Matematisk statistik TMS064/TMS063
Tentamen 2019-08-20

Tid: 14:00-18:00 **Tentamensplats:** SB
Hjälpmedel: Bifogad formelsamling och tabell samt Chalmersgodkänd räknare.
Kursansvarig: Olof Elias
Telefonvakt/jour: Olof Elias, 0762026293.
Betygsgränser: För betyg **3**, **4** resp. **5** krävs minst **12**, **18** resp. **24** poäng.
För att bli godkänd på kursen krävs godkänt på tentan **och** godkänt resultat på båda flervariabel-duggorna.
Till varje uppgift skall fullständig och välmotiverad lösning lämnas!

---

1. Låt A och B vara två godtyckliga händelser i ett utfallsrum Ω. Avgör om följande påståenden är sanna. **Ange endast svar** .

   (a) För alla händelser A, B så gäller: $(A \cap B)^c = A^c \cap B^c$.

   (b) För alla händelser A, B så gäller: $P(A \cup B) \leq P(A) + P(B)$

   (c) $A \cap B = \emptyset$ medför att A och B är oberoende.

   (d) $P(A|B) \leq P(A) \Rightarrow P(B|A) \leq P(B)$

   1 poäng för varje rätt svar, -1 poäng för varje felaktigt svar. Den totala poängen på uppgiften kan ej bli mindre än 0

   Lösning:

   (a) Falskt: $(A \cap B)^c = A^c \cup B^c$ (Rita ett Venn diagram för att se detta)

   (b) Sant: $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$ ty $P(A \cap B) \geq 0$

   (c) Falskt: $A \cap B = \emptyset$ ger $0 = P(A \cap B) = P(A)P(B)$ vilket implicerar att A eller B måste ha sannolikhet 0.

   (d) Sant: Detta fås av följande : $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ vilket ger att om $P(A|B) \leq P(A)$ så är högerledet givetvis mindre än $P(B)$.

2. Låt X vara en *kontinuerlig* stokastisk variabel med täthet

$$f_X(t) = \frac{c_X}{1+t}, t \in [0, 5]$$

   och låt Y vara en *diskret* stokastisk variabel med täthet

$$f_Y(s) = \frac{c_Y}{1+s}, s = 0, 1, 2, 3, 4, 5.$$

   Beräkna

   (a) c_X och c_Y. **1 +1 poäng**

   (b) $P(2 < X < 5)$ och $P(2 < Y < 5)$. **1 +1 poäng**

   (c) $E[X]$ och $E[Y]$. **1 +1 poäng**

Lösning :

(a) Konstanterna fås av följande:

$$c_X = \left(\int_0^5 \frac{1}{1+s}\mathrm{d}s\right)^{-1} = 1/\log(6),\ c_Y = \left(\sum_{i=0}^{5} 1/(1+i)\right)^{-1} = 1/2.45.$$

(b) För X så får vi

$$P(2 < X < 5) = \frac{1}{\log(6)}\int_2^5 \frac{1}{1+s}\mathrm{d}s = \frac{\log(2)}{\log(6)}.$$

För Y gäller

$$P(2 < Y < 5) = P(3 \leq Y \leq 4) = \frac{1}{2.45}\left(\frac{1}{1+3} + \frac{1}{1+4}\right) = 0.45/2.45.$$

(c) Väntevärdet av X fås av

$$E[X] = \frac{1}{\log(6)}\int_0^5 \frac{s}{1+s}\mathrm{d}s = \frac{1}{\log(6)}\int_0^5 \frac{1+s}{1+s} - \frac{1}{1+s}\mathrm{d}s = \frac{5}{\log(6)} - 1.$$

Liknande så fås väntevärdet av Y av

$$E[Y] = \frac{1}{2.45}\sum_{i=0}^{5}\frac{i}{1+i} = \frac{3.55}{2.45}.$$

3. Låt $Z = (X, Y)$ vara en stokastisk vektor som defineras enligt följande procedur.

Låt X vara Bernoulli-fördelad med parameter $p \in (0, 1)$, d.v.s X har marginaltäthet som ges av:

$$P(X = 0) = p, P(X = 1) = 1 - p.$$

Vidare, låt Y defineras av följande:

- Givet att $X = 0$ så är Y exponential-fördelad med parameter $\lambda = 1$.
- Givet att $X = 1$ så är Y likformigt fördelad på intervallet $[0, 1]$.

