

TENTAMEN I STATISTIKENS GRUNDER 1

2012-10-26

Skrivtid: kl 9.00 - 14.00
Godkända hjälpmedel: Miniräknare, språklexikon
Bifogade hjälpmedel: Formelblad, statistiska tabeller

- Tentamen består av 5 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per uppgift.
 - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
 - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
-

Uppgift 1.

Anta att A , B och C är tre händelser med sannolikheter $P(A) = 0,6$, $P(B) = 0,9$ och $P(C) = 0,3$. Dessutom gäller att $A \cup B = \Omega$ samt att A och C är disjunkta, dvs. ömsesidigt uteslutande.

- a) Beräkna sannolikheterna $P(A \cup B)$, $P(A \cap B)$, $P(A | B)$ och $P(B | A)$ samt ange om A och B är oberoende. (6p)
- b) Beräkna sannolikheterna $P(A | C)$, $P(C | A)$, $P(\bar{A} | C)$ och $P(A | \bar{C})$ samt ange om A och C är oberoende. (6p)
- c) Beräkna sannolikheten $P(\bar{A} \cup \bar{B})$ samt förklara varför $C \subset B$. (8p)

Uppgift 2.

Under 1970-talet gjordes i Storbritannien en analys av benägenheten att strejka i stora industriföretag. Det visade sig att för ett typiskt företag med ca 2000 anställda kunde antalet strejker per år med god approximation modelleras med en Poissonfördelning med väntevärde 0,4. Låt X = antalet strejker under ett år i ett sådant företag.

- a) Ange sannolikheten att ett företag råkar ut för minst en strejk. (2p)
- b) Anta att det i en viss region finns 10 företag av den angivna storleken. Anta att benägenheten att strejka är densamma i samtliga företag och att strejker sker oberoende av varandra. Låt Y = antal företag där strejker förekommer minst en gång. Hur är Y fördelad? Ange även väntevärde och varians för Y . (4p)
- c) Beräkna sannolikheten att (i) inga företag, (ii) samtliga företag samt (iii) givet att minst ett företag har strejkat, samtliga företag går ut i strejk. (6p)

- d) Beskriv kortfattat (max $\frac{1}{2}$ - 1 A4) de olika *skaltpyper* man förknippar med variabler och de värden som de kan anta. Ge ett realistiskt exempel på en variabel per skaltp. Vilken skaltp gäller i denna uppgift? (8p)

Uppgift 3.

Anta att två stokastiska variabler X och Y båda kan anta värdena 0 eller 1 och att den simultana frekvensfunktionen kan skrivas $f(x,y) = (1+x)(2-y)/9$.

- Beräkna sannolikheterna för hela det simultana utfallsrummet och sammanställ sannolikheterna i en tabell med marginalsannolikheterna för X resp Y . (6p)
- Ange de betingade fördelningarna samt beräkna väntevärde och varians för X givet att $Y=0$ respektive $Y=1$. (8p)
- Är X och Y oberoende eller beroende? Om de är beroende, kan det linjära sambandet sägas vara positivt eller negativt? (6p)

Uppgift 4.

Man har under säsongen observerat tiderna för två 100-meters löpare. Löpare X 's tider tenderar att vara normalfördelade med väntevärde 9,94 sekunder och standardavvikelse 0,05. Löpare Y är i snitt något snabbare med tider som också är normalfördelade med väntevärde 9,90 sekunder men en större variation med en standardavvikelse på 0,08. Anta att deras löptider är oberoende av varandra.

- Beräkna sannolikheten att båda löparna under ett träningslopp springer fortare än 9,92 sekunder. (6p)
- Beräkna sannolikheten att löpare X springer fortare än löpare Y . (8p)
- En *kontinuerlig* stokastisk variabel som vi kan beteckna med X , kan anta värden inom ett intervall. Förklara kortfattat varför $P(X \leq x) = P(X < x)$. (6p)

Uppgift 5.

En urna innehåller $N = 5$ kulor varav 2 är blå och 3 röda. Man drar $n = 3$ kulor från urnan. Definiera den stokastiska variabeln $Y =$ antal blå kulor.

