\hat{\beta}$.

b) Gör ett konfidensintervall för lutningen β_1. Lämplig statistisk felrisk 0.01. Från en datorkörning får vi bl.a. följande resultat:

Datamaterial: $n = 20$ data. $\hat{\beta}_0 = 4.6761$, $\hat{\beta}_1 = 0.7249$, $s = 11.3926$, $X'X = \begin{pmatrix} 20 & 540 \\ 540 & 17538 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 0.2965 & -0.0091 \\ -0.0091 & 0.00034 \end{pmatrix}$.

Lösningar

Tenta i matematisk statistik för F1, 2000-05-20.

UPPG 1: Y=startpunkt $\sim N(\mu_Y=1, \sigma_Y^2=4)$, $X\sim N(\mu_X=2, \sigma_X^2=9) \Rightarrow X+Y\sim N(\mu=3, \sigma^2=13)$

$P(\text{negativa sidan}) = P(X+Y<0) = P\left(\frac{X+Y-3}{\sqrt{13}} < \frac{-3}{\sqrt{13}}\right) = \Phi(-0.83) = 0.2033$

UPPG 2: $P(I\mid \text{sönder}) = \frac{P(I\cap \text{sönder})}{P(\text{sönder})} = \frac{P(\text{sönder}\mid I)P(I)}{P(\text{sönder}\mid I)P(I)+P(\text{sönder}\mid II)P(II)} =$

$= \frac{0.1\cdot 2/3}{0.1\cdot 2/3+0.4\cdot 1/3} = \frac{1}{3}$

UPPG 3: Om n oberoende försök utförs och var och ett lyckas med sannolikhet p, så är antalet lyckade $\sim Bin(n,p)$.

$X\sim Bin(n,p) \Rightarrow f(x)=P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x=0,1,\dots,n,$

ty det finns $\binom{n}{x}$ sätt att välja x lyckade ur n, var och ett av dessa x stycken lyckas med slh p och var och ett av övriga $n-x$ misslyckas med slh $(1-p)$.

Låt $Y_i=\begin{cases}1, & \text{om försök } i \text{ lyckas}\\ 0, & \text{annars}\end{cases} \Rightarrow \begin{cases}E[Y_i]=0\cdot(1-p)+1\cdot p=p\\ E[Y_i^2]=0^2(1-p)+1^2\cdot p=p\end{cases}$

$\Rightarrow Var(Y_i)=E[Y_i^2]-(E[Y_i])^2=p-p^2=p(1-p)$

$X=\sum_{i=1}^n Y_i \Rightarrow E[X]=E[\sum_{i=1}^n Y_i]=\sum_{i=1}^n E[Y_i]=\sum_{i=1}^n p=np$

$Var(X)=Var(\sum_{i=1}^n Y_i)=\sum_{i=1}^n Var(Y_i)=\sum_{i=1}^n p(1-p)=np(1-p)$

UPPG 4: (a) Låt X=atomlivslängd (i sekunder)

$f(x)=\lambda e^{-\lambda x},\ x>0 \Rightarrow F(x)=1-e^{-\lambda x},\ P(X>x)=e^{-\lambda x}$

Vet att $P(X>27\cdot 3600)=0.5 \Leftrightarrow e^{-\lambda\cdot 27\cdot 3600}=0.5$

$\Leftrightarrow -\lambda\cdot 27\cdot 3600=\ln 0.5 \Rightarrow \lambda=-\frac{\ln 0.5}{27\cdot 3600}\approx 7.13\cdot 10^{-6}$

(b) Låt n= # atomer vid $t=0$, Y= # sönderfall under första sekun[den]

$\Rightarrow Y\sim Bin(n,p)$, där $p=P(X\le 1)=1-e^{-\lambda}$

$E[Y]=np=8000 \Rightarrow n=\frac{8000}{p}=\frac{8000}{1-e^{-\lambda}}=1121844986$

(c) Låt W= # atomer som ej sönderfallit efter en månad (30 da[gar])

$\Rightarrow W\sim Bin(n=1121844986, p)$,

där $p=P(X>30\cdot 24\cdot 3600)=e^{-2592000\lambda}$

$P(W=0)=(1-p)^n\approx 2.67\cdot 10^{-5}$

UPPG5: (a) $L(p) = p^{53+61}(p^2)^{24+19}\left(\frac{1}{2}(1-2p-2p^2)\right)^{17+26} = p^{200}\cdot\left(\frac{1}{2}\right)^{43}(1-2p-2p^2)$

