

FACIT: Tentamen L9MA30, LGMA30

Matematik 3 för lärare, åk 7-9, Sannolikhetslära och statistik,
Matematik 3 för gymnasielärare, Sannolikhetslära och statistik

2016-08-22 kl. 14.00-18.00

Examinator och Jour: Henrike Häbel, tel. 073 0671252 alt. 031 772 8296

Hjälpmedel: Valfri miniräknare (tomt minne och ingen möjlighet att koppla upp sig till nätet; kontrolleras av examinatorn), (dubbelsidigt) A4 ark med egna anteckningar samt med skrivningen utdelade tabellsidor och formelsamling.

Maxpoäng: 30. För godkänd krävs minst 15 p, för Väl Godkänd 24 p. Eventuell bonuspoäng från 2015 adderas. Lösningar skall redovisas till varje uppgift. Uppgifterna är ej ordnade efter svårighetsgrad. Rättningsprotokoll anslås senast tre veckor efter tentamen.

1. Visa att om två händelser A och B är oberoende, då är även A och komplement B^c oberoende.

2p

Lösningsförslag:

$$P(A \cap B^c) \stackrel{\text{beting.}}{=} P(B^c|A)P(A) = (1 - P(B|A)) P(A) \stackrel{\text{obero.}}{=} (1 - P(B)) P(A) = P(A)P(B^c) \quad (2p)$$

2. Ett skolfotbollslag får pokalen om de vinner två matcher i rad. Laget tävlar med två andra skolor A och B. Turneringsregler säger att tre matcher spelas växelvis med A och B. Den första motståndaren får de välja själva. Antag att sannolikheterna för att vinna en match är $\frac{3}{5}$ mot A och $\frac{2}{5}$ mot B. Ska de välja A eller B som motståndare till första matchen? Ledning: Rita ett träd diagram.

2p

Lösningsförslag:

Första motståndare är A:

$$P(\text{Vinna 2 i rad} | A) = \underbrace{\frac{3}{5}}_{\text{seger mot A}} \cdot \underbrace{\frac{2}{5}}_{\text{seger mot B}} + \underbrace{\frac{2}{5}}_{\text{förlust mot A}} \cdot \underbrace{\frac{2}{5}}_{\text{seger mot B}} \cdot \underbrace{\frac{3}{5}}_{\text{seger mot A}} = \frac{42}{125}$$

Första motståndare är B:

$$P(\text{Vinna 2 i rad}|B) = \underbrace{\frac{2}{5}}_{\text{seger mot B}} \cdot \underbrace{\frac{3}{5}}_{\text{seger mot A}} + \underbrace{\frac{3}{5}}_{\text{förlust mot B}} \cdot \underbrace{\frac{3}{5}}_{\text{seger mot A}} \cdot \underbrace{\frac{2}{5}}_{\text{seger mot B}} = \frac{48}{125}$$

Det betyder att sannoligheten att vinna två matcher i rad och därmed får en chans att behålla pokalen är större om de börjar spela mot B. (2p)

3. Antag att skolfotbollslaget i uppgift 2. vann pokalen, men den stals från skolan och tjuven måste vara en av de 44 spelarna från skolan A och B. Varje spelare blir utfrågad från rektorn med hjälp av en lögn-detektor. Alla spelare säger att de inte är tjuven. Lögn-detektorn visar grönt om det är sant och rött om det var en lögn. Man får

- om personen talar sanning visar lögn-detektorn grönt i 99.9% av fallen och
- om personen ljuger visar detektorn rött i 90% av fallen.

Vad är sannolikheten att en spelare verkligen är tjuven om lögn-detektorn visa rött?

