Tentamenskrivning för TMS063, Matematisk Statistik.
Onsdag fm den 28 maj, 2014, V-huset.
Examinator: Marina Axelson-Fisk. Tel matstat: 070-2288113.
Tillåtna hjälpmedel: typgodkänd miniräknare, tabell- och formelhäfte (utdelas).

****

Tentan är på totalt 50 poäng: 40 poäng på matstat-delen, 10 poäng på flervariabel-delen. För att bli godkänd krävs godkänt på båda delarna. Betygsgränser för betyg 3, 4 och 5 är 40%, 60% och 80%, respektive. Betygen på de båda delarna vägs samman. **Fullständiga och välmotiverad lösningar skall ges till varje uppgift.**

Del I: Matematisk Statistik

1. För tre händelser A, B och C är följande givet: $P(A) = P(B) = P(C) = 0.5$ och $P(A \cap B) = 0.3$, och händelse C är oberoende av A, B och $A \cap B$.
   (a) Vad är sannolikheten att ingen av de tre händelserna inträffar? (2p)
   (b) Vad är sannolikheten att exakt en av de tre händelserna inträffar? (2p)
   (c) Vad är sannolikheten att åtminstone en av A och B inträffar, men inte C? (2p)
2. Antag att X är en stokastisk variabel med täthetsfunktion $f(x) = k(1 - x^2)$, $-1 \leq x \leq 1$.
   (a) Bestäm konstanten k. (3p)
   (b) Vad har X för väntevärde? (3p)
   (c) Vad har X för varians? (3p)
   (d) Antag att vi drar ett stickprov på $n = 20$ från X:s fördelning. Vad är den approximativa sannolikheten att stickprovsmedelvärdet ligger mellan -0.1 och 0.1? (4p)
3. Den Binomialfördelade variabeln X har väntevärde 5 och varians 4.
   (a) Bestäm parametrarna n och p. (2p)
   (b) Beräkna sannolikheten $P(X \geq 3)$. (3p)
4. En tillverkare producerar enheter i boxar om två. Känt är att
   - ca 95% av boxarna innehåller två felfria enheter.
   - ca 4% av boxarna innehåller exakt en defekt enhet.
   - i ca 1% av boxarna är båda enheterna defekta.

   (a) En box väljs slumpmässigt, och ur denna väljs en enhet slumpmässigt. Givet att den valda enheten är defekt, vad är sannolikheten att den återstående enheten är felfri? (2p)
   (b) Givet att den valda enheten är felfri, vad är sannolikheten att den återstående är defekt? (2p)
5. Man IQ-testade ett antal elever och fick resultaten för $n = 10$, att $\bar{x} = 109.4$ och $s = 26.81$. Antag normalfördelning.
   (a) Konsturera ett 95% konfidensintervall för medel-IQ. (3p)
   (b) Om vi nu antar att $\sigma = 25$, hur stort stickprov skulle behövas för att konstruera ett 99% konfidensintervall av samma längd som i (a)? (3p)

Var god vänd!

6. Låt X vara en stokastisk variabel som kan anta värdena $\{0, 1, 2, 3\}$. Antag att ett stickprov på $n = 800$ har givit 75 nollor, 206 ettor, 220 tvåor och 299 treor. Pröva hypotesen att X har fördelningen:

   $$\begin{aligned} P(X = 0) &= 0.1 \\ P(X = 1) &= 0.2 \\ P(X = 2) &= 0.3 \\ P(X = 3) &= 0.4 \end{aligned}$$

   på signifikansnivå $\alpha = 0.01$. (6p)

Del II: Flervariabelanalys

7. Bestäm tangentplanet till funktionsytan $z = \arctan(xy)$ genom punkten $(2, \frac{1}{2}, \frac{\pi}{4})$. (3p)

8. Beräkna $\iint_D \sqrt{x^2 + y^2}\, dxdy$,

   då D är den del av cirkelskivan $x^2 + y^2 \leq 3$ där $y \geq 0$. (4p)

9. Antag att en partikels position i xy-planet vid en tidpunkt t ges av $\mathbf{r}(t) = t^3\mathbf{i} + t\mathbf{j}$. Skissa partikelns rörelsebana för $-1 \leq t \leq 1$ och ange partikelns fart vid varje tidpunkt under detta tidsintervall. (3p)

Lösningar

Del I: Matematisk Statistik

1. Rita Venn-diagram över uppgifterna så blir det lättare. Oberoendena ger att $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$ och $P(A \cap B \cap C) = P(A \cap B)P(C)$.

