LINKÖPINGS UNIVERSITET

Matematiska institutionen

01 juni 2020, kl. 8.00-12.00
Examinator: Jörg-Uwe Löbus

Alla hjälpmedel är tillåtna.

Betygsgränser 3: minst 8 poäng, 4: minst 11.5 poäng, 5: minst 15 poäng (av 18)

(1) Antag att avstånd mellan bilar på en landsväg (i en riktning) kan antas vara oberoende och exponentialfördelade med väntevärde 100 meter. Tänk dig att du befinner dig i en bil på denna väg.

(a) Vilken approximativ fördelning (inklusive parametrar) har avståndet till den femtionde bilen framför dig? (1p)

(b) Vad är sannolikheten att avståndet till den femtionde bilen är mellan 5 km och 5.2km? (2p)

Ledning: Använd centrala gränsvärdessatsen.

(2) Antag att andelen bilister som är alkoholpåverkade en lördagskväll är 0.4%. En viss typ av alkotest ger utslag i 99% av fallen då testpersonen verkligen är berusad. Sannolikheten är även 0.6% att testet ger utslag givet att testpersonen är nykter. Polisen stoppar slumpmässigt en bilist en lördagskväll.

(a) Vad är sanolikhet att alkotestet ger utslag. (1.5p)

(b) Om alkotestet ger utslag, hur stor är sannolikheten att bilisten verkligen är berusad? (1.5p)

(3) Två bussar med olika destinationer anländer oberoende och likformigt fördelade mellan kl. 8.05 och kl. 8.10 på en viss busshållplats. Bussarna stannar precis två minuter innan de åker iväg. Beräkna sannolikheten att det finns ett tidsintervall av minst 1 minut då båda bussarna stannar samtidigt på busshållplatsen. (3p)

(4) I en undersökning om belysningens effekt på den mänskliga förmågan att genomföra koncentrationskrävande arbetsuppgifter gjordes följande experiment. Nio personer med normal, men varierande, synförmåga fick som sin uppgift att så snabbt som möjligt dra en tunn tråd genom tio nålsögon. Detta gjordes både mot svart bakgrund under lägre ljusstyrka och mot vit bakgrund under högre ljusstyrka. Man mätte tiden att fullborda uppgiften. Resultatet (enhet: sek) blev:

Person nr	1	2	3	4	5	6	7	8	9
Högre ljusstyrka	25.85	28.84	32.05	25.74	20.89	41.05	25.01	24.96	27.47
Lägre ljusstyrka	18.23	20.84	22.96	19.68	19.50	24.98	16.61	16.07	24.59

Vi antar som modell att observationerna är oberoende, normalfördelade med samma varians. Vi antar också att de förväntade tidsskillnaderna är desamma ($= \Delta$) för varje person.

(a) Ange det generella (tvåsidiga) konfidensintervallet för väntevärdet Δ av differensen av parvisa mätningar. (1p)

(b) Gör ett 99.9% :igt tvåsidigt konfidensintervall för denna skillnad Δ. (1.5p)

(c) Får man hävda att det finns en statistiskt säkerställd skillnad mellan de båda mätningarna på konfidensnivå 99.9%? (0.5p)

(5) De två sämsta segertiderna på Stockholm Marathon har registrerats under de varmaste dagarna (1982 och 2007). I nedanstående tabell finns temperatur och segerrtid (i minuter) till och med 2008.

År	Temp	Tid	År	Temp	Tid
1979	19	137.58	1994	17	134.10
1980	17	135.82	1995	17	134.48
1981	19	133.43	1996	17	135.08
1982	31	139.33	1997	26	137.37
1983	15	131.62	1998	19	136.20
1984	19	133.78	1999	23	134.87
1985	22	137.30	2000	16	138.82
1986	11	132.55	2001	17	138.28
1987	7	133.87	2002	25	138.33
1988	11	134.43	2003	16	138.23
1989	13	133.57	2004	18	136.20
1990	19	133.07	2005	13	133.50
1991	12	132.63	2006	22	137.02
1992	28	135.97	2007	29	140.93
1993	22	136.50	2008	25	136.05

Låt x_i beteckna temperatur och y_i segertid och anta att data följer modellen enkel linjär regression (Stockholms universitet)

$$Y_i = \alpha + \beta x_i + \varepsilon_i\,, \quad i = 1, 2, \ldots, n\,,$$

där $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ är oberoende och normalfördelade med väntevärde 0 och varians σ^2.