3p

Lösningförslag:

B_1 := 'spelare = tjuven', B_2 := 'spelare $\neq$ tjuven', A := 'röd'.

$$P(B_1) = \frac{1}{44}, P(B_2) = \frac{43}{44}, P(A|B_1) = 0.9, P(A|B_2) = 0.001 \quad (1p)$$

$$\begin{aligned} P(B_1|A) &\stackrel{\text{betning.}}{=} \frac{P(B_1 \cap A)}{P(A)} \stackrel{\text{Bayes}}{=} \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} \quad (1p) \\ &= \frac{\frac{1}{44} \cdot 0.9}{\frac{1}{44} \cdot 0.9 + \frac{43}{44} \cdot 0.001} \approx 0.954 \quad (1p) \end{aligned}$$

4. En maskin producerar 2% felaktiga pennor. 100 slumpmässigt utvalda pennor kontrolleras.

- a) Hur många felaktiga pennor kan man förvänta sig att få?

1p

- b) Vilken fördelning är rimlig att anta för den stokastiska variabeln X = 'antal felaktiga pennor'? Motivera ditt svar.

1p

- c) Vad är sannolikheten att minst 3 pennor är felaktiga?

1p

- d) Minst hur många pennor måste kontrolleras för att sannolikheten att få mindre än en felaktig penna är som mest 1%?

3p

Lösningförslag:

- a)

$$\mathbb{E}(X) = \lambda = n \cdot p = 100 \cdot 0.02 = 2 \quad (1p)$$

- b) X är en stokastisk variabel med två olika utfall felaktig eller inte. Alltså kan man anta en Binomial-fördelning, men eftersom $p = 0.02 < 0.1$ är så litet är en Poisson-fördelning det bättre valet. (1p)

c)

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) \\ &= 1 - (1 + 2 + 2^2/2)e^{-2} = 1 - 5e^{-2} = 0.323 \quad (1p) \end{aligned}$$

- d) Vi har $\lambda = n \cdot 0.02$ och söker n så att $P(X < 1) = P(X = 0) = e^{-n \cdot 0.02} \leq 0.01$. (1.5p):

$$e^{-n \cdot 0.02} \stackrel{!}{\leq} 0.01 \iff -n \cdot 0.02 \leq \log(0.01) \iff n \geq -\frac{\log(0.01)}{0.02} = 230.259 \quad (1p)$$

D.v.s. att man måste kontrollera minst 231 pennor. (0.5p)

5. Låt X och Y vara två stokastiska variabler. Visa följande formel för kovariansen med konstanter $a, b, c \in \mathbb{R}$:

$$\mathbb{C}(aX + c, bY + c) = ab\mathbb{C}(X, Y).$$

Ledning: Du kan använda att $\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

2p

Lösningsförslag:

$$\begin{aligned} \mathbb{C}(aX + c, bY + c) &\stackrel{led}{=} \mathbb{E}((aX + c)(bY + c)) - \mathbb{E}(aX + c)\mathbb{E}(bY + c) \quad (0.5p) \\ &\stackrel{reg}{=} \mathbb{E}(abXY + acX + bcY + c^2) - (a\mathbb{E}(X) + c)(b\mathbb{E}(Y) + c) \quad (0.5p) \\ &\stackrel{reg}{=} ab\mathbb{E}(XY) + ac\mathbb{E}(X) + bc\mathbb{E}(Y) + c^2 \\ &\quad - ab\mathbb{E}(X)\mathbb{E}(Y) - ac\mathbb{E}(X) - bc\mathbb{E}(Y) - c^2 \quad (0.5p) \\ &= ab(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \stackrel{led}{=} ab\mathbb{C}(X, Y) \quad (0.5p) \end{aligned}$$

6. En tvådimensionell slumpvariabel (X, Y) har täthetsfunktion

$$f_{XY}(x, y) = \begin{cases} k(x + y) & , 0 \leq x, 0 \leq y, x + y \leq 2 \\ 0 & , \text{annars} \end{cases}.$$

- a) Beräkna konstanten k , så att f_{XY} verkligen är en täthetsfunktion.