   (a) Vi söker sannolikheten P(ingen händelse inträffar), dvs regionen utanför alla tre händelserna A, B och C. Dvs $P((A \cup B \cup C)^c) = 1 - P(A \cup B \cup C)$. Vi får

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A)P(C) - P(B)P(C) \\ &\quad + P(A \cap B)P(C) \\ &= 0.5 + 0.5 + 0.5 - 0.3 - 0.5 \cdot 0.5 - 0.5 \cdot 0.5 + 0.3 \cdot 0.5 \\ &= 0.85 \end{aligned}$$

   Svar: $P(\text{ingen händelse}) = 1 - 0.85 = 0.15$

   (b) P(exakt en händelse) motsvaras av regionerna i A, B och C som inte överlappar med någon annan händelse. Dvs

$$\begin{aligned} P(\text{bara } A) &= P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) \\ &= 0.5 - 0.3 - 0.5 \cdot 0.5 + 0.3 \cdot 0.5 = 0.10 \\ P(\text{bara } B) &= P(B) - P(A \cap B) - P(B \cap C) + P(A \cap B \cap C) \\ &= 0.5 - 0.3 - 0.5\dot{0}.5 + 0.3 \cdot 0.5 = 0.10 \\ P(\text{bara } C) &= P(C) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= 0.5 - 0.5 \cdot 0.5 - 0.5 \cdot 0.5 + 0.3 \cdot 0.5 = 0.15 \end{aligned}$$

   Så vi får

$$\begin{aligned} P(\text{exakt en}) &= P(\text{bara } A) + P(\text{bara } B) + P(\text{bara } C) \\ &= 0.10 + 0.10 + 0.15 = 0.35 \end{aligned}$$

   Alternativt får man samma svar med hjälp av

$$\begin{aligned} P(\text{exakt en händelse}) &= P(A \cup B \cup C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + 2P(A \cap B \cap C) \end{aligned}$$

   Svar: $P(\text{exakt en händelse}) = 0.35$

   (c) A eller B, men inte C. Vi använder att $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

$$\begin{aligned} &P(A \text{ eller } B, \text{ men inte } C) = \\ &= P(A \cup B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= P(A) + P(B) - P(A \cap B) - P(A)P(C) - P(B)P(C) + P(A \cap B)P(C) \\ &= 0.5 + 0.5 - 0.3 - 0.5 \cdot 0.5 - 0.5 \cdot 0.5 + 0.3 \cdot 0.5 \\ &= 0.35. \end{aligned}$$

   Svar: $P(A \text{ eller } B, \text{ men inte } C) = 0.35$

2. (a) För att $f(x)$ ska vara en täthetsfunktion måste den integrera sig till 1. Alltså vill vi hitta det värde på k som uppfyller detta.

$$\begin{aligned} \int_{-1}^{1} k(1 - x^2)dx &= k\left[x - \frac{x^3}{3}\right]_{-1}^{1} \\ &= k\left(1 - \frac{1^3}{3} - \left((-1) - \frac{(-1)^3}{3}\right)\right) \\ &= k\frac{4}{3} \end{aligned}$$

   Så för att få $\int f(x)\,dx = 1$ måste $k = \frac{3}{4}$.
   Svar: $k = 3/4$.

(b) Väntevärdet ges av

$$E[X] = \int_{-1}^{1} x\, f(x)\, dx = \int_{-1}^{1} \frac{3}{4} x(1-x^2)\, dx = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^4}{4}\right]_{-1}^{1} = 0$$

Svar: $E[X] = 0$

(c) Variansen för X får vi enklast med hjälp av formeln

$$V(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \int_{-1}^{1} \frac{3}{4} x^2(1-x^2)\, dx = \frac{3}{4}\left[\frac{x^3}{3} - \frac{x^5}{5}\right]_{-1}^{1} = \frac{3}{4}\left(\frac{2}{3} - \frac{2}{5}\right) = \frac{1}{5}$$

Svar: $V(X) = 1/5 - 0 = 1/5 = 0.2$

(d) Centrala gränsvärdessatsen ger oss att stickprovsmedelvärdet $\bar{X}$ är approximativt normalfördelat med väntevärde $E[X] = 0$ och varians $V(X)/n = 0.4/20 = 0.02$. Dvs

$$Z = \frac{\bar{X} - 0}{\sqrt{0.02}} \sim N(0,1)$$

Vi får

$$\begin{aligned}
P(-0.1 \le \bar{X} \le 0.1) &= P(\frac{-0.1}{\sqrt{0.02}} \le Z \le \frac{0.1}{\sqrt{0.02}}) \\
&= P(-0.707 \le Z \le 0.707) \\
&= 2 \cdot \Phi(0.707) - 1 \\
&= 2 \cdot 0.55 - 1 = 0.1
\end{aligned}$$

Svar: $P(-0.1 \le \bar{X} \le 0.1) \approx 0.1$

3. (a) $Bin(n,p)$ har väntevärde np och varians $np(1-p)$ så vi behöver lösa ekvationssystemet:

$$\begin{aligned}
np &= 5 \\
np(1-p) &= 4
\end{aligned}$$

Svar: $p = 0.2$ och $n = 25$

(b) Här kan man välja om man vill göra normalapproximation eller inte. Exakt beräkning ger

$$\begin{aligned}
P(X \ge 3) &= 1 - P(X < 3) \\
&= 1 - \binom{n}{0} p^0(1-p)^n - \binom{n}{1} p^1(1-p)^{n-1} - \binom{n}{2} p^2(1-p)^{n-2} \\
&= 1 - 0.8^{25} - 25 \cdot 0.2 \cdot 0.8^{24} - \frac{25 \cdot 24}{2} \cdot 0.2^2 \cdot 0.8^{23} \\
&= 1 - 0.098 = 0.902
\end{aligned}$$

Svar: $P(X \ge 3) = 0.902$

4. Vi definierar händelserna

$$\begin{aligned}
A_1 &= \text{första är felfri} \\
A_2 &= \text{andra är felfri}
\end{aligned}$$

(a)

$$\begin{aligned}
P(\text{andra är felfri} \,|\, \text{första är defekt}) &= P(A_2 | A_1^c) \\
&= \frac{P(A_1^c \cap A_2)}{P(A_1^c)} \\
&= \frac{0.04}{0.04 + 0.01} = 0.8
\end{aligned}$$

Svar: Sannolikheten är 0.8.