Beräkna

(a) $P(Y \geq 1, X = 0)$ **1 poäng**

(b) Marginaltätheten av Y **3 poäng**
(Ledning: Använd totala sannolikhetslagen)

(c) Väntevärdet av Y **1 poäng**

Lösning:

(a) Bayes ger

$$P(Y \geq 1, X = 0) = P(Y \geq 1|X = 0)P(X = 0) = \mathrm{e}^{-1}p$$

då $P(Y \geq 1|X = 0)$ är exponential fördelad med parameter $\lambda = 1$.

(b) Vi söker :

$$P(Y \leq y) = P(Y \leq y|X = 0)P(X = 0) + P(Y \leq y|X = 1)P(X = 1).$$

Om $y < 0$ så är högerledet 0. Om $y > 1$ så är $P(Y \leq y|X = 1) = 1$ vilket ger att för $y > 1$ så har vi

$$P(Y \leq y) = (1 - \mathrm{e}^{-y})p + (1 - p).$$

För $y \in [0, 1]$ så tillkommer endast fördelningsfunktionen för likformiga fördelningen:

$$P(Y \leq y) = (1 - \mathrm{e}^{-y})p + y(1 - p).$$

Sammanställer man allt så får man följande

$$P(Y \leq y) = \begin{cases} 0 & y < 0 \\ (1 - \mathrm{e}^{-y})p + y(1 - p) & y \in [0, 1] \\ (1 - \mathrm{e}^{-y})p + (1 - p) & y > 1 \end{cases}$$

Vilket efter derivering ger täthetsfunktion

$$f(y) = \frac{\mathrm{d}}{\mathrm{d}y} P(Y \leq y) = \begin{cases} 0 & y < 0 \\ \mathrm{e}^{-y}p + (1 - p) & y \in [0, 1] \\ \mathrm{e}^{-y}p & y > 1 \end{cases}$$

(c) Väntevärdet fås m.h.a (b) :

$$\begin{aligned} E[Y] &= \int_0^\infty y f(y)\mathrm{d}y = \int_0^1 py\mathrm{e}^{-y} + y(1 - p)\mathrm{d}y + \int_1^\infty py\mathrm{e}^{-y}\mathrm{d}y \\ &= p\frac{\mathrm{e} - 2}{\mathrm{e}} + (1 - p) + p\frac{2}{\mathrm{e}} = p + (1 - p) = 1. \end{aligned}$$

4. (a) Låt $x = (x_1, x_2, ..., x_n)$ vara ett observerat stickprov från Geom(p). Härled maximumlikelihoodskattaren av p. **4 poäng**

(b) Är din skattare väntevärdesriktig? Motivera. **1 poäng**
(Ledning: räcker att titta på fallet $n = 1$)

Lösning

(a) log-likelihoodfunktionen ges av

$$l(p|x) = n\log(p) + \left(\sum_1^n x_i - n\right)\log(1 - p)$$

vilket ger att maximum fås av den stationära punkten

$$l'(p|x) = 0 \Leftrightarrow \frac{n}{p} - \frac{\sum_1^n x_i - n}{1 - p} = 0 \Rightarrow p_{MLE} = n/\sum_1^n x_i.$$

(b) Av ledningen så ser vi att för $n = 1$ så måste

$$\mathrm{E}[1/X] = p$$

gälla, för $X \sim \mathrm{Geom}(p)$. Detta är samma sak som

$$\sum_{k=1}^{\infty} \frac{1}{k} p(1-p)^{k-1} = p$$

Vi ser att vänsterledet kan skrivas som

$$\frac{1}{1} p(1-p)^{1-1} + \sum_{k=2}^{\infty} \frac{1}{k} p(1-p)^{k-1} = p + \sum_{k=2}^{\infty} \frac{1}{k} p(1-p)^{k-1} > p$$

vilket implicerar att vår skattare ej är väntevärdesriktig.