2p

- b) Beräkna de marginella täthetsfunktionerna $f_X(x)$ och $f_Y(y)$. Ledning: För X , delar upp i två fall, $0 \leq x \leq 2$ och $x < 0, 2 < x$. Gör samma sak för Y .

2p

- c) Är X och Y oberoende?

1p

Lösningsförslag:

a)

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x,y) dy dx \stackrel{!}{=} 1 \quad (0.5p)$$

Så vi får

$$\begin{aligned} \int_0^2 \int_0^{2-x} k(x+y) dy dx &= k \int_0^2 \left(2x - x^2 + \frac{1}{2}(2-x)^2 \right) dx \quad (0.5p) \\ &= k \int_0^2 \left(2x - x^2 + \frac{1}{2}(2-x)^2 \right) dx \\ &= k \int_0^2 \left(2 - \frac{1}{2}x^2 \right) dx \\ &= k \left[2x - \frac{1}{2} \frac{x^3}{3} \right]_0^2 \quad (0.5p) \\ &= k \frac{8}{3} = 1 \end{aligned}$$

och därmed

$$k = \frac{3}{8} \quad (0.5p)$$

b)

För X : $0 \leq x \leq 2$

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y) dy = \frac{3}{8} \int_0^{2-x} (x+y) dy \quad (0.5p) \\ &= \frac{3}{8} \left[xy + \frac{y^2}{2} \right]_0^{2-x} \\ &= \frac{3}{16} (4 - x^2) \quad (0.5p) \end{aligned}$$

och för $x < 0$, $2 < x$

$$f_X(x) = 0 \quad (0.5p)$$

Vi får detsamma för Y :

$$f_Y(y) = \begin{cases} \frac{3}{16}(4-y^2) & , \quad 0 \leq y \leq 2 \\ 0 & , \quad \text{annars} \end{cases} \quad (0.5p)$$

c) Nej, det ser man direkt från betingning $x + y \leq 2$ eller från $f_{XY}(x,y) \neq f_X(x)f_Y(y)$ (1p).

7. En resebyrå undrar om antal sålda resor x och omsättning y (i 10000 kr) är positivt linjärt korrelerade under de senaste sju åren:

i	År	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2009	200	470	94000	40000	220900
2	2010	250	630	157500	62500	396900
3	2011	300	640	192000	90000	409600
4	2012	280	630	176400	78400	396900
5	2013	350	790	276500	122500	624100
6	2014	400	930	372000	160000	864900
7	2015	425	950	403750	180625	902500
Summan		2205	5040	1672150	734025	3815800

- a) Beräkna och tolka korrelationskoefficienten

$$\rho = \frac{s_{xy}^2}{s_x \cdot s_y}.$$

Är det rimligt att anta ett (statistiskt signifikant) linjärt samband mellan x och y ?

2p

- b) Antag att det finns ett linjärt samband mellan x och y , så att $\hat{y}_i = 45 + 2.14x_i$. Vilken omsättning kan resebyrån förvänta sig i 2016 om de säljer 500 resor detta år?

1p

Lösningsförslag:

- a)

$$\begin{aligned}\bar{x} &= \frac{2205}{7} = 315, & \bar{y} &= \frac{5040}{7} = 720, & \bar{x}\bar{y} &= 226800, & \bar{x}^2 &= 99225, & \bar{y}^2 &= 518400 \\ s_{xy}^2 &= \frac{1}{n} \sum_{i=1}^7 x_i y_i - \bar{x}\bar{y} = \frac{1672150}{7} - 226800 = 12078.571 \\ s_x^2 &= \frac{1}{n} \sum_{i=1}^7 x_i^2 - \bar{x}^2 = \frac{734025}{7} - 99225 = 5635.714 \\ s_y^2 &= \frac{1}{n} \sum_{i=1}^7 y_i^2 - \bar{y}^2 = \frac{3815800}{7} - 518400 = 26714.286 \\ \rho &= \frac{s_{xy}^2}{s_x \cdot s_y} = \frac{12078.571}{\sqrt{5635.714} \sqrt{26714.286}} = 0.984 \quad (2p)\end{aligned}$$