(b)

$$\begin{aligned} P(\text{andra är defekt} \,|\, \text{första är felfri}) &= P(A_2^c|A_1) \\ &= \frac{P(A_1 \cap A_2^c)}{P(A_1)} \\ &= \frac{0.04}{0.95+0.04} \\ &= 0.0404 \end{aligned}$$

Svar: Sannolikheten är 0.0404.

5. (a) Vi har fått en skattning för s och antar alltså att σ är okänd. Dvs vi får använda t-fördelningen i konfidensintervallet. För $1-\alpha = 0.95$ och $n = 10$ ska vi slå upp värdet för $t_{\alpha/2,n-1} = t_{0.025,9} = 2.262$. Konfidensintervallet blir

$$\begin{aligned} \mu &= \bar{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \\ &= 109.4 \pm 2.262 \frac{26.81}{\sqrt{10}} \\ &= 109.4 \pm 19.177 \end{aligned}$$

Svar: $\mu = 109.4 \pm 19.177$ eller $90.22 \le \mu \le 128.58$.

(b) Nu är alltså σ känt, så vi kan använda normalfördelningen. Nu är $1-\alpha = 0.99$ så vi slår upp tabellvärdet $z_{\alpha/2} = z_{0.995} = 2.58$. Vi vill hitta det värde n som gör att

$$z_{0.995} \frac{\sigma}{\sqrt{n}} \le 19.177$$

dvs

$$n \ge \left(z_{0.995} \frac{\sigma}{19.177}\right)^2 = \left(2.58 \frac{25}{19.177}\right)^2 = 11.31$$

Svar: n måste vara minst 12.

6. För detta använder vi ett Goodness-of-fit-test, med statistika

$$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-p-1}$$

Vi har $n = 800$ observationer, $k = 4$ kategorier och har fått observationerna

$$O_1 = 75, O_2 = 206, O_3 = 220, O_4 = 299$$

Förväntat antal för den givna fördelningen är $E_i = n \cdot p_i$ där $p_1 = P(X = 0), ..., p_4 = P(X = 3)$, exempelvis $E_1 = 800 \cdot 0.1$. Vi får

	$(X=0)$	$(X=1)$	$(X=2)$	$(X=3)$
Kategori i:	1	2	3	4
Observerat antal O_i:	75	206	220	299
Förväntat antal E_i:	80	160	240	320
Differens $(O_i - E_i)$:	-5	46	-20	-21

Statistikan blir $X_0^2 = 16.58$.

Vi förkastar hypotesen att våra observationer kommer från den här fördelningen om $X_0^2 > \chi^2_{\alpha,k-p-1}$. Här är $p = 0$ eftersom vi inte behövde skatta några parametrar. $\alpha = 0.01$. Vi får från tabell $\chi^2_{0.01,3} = 11.34$. Och $X_0^2 > 11.34$ så vi förkastar hypotesen.
Svar: Vi förkastar hypotesen att variabeln X kommer från den givna fördelningen.

Del II: Flervariabelanalys

7. Med $f(x,y) = \arctan(xy)$ är $f_1'(x,y) = \dfrac{y}{1+(xy)^2}$ och $f_2'(x,y) = \dfrac{x}{1+(xy)^2}$
   så $f(2,\frac{1}{2}) = \frac{\pi}{4}$, $f_1'(2,\frac{1}{2}) = \frac{1}{4}$, $f_2'(2,\frac{1}{2}) = 1$, och tangentplanet kan således beskrivas med ekvationen $\underline{z = \frac{\pi}{4} + \frac{1}{4}(x-2) + (y - \frac{1}{2})}$ $\left(\Leftrightarrow\ z = \frac{\pi}{4} - 1 + \frac{1}{4}x + y\right)$

8. Övergång till polära koordinater ger att;

$$\iint_D \sqrt{x^2+y^2}\,dxdy = \int_0^{\sqrt{3}} \left(\int_0^{\pi} r \cdot r\,d\theta\right) dr = \pi\left[\frac{1}{3}r^3\right]_0^{\sqrt{3}} = \underline{\sqrt{3}\pi}$$

9. Partikelns hastighet är $\mathbf{r}'(t) = 3t^2\mathbf{i} + \mathbf{j}$, så dess fart är $v(t) = |\mathbf{r}'(t)| = \underline{\sqrt{9t^4+1}}$
   Vidare har partikelns rörelsebana följande form;