- Ange utfallsrummet för Y när man drar utan återläggning, dvs. när en kula dragits läggs den inte tillbaka inför nästa dragning, samt ange sannolikhetsfunktionen för Y i tabellform. TIPS: Lista alla sekvenser av kulfärger som kan inträffa och beräkna sannolikheten för var och en. (8p)
- Låt $p =$ andelen blå kulor från start i urnan. Bekräfta att väntevärde och varians för Y kan skrivas som $E(Y) = np$ respektive $V(Y) = np(1-p)\left(\frac{N-n}{N-1}\right)$. (6p)
- Ange hur Y är fördelad om dragningen sker med återläggning samt ange väntevärde och varians. Förklara skillnaden mellan varianserna i de båda fallen. (6p)

FORMLER

Additionssatsen: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Multiplikationssatsen: $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Väntevärde för diskret s.v. X :

$$E(X) = \sum_{x \in \Omega_x} x f(x)$$

Varians för diskret s.v. X :

$$V(X) = \sum_{x \in \Omega_x} [x - E(X)]^2 f(x) = \sum_{x \in \Omega_x} x^2 f(x) - [E(X)]^2 = E(X^2) - E(X)^2$$

Kovarians och korrelation mellan två diskreta s.v. X och Y , ($\mu_x = E(X)$ och $\mu_y = E(Y)$):

$$\text{Cov}(X, Y) = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} (x - \mu_x)(y - \mu_y) f(x, y) = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} xy f(x, y) - \mu_x \mu_y$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

Räkneregler för väntevärden och varianser (a, b och c är konstanter och X och Y är stokastiska variabler)

$$E(c) = c$$

$$V(c) = 0$$

$$E(X + c) = E(X) + c$$

$$V(X + c) = V(X)$$

$$E(aX) = aE(X)$$

$$V(aX) = a^2 V(X)$$

$$E(aX + bY + c) = aE(X) + bE(Y) + c \quad V(aX + bY + c) = a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X, Y)$$

Binomialfördelningen:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x! (n-x)!} p^x (1-p)^{n-x}$$

Poissonfördelningen:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

LÖSNINGSFÖRSLAG (kortfattade)

Uppgift 1.

- a) Givet i fetstil samt $PA \cup B = \Omega$ vilket betyder att $P(A \cup B) = P(\Omega) = 1$

$$P(A \cap C) = [\text{additionssatsen}] = P(A) + P(B) - P(A \cup B) = 0,6 + 0,9 - 1 = 0,5$$

$$P(A | B) = P(A \cap B) / P(B) = 0,5/0,9 = 5/9 \approx 0,556$$

$$P(B | A) = P(A \cap B) / P(A) = 0,5/0,6 = 5/6 \approx 0,833$$

A och B är beroende ty t.ex. $P(A | B) \neq P(A)$

	B	$\bar{B}$	
A	0,5	0,1	0,6
$\bar{A}$	0,4	0	0,4
	0,9	0,1	1,0

- b) Givet i fetstil samt $A \cap C = \emptyset$ vilket betyder att $P(A \cap C) = P(\emptyset) = 0$

$$P(A | C) = P(A \cap C) / P(C) = [\text{ty } A \text{ och } C \text{ är disjunkta}] = 0/0,3 = 0$$