Korrelationskoefficienten är över 98% och därmed finns en nästan perfekt positivt linjär korrelation mellan x och y . Det betyder att om antalet resor stiger, så stiger också omsättningen och det är rimligt att anta ett linjärt samband här. (1p)

- b)

$$\hat{y}_8 = 45 + 2.14 \cdot 500 = 1115 \text{ [10000 kr]} \quad (1p)$$

8. En undersökning av storlek $n_1 = 144$ skolor i staden A visade att 80% av skolorna har en modern datorsal. Ett jämförbart stickprov av storlek $n_2 = 100$ från staden B gav ett värde av 70%. Testa hypotesen att andelen skolor med en modern datorsal är detsamma i städerna A och B med hjälp av ett 95% konfidensintervall.

Lösningsförslag:

Antal skolor med en modern datorsal antas vara $\text{Bin}(n_i, p_i)$ -fördelade för $i = 1, 2$, där p_i är den sanna proportionen skolor i hela (oändliga) populationen i staden A ($i = 1$) respektive B ($i = 2$). Vi vet $\hat{p}_1 = 0.8$, $\hat{p}_2 = 0.7$. Eftersom $n_i \hat{p}_i (1 - \hat{p}_i) (= 23.04 \text{ och } 21) > 10$ för $i = 1, 2$ använder vi normalapproximationen (1.5p) och får ett approximativt $100(1 - \alpha)\%$ konfidensintervall med $\alpha = 0.05$ och $\lambda_{0.025} = 1.96$ för $p_1 - p_2$ i en oändlig population:

$$\begin{aligned} CI &= \left[\hat{p}_1 - \hat{p}_2 \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \right] \\ &= \left[0.8 - 0.7 \pm 1.96 \sqrt{\frac{0.8(1 - 0.8)}{144} + \frac{0.7(1 - 0.7)}{100}} \right] \\ &= \left[0.1 \pm 1.96 \sqrt{0.003211111} \right] \\ &= [-0.011, 0.211] \quad (1.5p) \end{aligned}$$

Konfidensintervallet innehåller alltså värdet 0, så det finns ingen grund för påståendet att det finns ett (statistiskt signifikant) skillnad mellan A och B. Hypotesen att andelen skolor med en modern datorsal är detsamma förkastas inte. ($H_0 : p_1 - p_2 = 0$ vs. $H_1 : p_1 - p_2 \neq 0$, H_0 förkastas inte.) (1p)

9. Antag att vi har ett stickprov av storlek n av en $N(\mu, \sigma)$ -fördelad stokastisk variabel X .

a) Hur är punktskattaren $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ för medelvärdet μ fördelad? Motivera ditt svar.

1p

b) Hur gör du om du vill konstruera ett konfidensintervall för μ om σ är okänd? Vilken fördelning används och varför? Beskriv med egna ord.

2p

Lösningsförslag:

a) Linjärkombinationer av normal-fördelade stokastiska variabler är i sig normal-fördelade med

$$\mathbb{E}(\bar{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) = \mu \quad \text{och} \quad V(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^n V(x_i) = \frac{\sigma^2}{n}$$

(1p)

b)

$$I_\mu = \left[\bar{x} \pm t(n-1)_{\alpha/2} \frac{s}{\sqrt{n}} \right] \quad (1p)$$

där variansen σ^2 är skattat med stickprovsvariansen s^2 . Eftersom σ^2 måste skattas, har vi ingen exakt $N(\mu, \sigma/n)$ -fördelning för $\bar{x}$ längre, utan en approximativt normal-fördelad Student t-fördelning som beror på hur många observationer n man har. Ju större n desto närmare kommer t-fördelningen den standard-normal-fördelningen. (1p)

Lycka till!