(a) Beräkna den skattade regressionslinjen $y = \alpha^*_{obs} + \beta^*_{obs} x$. (2p)

Räknehjälp: $\sum x_i = 565$, $\sum y_i = 4070.91$, $\sum x_i^2 = 11577$, $\sum y_i^2 = 552563.70$, $\sum x_i y_i = 76909.63$.

(b) Antag att temperaturen under Stockholm Marathon nästa år är $x_0 = 10\ ^oC$. Ange en punktprediktion för segertiden Y_0 vid temperaturen x_0. (1p)

(6) Vi har observationer från n oberoende identiskt Poisson fördelade slumpvariabler $X_1, \ldots, X_n$ med parameter $\lambda > 0$, dvs.

$$P(X_1 = x) = e^{-\lambda}\frac{\lambda^x}{x!}\,, \quad x = 0, 1, \ldots\,.$$

(a) Härled Maximum-Likelihood-skattningen för λ. (2p)

(b) Avgör om skattaren i (a) är väntevärdesriktig. (1p)

(c) Avgör om skattaren i (a) är konsistent. (1 extrappoäng)

Lösningar

(1) Låt X_i beteckna avståndet mellan bil $i-1$ och bil i framför oss. Eftersom $E[X_i] = 0.1$ km och X_i är exponentialfördelad så gäller $X_i \sim Exp(10)$. Således $E[X_i] = 0.1$ km och $Var(X_i) = 0.01$ km^2.

(a) Centrala gränsvärdessatsen medför att $\sum_{i=1}^{50} X_i$ är approximativt normalfördelad med$E\left[\sum_{i=1}^{50} X_i\right] = 50 \cdot 0.1 = 5$ km and $Var\left(\sum_{i=1}^{50} X_i\right) = 50 \cdot 0.1^2$ km^2 där vi tar hänsyn till att $X_1, \ldots, X_{50}$ är oberoende.

(b)

$$P\left(5 \leq \sum_{i=1}^{50} X_i \leq 5.2\right) = P\left(\frac{5-5}{\sqrt{50} \cdot 0.1} \leq \frac{\sum_{i=1}^{50} X_i - 5}{\sqrt{50} \cdot 0.1} \leq \frac{5.2-5}{\sqrt{50} \cdot 0.1}\right)$$
$$= P(0 \leq Z \leq 0.2828) = 0.6110 - 0.5 = 0.1110$$

där $Z = \frac{\sum_{i=1}^{50} X_i - 5}{\sqrt{50} \cdot 0.1} \sim N(0,1)$.

(2) Låt A beteckna händelsen att föraren är alkoholpåverkad och låt S beteckna händelsen att föraren inte är alkoholpåverkad. Låt D vara händelsen att alkotestet ger utslag.

(a) $P(D) = P(D|A)P(A) + P(D|S)P(S) = 0.99 \cdot 0.004 + 0.006 \cdot 0.996 \approx 0.01$ pga lagen om total sannolikhet.

(b) Bayes' formel medför att

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.99 \cdot 0.004}{0.01} \approx 0.40\,.$$

(3) Låt

X : ankomsttid av bussen med destination A, i minuter efter kl. 8.05

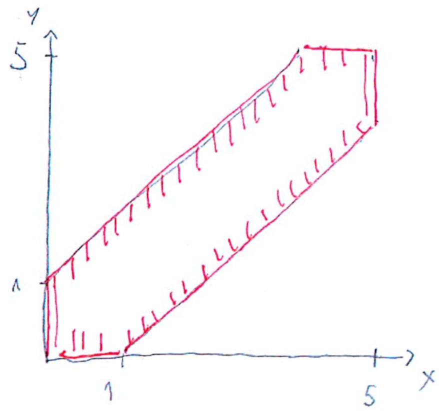
och

Y : ankomsttid av bussen med destination B, i minuter efter kl. 8.05

Oberoendet av X och Y medför att

$$f(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{5} \cdot \frac{1}{5} & \text{om} \quad 0 \leq x \leq 25, \quad 0 \leq y \leq 5 \\ 0 & \text{annars} \end{cases} .$$

Sökt är $P(\{|X - Y| \leq 1\}$.