$$P(C | A) = P(A \cap C) / P(A) = 0/0,3 = 0$$

$$P(\bar{A} | C) = 1 - P(A | C) = 1 - 0 = 1$$

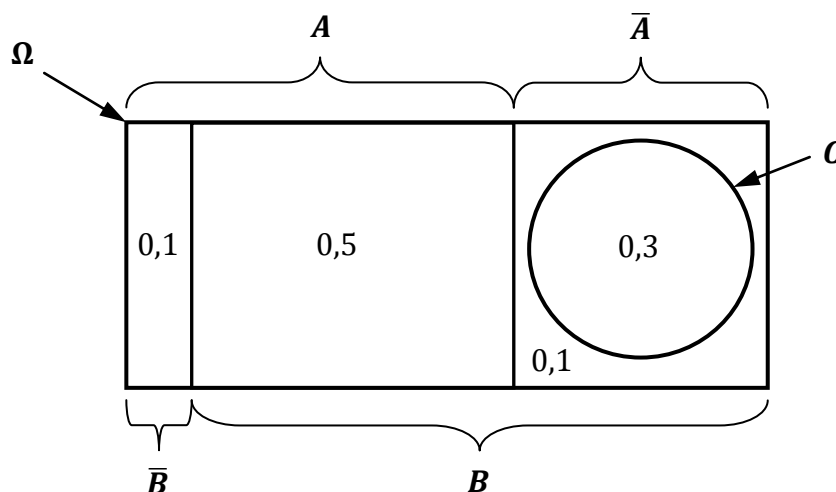
$$P(A | \bar{C}) = P(A \cap \bar{C}) / P(\bar{C}) = P(A) / [1 - P(C)] = 0,6/0,7 = 6/7 \approx 0,857$$

A och C är beroende ty t.ex. $P(A | C) \neq P(A)$

	C	$\bar{C}$	
A	0	0,6	0,6
$\bar{A}$	0,3	0,1	0,4
	0,3	0,7	1,0

- c) $P(\bar{A} \cup \bar{B}) = [\text{additionssatsen}] = P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cap \bar{B}) = P(\bar{A}) + P(\bar{B}) - 0 = 1 - P(A) + 1 - P(B) = 2 - 0,6 - 0,9 = 0,5$

Eftersom A och C är disjunkta måste C ligga i $\bar{A}$. Och eftersom $A \cup B = \Omega$ måste C ligga i den del av Ω som endast täcks in av B . Alltså är C en delmängd till B .



Uppgift 2.

- a) $X \sim Po(0,4)$. $P(X > 1) = 1 - P(X = 0) = 1 - \frac{0,4^0 e^{-0,4}}{0!} = [\text{alt. enl. tabell}] = 1 - 0,67032 = 0,32968 \approx 0,33$

- b) $Y \sim Bin(n, \pi)$ där $n = 10$ och $\pi = 0,33$ från a-uppgiften.
 $E(Y) = n\pi = 3,3$ och $V(Y) = n\pi(1 - \pi) = 2,21$.

- c) (i) $P(Y = 0) = \binom{10}{0} 0,33^0 0,67^{10} = 1 \cdot 1 \cdot 0,67^{10} \approx 0,0182$
(ii) $P(Y = 10) = \binom{10}{10} 0,33^{10} 0,67^0 = 1 \cdot 0,33^{10} \cdot 1 \approx 0,0000153$
(iii) $P(Y = 10 | Y > 0) = P(Y = 10 \cap Y > 0) / P(Y > 0) = P(Y = 10) / [1 - P(Y = 0)] =$
 $= 0,33^{10} / (1 - 0,0182) \approx 0,0000156$
- d) Se kurslitteratur och föreläsningssanteckningar. Kortfattat: nominal-, ordinal-, intervall- och kvotskala. I detta fall gäller kvotskala för både X och Y; det är meningsfullt att tala om t.ex. "dubbelt så många strejker".

Uppgift 3.

- a) Simultana sannolikheter och marginalsannolikheter:

$f(x,y)$	$Y = 0$	1	$f_Y(y)$
$X = 0$	2/9	1/9	1/3
1	4/9	2/9	2/3
$f_X(x)$	2/3	1/3	1

- b) $P(X = 0 | Y = 0) = f(0,0) / f_Y(0) = (2/9)/(2/3) = 1/3$
 $P(X = 1 | Y = 0) = f(1,0) / f_Y(0) = (4/9)/(2/3) = 2/3$
 $E(X | Y = 0) = 2/3; V(X | Y = 0) = 2/9$
 $P(X = 0 | Y = 1) = f(0,1) / f_Y(1) = (1/9)/(1/3) = 1/3$
 $P(X = 1 | Y = 1) = f(1,1) / f_Y(1) = (2/9)/(1/3) = 2/3$
 $E(X | Y = 1) = 2/3; V(X | Y = 1) = 2/9$
- c) De är oberoende eftersom de betingade sannolikheterna för X givet Y i b-uppgiften är identiska sinsemellan samt med marginalsannolikheterna för X i a-uppgiften. Det linjära sambandet är alltså noll (varken positivt eller negativt).