$\ln L(p) = 200\ln p + 43\ln 0.5 + 43\ln(1-2p-2p^2)$

$\frac{\partial \ln L(p)}{\partial p} = 0 \Leftrightarrow \frac{200}{p} - \frac{86(1+2p)}{1-2p-2p^2} = 0 \Leftrightarrow 200(1-2p-2p^2) - 86p(1+2p) = 0$

$\Leftrightarrow 592p^2 + 486p - 200 = 0 \Rightarrow \hat{p} = -\frac{243}{592} + \sqrt{\left(\frac{243}{592}\right)^2 + \frac{200}{592}} \approx 0.30$

(b) $\mu = \mu(p) = \sum_{x=0}^{5} x f(x) = 0\cdot p + 1\cdot p^2 + 2\cdot\frac{1}{2}(1-2p-2p^2) + 3\cdot\frac{1}{2}(1-2p-2p^2)$
$+ 4\cdot p^2 + 5\cdot p = p^2 + 1 - 2p - 2p^2 + 1.5 - 3p - 3p^2 + 4p^2 + 5p = 2.5$

Momentmetoden skattar $E[X]$ med $\bar{X}$, dvs 2.5 med 2.585 ?!

UPPG6: Låt X = # "A-vinster" $\Rightarrow X \sim Bin(n=30, p=P(\text{"A-vinst"}))$

Om A och B lika starka är $p=0.5$, Om A starkare är $p>0.5$.

$X \overset{appr.}{\sim} N(\mu=np, \sigma^2=np(1-p)) \Rightarrow \hat{p} = \frac{X}{n} \sim N\left(\mu=p, \sigma^2=\frac{p(1-p)}{n}\right)$

$\Rightarrow P\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le 1.645\right) = 0.95 \Leftrightarrow P\left(p \ge \hat{p} - 1.645\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$

$\Rightarrow \left(\hat{p} - 1.645\sqrt{\frac{p(1-p)}{n}}, 1\right) \approx \left(\hat{p} - 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1\right)$ är ett nedåt begränsat 95% k.i. för p. Vi får intervallet $\left(\frac{21}{30} - 1.645\sqrt{\frac{21/30\cdot 9/30}{30}}, 1\right) = (0.56, 1)$, vilket tyder på att A är starkare.

UPPG7: $X_1 \sim N(\mu_1, \sigma^2)$, $X_2 \sim N(\mu_2, \sigma^2) \Rightarrow \bar{X}_1 \sim N(\mu_1, \frac{\sigma^2}{25})$, $\bar{X}_2 \sim N(\mu_2, \frac{\sigma^2}{20})$

$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1-\mu_2, \sigma^2\left(\frac{1}{25}+\frac{1}{20}\right)\right) \Rightarrow \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sigma\sqrt{1/25+1/20}} \sim N(0,1) \Rightarrow \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{25}+\frac{1}{20}}} \sim t_{43}$

Testa $H_0: \mu_1=\mu_2$ mot $H_1: \mu_1<\mu_2$. Om H_0 sann är $T = \frac{\bar{X}_1-\bar{X}_2}{S_p\sqrt{1/25+1/20}} \sim t_{23}$. Förkasta på α-nivå om $T< -t_{\alpha,23}$, dvs på 0.01-nivå om $T<-2.416$. Vi får $s_p^2 = \frac{24\cdot 2.1^2 + 19\cdot 2.48^2}{43}$ och $t = -3.24 \Rightarrow$ Förk. H_0 på 0.01-nivå.

UPPG8: (a) ε_i oberoende $\sim N(0, \sigma^2)$. $Y = X\beta + \varepsilon$, där $Y=[Y_1,\dots,Y_n]^T$, $\beta=[\beta_0,\beta_1]^T$, $\varepsilon=[\varepsilon_1,\dots,\varepsilon_n]^T$, $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\hat{\beta} = (X^TX)^{-1}X^TY$

$Var(\hat{\beta}) = (X^TX)^{-1}X^T Var(Y)\left((X^TX)^{-1}X^T\right)^T = (X^TX)^{-1}X^TX(X^TX)^{-1}\sigma^2 = (X^TX)^{-1}\sigma^2$

(b) 99% k.i. för β_1:

$\hat{\beta}_1 \pm t_{0.005,18}\cdot S\cdot\sqrt{0.00034} = 0.7249 \pm 2.878\cdot 11.3426\sqrt{0.00034} =$
$= 0.7249 \pm 0.6048 \quad (0.1203, 1.3295)$