

FACIT: Tentamen L9MA30, LGMA30

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

2013-10-25 kl. 8.30-12.30

Examinator och Jour: Henrike Häbel, tel. 073 0671252 alt. 031 772 5380

Hjälpmedel: Valfri miniräknare (minnet måste vara tomt och apparaten kan inte ha möjligheten att koppla upp sig till nätet), (dubbelsidigt) A4 ark med egna anteckningar samt utdelade tabellsidor och formelsamling.

Maxpoäng: 30. För godkänd krävs minst 12 p, för Väl Godkänd 21 p. Eventuell bonuspoäng från hemdugga 2013 adderas. Lösningar skall redovisas till varje uppgift. Uppgifterna är ej ordnade efter svårighetsgrad. Rättningsprotokoll anslås senast tre veckor efter tentamen.

1. Vi kastar en tärning två gånger, d.v.s. utfallsrummet är $\Omega = \{(1,1), (1,2), (1,3), \dots, (6,5), (6,6)\}$.

a) Beräkna sannolikheten för händelsen $A = \text{'det första talet är större än 3'}$.

$$P(A) = \frac{3 \cdot 6}{36} = \frac{1}{2}$$

1p

b) Beräkna sannolikheten för händelsen $B = \text{'summan av båda talen är 7'}$.

$$P(B) = \frac{6}{36} = \frac{1}{6}$$

1p

c) Är händelserna A och B oberoende?

$$P(A)P(B) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} = \frac{3}{36} = P(A \cap B) \Rightarrow \text{oberoende}$$

1p

2. Uppskattningsvis har var femte vuxen i Sverige någon form av allergi och 40% av alla som har allergi är allergiska mot pollen. Anta att ett blodprov används för att avgöra om en person är allergisk mot pollen. Testet ger positivt svar (allergisk) med 95% sannolikhet om man verkligen är allergisk, och negativt svar (ej allergisk) med 99% sannolikhet om man är inte allergisk mot pollen.

- a) Vad är sannolikheten att en slumpmässigt vald person är allergisk mot pollen?

$$P(\text{pollen}) = P(\text{pollen} \cap \text{allergisk}) = P(\text{allergisk}) \cdot P(\text{pollen}|\text{allergisk}) = \frac{1}{5} \frac{4}{10} = \frac{2}{25} = 0.08$$

1p

- b) Anta att ett test ger positivt svar på en slumpmässigt vald person. Hur stor är sannolikheten att personen är allergisk mot pollen?

A_1 = 'allergisk mot pollen', A_2 := 'ej allergisk mot pollen', B := 'testet ger positivt svar'.

$$P(A_1) = \frac{2}{25} = 0.08, P(A_2) = \frac{23}{25} = 0.92, P(B|A_1) = 0.95, P(B|A_2) = 1 - 0.99 = 0.01$$

$$\begin{aligned} P(A_1|B) &\stackrel{\text{betingad}}{=} \frac{P(A_1 \cap B)}{P(B)} \stackrel{\text{Bayes}}{=} \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{\frac{2 \cdot 0.95}{25}}{\frac{2 \cdot 0.95 + 0.01 \cdot 23}{25}} \approx 0.89 \end{aligned}$$

3p

3. För en viss försäkring betalar man 700 kr per år och får 1000 kr vid ett skadetillfälle. Låt det årliga antalet skador per kund, X , vara Poisson-fördelat med parameter $\lambda > 0$. Beteckna vinsten som försäkringen gör från en viss kund per år $Y = 700 - 1000X$.

- a) Hur stor kan parametern λ av Poisson-fördelningen av de årliga skadetillfällena vara, så att det med 90% sannolikhet inte finns något skadetillfälle? Ledning: Du kan använda sannolikhetsfunktion $p_X(k) = \lambda^k \frac{e^{-\lambda}}{k!}$, $k = 0, 1, \dots$

$$P(X = 0) = 0.9 \Leftrightarrow p_X(0) = \lambda^0 \frac{e^{-\lambda}}{0!} = 0.9 \Leftrightarrow \lambda = -\log(0.9) \approx 0.105$$

2p

- b) Hur stor kan parametern λ vara som mest, om försäkringsbolaget inte ska gå med förlust i genomsnitt?

$$\mathbb{E}(700 - 1000X) \geq 0 \Leftrightarrow \mathbb{E}(X) = \lambda \leq \frac{-700}{-1000} = 0.7$$

1p

4. Antag att X är en kontinuerlig stokastisk variabel med täthetsfunktion

$$f_X(x) = \begin{cases} 6x(1-x) & , 0 \leq x \leq 1, \\ 0 & , \text{annars} \end{cases}$$

- a) Visa att $f_X(x)$ är en täthetsfunktion.

i) $6x(1-x) \in [0, 1/2] \geq 0 \quad \forall x$

ii)

$$\int_{-\infty}^{\infty} f_x(x) = \int_0^1 6x(1-x) = 6 \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = 6 \left(\frac{3-2}{6} \right) = 1$$

1p

b) Beräkna väntevärdet $\mathbb{E}(X)$.

$$\int_{-\infty}^{\infty} x f_x(x) = \int_0^1 6x^2(1-x) = 6 \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 = 6 \left(\frac{4-3}{12} \right) = \frac{1}{2}$$

1p

c) Beräkna sannolikhetsfunktionen $F_X(x)$ och sannolikheten $P(1/2 < x < 3/2)$.

$$F_X(x) \stackrel{a)}{=} \begin{cases} 0 & , x < 0 \\ 3x^2 - 2x^3 & , 0 \leq x \leq 1, \\ 1 & , x > 1 \end{cases}$$

och

$$P(1/2 < x < 3/2) = F_X(3/2) - F_X(1/2) \stackrel{c)}{=} 1 - (3(1/2)^2 - 2(1/2)^3) = 1 - \frac{6-2}{8} = \frac{1}{2}$$

1p

5. Betrakta den följande simultana sannolikhetsfunktionen för X med värden i $\{1, 2, 4\}$ och Y med värden i $\{-1, 0, 1\}$.