$\{(x,y) \in [0,5] \times [0,5] \;:\; |x-y| \leq 1\}$

Vi får

$$P(\{|X-Y| \leq 1) \\ = \cdot \frac{\text{markerad area}}{\text{total area}} = \frac{25 - 2 \cdot 4 \cdot 4/2}{25} = \frac{9}{25}.$$

(4) (a) Vi betraktar de nio skillnaderna som nio oberoende utfall av en slumpvariabel, vars väntevärde vi önskar ett konfidensintervall för. Medelvärdet antas vara normalfördelat, så att vi kan beräkna konfidensintervallets gränser ur formeln

$$I_\Delta = \left(\bar{z} - t(n-1)_{\alpha/2} \cdot \frac{s_z}{\sqrt{n}}, \, \bar{z} + t(n-1)_{\alpha/2} \cdot \frac{s_z}{\sqrt{n}}\right).$$

(b) Detta är hypotesprövning med stickprov i par. Vi betecknar tiderna under lägre ljusstyrka med x_i och tiderna under högre ljusstyrka med y_i. Vi bildar skillnaden mellan dessa som $z_i = y_i - x_i$, $i = 1, \ldots, 9$. Vi ser z_i som utfall av oberoende normalfördelade stokastiska variabler $Z_i \sim N(\Delta, \sigma^2)$, $i = 1, \ldots, 9$, respektive. Vi bildar ett konfidensintervall för Δ. Det aritmetiska medelvärdet är en punktskattning av *Delta* med

$$\bar{z} = \frac{1}{9} \sum_{i=1}^{9} z_i = 7.60\,.$$

Vi skattar σ med

$$s_z = \sqrt{\frac{1}{8} \sum_{i=1}^{9} (z_i - \bar{z})^2} = 4.1779\,.$$

Det sökta konfidensintervallet är av formen

$$I_\Delta = \left(\bar{z} - t_{0.0005}(8) \cdot \frac{s_z}{\sqrt{3}}, \, \bar{z} + t_{0.0005}(8) \cdot \frac{s_z}{\sqrt{3}}\right).$$

där $t_{0.0005}(8)$ är 0.0005-kvantilen för t-fördelningen med 8 frihetsgrader. Tabellslagning ger $t_{0.0005}(8) = 5.04$. Med insättning fås $I_\Delta = (0.58, 14.6)$.

<u>*Anmärkning:*</u> Om man väljer att beteckna tiderna under lägre ljusstyrka med y_i och tiderna under högre ljusstyrka med x_i så får man $I_\Delta = (-14.6, -0.58)$.

(c) Eftersom $0 \notin I_\Delta$, så är det med konfidensgrad 99,9% statistiskt säkerställt att det finns en skillnad.

(5) (a) Enligt formeln ges skattningarna av α^*_{obs} och β^*_{obs} av

$$\alpha^*_{obs} = 130.85\,, \quad \beta^*_{obs} = 0.257$$

som betyder att regressionslinjen är

$$y = 130.85 + 0.257x\,.$$

Med andra ord ökar segertiden i genomsnitt med 0.257 minuter (15 sekunder) för varje temperaturökning med 1^oC. Anmärkning: Avrundningsfel för α^*_{obs} kan vara stor.

(b)

$$y^*_{0\,obs} = 130.85 + 0.257x_0 = 133.420\,.$$

(6) (a) Likelihoodfunktionen blir

$$L(\lambda) = \prod_{j=1}^{n} e^{-\lambda}\frac{\lambda^{x_j}}{x_j!} = e^{-n\lambda}\,\frac{\lambda^{\sum_{j=1}^n x_j}}{\prod_{j=1}^n x_j!}$$

och

$$\ln L(\lambda) = -n\lambda + \ln\lambda\cdot\sum_{j=1}^{n} x_j - \sum_{j=1}^{n}\ln(x_j!)\,.$$

Derivation ger

$$\frac{d\ln L(\lambda)}{d\lambda} = \frac{\sum_{j=1}^n x_j}{\lambda} - n\,.$$

Av $\frac{d\ln L(\lambda)}{d\lambda} = 0$ fås

$$\lambda^*_{obs} = \bar{x}\,.$$

Eftersom $\ln L(\lambda) = -\infty$ då $\lambda \to 0$ eller $\lambda \to \infty$ så följer att extremvärdet är ett maximum och att $\lambda^*_{obs} = \bar{x}$ är ML-skattningen av λ.

(b) Skattaren är väntevärdesriktig eftersom

$$E[\lambda^*] = E\left[\overline{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\cdot E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\cdot n\lambda = \lambda\,.$$

(c) Vi har visat att skattaren $\lambda^* = \overline{X}$ är väntevärdesriktig. Dessutom gäller det att

$$Var(\lambda^*) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\lambda}{n} \longrightarrow 0$$

där $n \to \infty$. Detta medför att skattaren är konsistent.