5. Ett företag ska konstruera hissarna till Nya Karolinska. Som ett led i konstruktionen behöver man därav känna till den genomsnittliga vikten, μ, hos en besökare samt proportionen, p, av alla besökare som väger minst 95 kg.

Man drar ett stickprov, x, på 100 personer och skattar följande kvantiter

$$\bar{x} = 74.8, s = 15.2, n_{95} = 9$$

där n_{95} betecknar antalet personer vars vikt överskrider 95 kg.

(a) Antag att vikten hos en person är normalfördelad med parametrar μ och σ. Beräkna ett två-sidigt konfidensinterval för μ vid en signifikansnivå 0.01. **2 poäng**

(b) Beräkna ett två-sidigt intervall för proportionen p vid en signifikansnivå på 0.005. **2 poäng**

(c) Testa hypotesen för väntevärde

$$H_0 : \mu = 77$$

med hjälp av konfidensintervallen i (a) . **1 poäng**

Lösning:

(a) Ett två-sidigt konfidensintervall ges av

$$\bar{x} \pm t_{\alpha/2}(100-1)\frac{s}{\sqrt{n}} \approx 74.8 \pm 2.617\frac{15.2}{\sqrt{100}} \approx [70.8, 78.7]$$

(b) Motsvarande konfidensintervall för p ges av

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = \frac{9}{100} \pm 2.86\sqrt{\frac{9}{100}\left(1 - \frac{9}{100}\right)/100} \approx [0.0081, 0.17]$$

(c) Vi ser att $77 \in [70.8, 78.7]$ vilket ger att vi inte kan förkasta H_0.

6. Låt $X_1, ..., X_n$ vara ett stickprov draget från en stokastisk variabel X med $E[X] = \mu$ och $\text{Var}(X) = 1$. Antag att vi har följande hypoteser

$$H_0 : \mu = 1,$$
$$H_1 : \mu < 1,$$

samt följande test:

Förkasta H_0 om

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \leq \gamma,$$

där γ är en fix parameter.

(a) Vad blir testets (approximativa) signifikansnivå om $n = 40$ och $\gamma = 0.8$? **3 poäng**

(b) Vad blir testets (approximativa) styrka om det sanna väntevärdet är $\mu = 0.90$ och $n = 60$ och $\gamma = 0.8$? **2 poäng**

(Ledning: Använd centrala gränsvärdessatsen)
Lösning:

(a) Den approximativa signifikansnivån ges av

$$\alpha = P\left(\text{Förkasta } H_0 | H_0 \text{ sann}\right) = P\left(\bar{X} \leq \gamma | \mu = 1\right)$$

Centrala gränsvärdessatsen säger nu att

$$\bar{X} \overset{approx}{\sim} N\left(\mu, \frac{1}{n}\right)$$

vilket ger att signifikansnivån är lika med:

$$\alpha \approx P\left(Z \leq \sqrt{n}(\gamma - \mu)\right) = P(Z \leq -1.2649) \approx 0.1.$$

(b) Testets approximativa styrka vid $\mu = 0.90$ ges av

$$1 - \beta = P\left(\text{Förkasta } H_0 | H_0 \text{ falsk}\right) = P\left(\bar{X} \leq \gamma | \mu = 0.90\right)$$

I likhet med (a) så fås att centrala gränsvärdessatsen ger att

$$\bar{X} \overset{approx}{\sim} N\left(\mu, \frac{1}{n}\right)$$

Styrkan ges därmed i likhet med ovanstående uträkning av

$$1 - \beta \approx P\left(Z \leq \sqrt{n}(\gamma - \mu)\right) = P(Z \leq -0.7746) \approx 0.22$$