(X, Y)		k			
		1	0	-1	
j	1	1/8	1/12	1/6	a
	2	1/24	1/4	0	7/24
	4	1/6	b	1/12	1/3
		1/3	5/12	1/4	1

a) Bestäm konstanterna $a, b \in \mathbb{R}$.

$$a \stackrel{t.ex.}{=} 1 - 1/3 - 7/24 = 3/8, \quad b \stackrel{t.ex.}{=} 1/3 - 1/12 - 1/6 = 1/12$$

1p

b) Beräkna $C(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

$$\mathbb{E}(X) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{7}{24} + 4 \cdot \frac{1}{8} = \frac{55}{24}, \quad \mathbb{E}(Y) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{5}{12} - 1 \cdot \frac{1}{4} = \frac{1}{12}$$

och

$$\mathbb{E}(XY) = 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{24} + 4 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} - 2 \cdot 0 - 4 \cdot \frac{1}{12} = \frac{3}{8}$$

så att

$$C(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{3}{8} - \frac{55}{24} \cdot \frac{1}{12} = \frac{53}{288}$$

3p

c) Är X och Y oberoende?

Nej, eftersom

$$p_{XY}(2, -1) = 0 \neq \frac{7}{24} \frac{1}{4} = p_X(2)p_Y(-1)$$

1p

6. Låt X vara en $N(\mu, \sigma)$ -fördelad stokastisk variabel. Visa, genom att beräkna väntevärde och varians, att

$$Z = \frac{X - \mu}{\sigma}$$

är standardnormalfördelad, d.v.s. $Z \sim N(0, 1)$.

$$\mathbb{E}(Z) = \mathbb{E}\left(\frac{X - \mu}{\sigma}\right) = \frac{\mathbb{E}(X)}{\sigma} - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

och

$$V(Z) = V\left(\frac{X - \mu}{\sigma}\right) = \frac{V(X)}{\sigma^2} - 0 = \frac{\sigma^2}{\sigma^2} = 1$$

2p

7. Betrakta följande del av en artikel från 2013-09-10 publicerad på DN.se.

‘Var tionde förstaårsstudent på Harvard har någon gång fuskat vid ett provtillfälle. Knappt hälften av [de 1300 tillfrågade studenterna], 42 procent, uppger att de har använt sig av hjälpmedel eller otillåtna källor på hemduggan.’

Konstruera ett 95 % konfidensintervall för andelen studenter som fuskar för hemduggan. Ger detta intervall någon grund för att påstå att inte hälften av studenterna fuskar?

Antal studenter som fuskar antas $\text{Bin}(1300, p)$ -fördelade, där p är den sanna proportionen studenter som fuskar i hela (oändlig) populationen. Skatta p med $\hat{p} = 0.42$. Eftersom $n\hat{p}(1 - \hat{p}) \approx 316 \geq 10$ använder vi normalapproximationen och får approximativt 100(1 - α)% konfidensintervall med $\alpha = 0.05$ för p :

$$CI = \left[\hat{p} \pm z_{0.005} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

och eftersom $z_{0.025} = 1.96$

$$CI = [0.3932, 0.4468].$$

Det vill säga att konfidensintervallet inte innehåller värde = 0.5, så det ger grund för påståendet att inte (mindre än) hälften fuskar.

4p

8. Följande mätvärden för intelligenskvot (IK) kommer från två grupper av personer A och B.

A	86	101	93	98	89	101	100	104	95	99
B	101	97	96	99	103	113	106	97		

Härvid anses de uppmätta värdena vara två oberoende stickprov från en $N(\mu_A, \sigma)$ - respektive $N(\mu_B, \sigma)$ -fördelning med $\sigma = 6$. Beräkna ett 95% konfidensintervall för skillnaden. Ger detta intervall någon grund för att påstå att de IK-väntevärdena skiljer sig mellan grupperna?

$$n_A = 10, n_B = 8, \quad \bar{X}_A = 96.6, \bar{x}_B = 101.5, \quad \lambda_{0.025} = 1.96$$

100(1 - α)% konfidensintervall med $\alpha = 0.05$:

$$CI = \left[\bar{x}_A - \bar{X}_B \pm \lambda_{0.025} \sigma \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B} \right)} \right] = [-10.48, 0.68]$$

Det vill säga att konfidensintervallet innehåller värde 0, så det ger ingen grund för påståendet att de IK-väntevärdena skiljer sig mellan grupperna.

3p

9. Antag att man har ett stickprov $(x_1, \dots, x_n)$ av storlek n , d.v.s. x_i är observationer av oberoende och likafördelade stokastiska variabler X_i med väntevärde μ och varians σ^2 .

- a) Vad är anledningen till att man brukar använda

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{istället för} \quad s_*^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

när man skattar variansen?

s_*^2 är inte en väntevärdesriktig skattning, men s^2 är det.

1p

- b) Utgå från uttrycket i uppgift 9.a) och visa att s^2 kan skrivas som

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2) - n\bar{x}^2 \right).$$

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2) - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2) - 2n\bar{x}^2 + n\bar{x}^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2) - n\bar{x}^2 \right) \end{aligned}$$

2p

Lycka till!