Uppgift 4.

- a) $P(X < 9,92 \cap Y < 9,92) = [\text{ty } X \text{ och } Y \text{ oberoende}] = P(X < 9,92) \cdot P(Y < 9,92) =$
 $= P\left(Z < \frac{9,92-9,94}{0,05}\right) \cdot P\left(Z < \frac{9,92-9,90}{0,08}\right) = P(Z < -0,4) \cdot P(Z < 0,25) =$
 $= [1 - \Phi(0,4)] \cdot \Phi(0,25) = [\text{enl. tabell}] = (1 - 0,65542) \cdot 0,59871 = 0,20630 \approx 0,21$
- b) $P(X < Y) = P(X - Y < 0) = P(W < 0) = [\text{där } W \sim N(0,04;0,089)] = P\left(Z < \frac{0-0,04}{0,09434}\right) =$
 $= \Phi(-0,423999) \approx 1 - \Phi(0,42) = [\text{enl. tabell}] = 1 - 0,66276 \approx 0,337$
- c) För en kontinuerlig variabel ligger utfallen »oändligt tätt» och inget enskilt utfall kan antas med positiv sannolikhet. Det är istället *arean* under täthetsfunktionen (frekvensfunktionen) ner till x-axeln i ett (del-)intervall till utfallsrummet som anger sannolikheten. Att inkludera en ändpunkt eller inte i ett intervall spelar alltså ingen roll eftersom den enskilda linjen från punkten upp till funktionen $f(x)$ punkten inte har någon »areamässig» utsträckning.

Uppgift 5.

- a) Lista samtliga sekvenser och beräkna sannolikheterna.

Sekvens nr i	Färg dragning nr			Sannolikhet vid			$P(\text{Sekvens})$	y	$P(Y=y)$
	1:a	2:a	3:e	1:a	2:a	3:e			
1	Röd	Röd	Röd	3/5	2/4	1/3	6/60	0	0,1
2	Röd	Röd	Blå	3/5	2/4	2/3	12/60	1	0,6
3	Röd	Blå	Röd	3/5	2/4	2/3	12/60	1	
4	Blå	Röd	Röd	2/5	3/4	2/3	12/60	1	
5	Röd	Blå	Blå	3/5	2/4	1/3	6/60	2	0,3
6	Blå	Röd	Blå	2/5	3/4	1/3	6/60	2	
7	Blå	Blå	Röd	2/5	1/4	3/3	6/60	2	
Summa							1		1

Observera att sannolikheterna för röd och blå kula förändras vid varje dragning eftersom man ändrar på sammansättningen av röda och blå kulor.

- b) $N = 5; n = 3; p = 2/5 = 0,4$

$$E(Y) = 0 \cdot 0,1 + 1 \cdot 0,6 + 2 \cdot 0,3 = 1,2;$$

$$np = 3 \cdot 0,4 = 1,2$$

$$V(Y) = (0-1,2)^2 \cdot 0,1 + (1-1,2)^2 \cdot 0,6 + (2-1,2)^2 \cdot 0,3 = 0,36$$

$$np(1-p) \left(\frac{N-n}{N-1} \right) = 3 \cdot 0,4 \cdot 0,6 \cdot \left(\frac{5-3}{5-1} \right) = 0,36$$

- c) Om dragning sker med återläggning är Y binomialfördelad med väntevärde $E(Y) = np = 1,2$ och varians $V(Y) = np(1-p) = 0,72$.

Väntevärdet är identiskt men variansen är större när man drar med återläggning. Det första man kan observera är att med återläggning blir utfallsrummet $\{0,1,2,3\}$ men utan återläggning kan man aldrig observera 3 blå kulor. Detta drar i sig ner variansen men framförallt är det sannolikheterna som förändras efter varje enskild dragning när man drar med eller utan återläggning som påverkar variansen.