STOCKHOLMS UNIVERSITET HT2012

Statistiska institutionen

Michael Carlson

**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* ⋃ *B* = Ω samt att *A* och *C* är disjunkta, dvs. ömsesidigt uteslutande.

1. Beräkna sannolikheterna *P*(*A* ⋃ *B*), *P*(*A* ⋂ *B*), *P*(*A* | *B*) och *P*(*B* | *A*) samt ange om *A* och *B* är oberoende. (6p)
2. Beräkna sannolikheterna *P*(*A* | *C*), *P*(*C* | *A*), *P*($\overbar{A}$ | *C*) och *P*(*A* | $\overbar{C}$) samt ange om *A* och *C* är oberoende. (6p)
3. Beräkna sannolikheten $P\left(\overbar{A}∪\overbar{B}\right)$ samt förklara varför C ⊂ 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.

1. Ange sannolikheten att ett företag råkar ut för minst en strejk. (2p)
2. 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)
3. 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)
4. Beskriv kortfattat (max ½ - 1 A4) de olika *skaltyper* man förknippar med variab­ler och de värden som de kan anta. Ge ett realistiskt exempel på en variabel per skaltyp. Vilken skaltyp 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.

1. Beräkna sannolikheterna för hela det simultana utfallsrummet och sammanställ sannolikheterna i en tabell med marginalsannolikheterna för *X* resp *Y*. (6p)
2. Ange de betingade fördelningarna samt beräkna väntevärde och varians för *X* givet att *Y* =0 respektive *Y* = 1. (8p)
3. 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.

1. Beräkna sannolikheten att båda löparna under ett träningslopp springer fortare än 9,92 sekunder. (6p)
2. Beräkna sannolikheten att löpare X springer fortare än löpare Y. (8p)
3. En *kontinuerlig* stokastisk variabel som vi kan beteckna med *X*, kan anta värden inom ett intervall. Förklara kortfattat varför *P*(*X* ≤ *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.

1. 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)
2. 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)
3. 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: $P\left(A⋃B\right)=P\left(A\right)+P\left(B\right)-P\left(A⋂B\right)$

Multiplikationssatsen: $P\left(A⋂B\right)=P\left(B\right)∙P\left(B\right)=P(B|A)∙P(A)$

Väntevärde för diskret s.v. *X*:

$$E\left(X\right)=\sum\_{x\in Ω\_{x}}^{}xf(x)$$

Varians för diskret s.v. *X*:

$$V\left(X\right)=\sum\_{x\in Ω\_{x}}^{}[x-E\left(X\right)]^{2}f(x)=\sum\_{x\in Ω\_{x}}^{}x^{2}f\left(x\right)-[E\left(X\right)]^{2}=E\left(X^{2}\right)-E(X)^{2}$$

Kovarians och korrelation mellan två diskreta s.v. *X* och *Y*, ($μ\_{x}=E(X)$ och $μ\_{y}=E(Y)$):

$$Cov\left(X,Y\right)=\sum\_{x\in Ω\_{x}}^{}\sum\_{y\in Ω\_{y}}^{}\left(x-μ\_{x}\right)\left(y-μ\_{y}\right)f\left(x,y\right)=\sum\_{x\in Ω\_{x}}^{}\sum\_{y\in Ω\_{y}}^{}xyf\left(x,y\right)-μ\_{x}μ\_{y}$$

$$Corr\left(X,Y\right)=\frac{Cov(X,Y)}{SD(X)∙SD(Y)}=\frac{Cov(X,Y)}{\sqrt{V\left(X\right)V(Y)}}$$

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

|  |  |
| --- | --- |
| $$E\left(c\right)=c$$ | $$V\left(c\right)=0$$ |
| $$E\left(X+c\right)=E\left(X\right)+c$$ | $$V\left(X+c\right)=V(X)$$ |
| $$E\left(aX\right)=aE(X)$$ | $$V\left(aX\right)=a^{2}V(X)$$ |
| $$E\left(aX+bY+c\right)=aE\left(X\right)+bE\left(Y\right)+c$$ | $$V\left(aX+bY+c\right)=a^{2}V\left(X\right)+b^{2}V\left(Y\right)+2abCov\left(X,Y\right)$$ |

Binomialfördelningen:

$$f\left(x\right)=\left(\begin{matrix}n\\x\end{matrix}\right)p^{x}(1-p)^{n-x}=\frac{n!}{x!\left(n-x\right)!}p^{x}(1-p)^{n-x}$$

Poissonfördelningen:

$$f\left(x\right)=\frac{λ^{x}e^{-λ}}{x!}$$

**LÖSNINGSFÖRSLAG (kortfattade)**

**Uppgift 1.**

1. Givet i fetstil samt *PA* ⋃ *B* = Ω vilket betyder att *P*(*A* ⋃ *B*) = *P*(Ω) = 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | *B* | $$\overbar{B}$$ |  |
| *A* | 0,5 | 0,1 | **0,6** |
| $$\overbar{A}$$ | 0,4 | 0 | 0,4 |
|  | **0,9** | 0,1 | 1,0 |

*P*(*A* ∩ *C*) = [additionssatsen] = *P*(*A*) + *P*(*B*) - *P*(*A* ⋃ *B*) = 0,6 + 0,9 -1 = 0,5

*P*(*A* | *B*) = *P*(*A* ∩ *B*) / *P*(*B*) = 0,5/0,9 = 5/9 ≈ 0,556

*P*(*B* | *A*) = *P*(*A* ∩ *B*) / *P*(*A*) = 0,5/0,6 = 5/6 ≈ 0,833

*A* och *B* är beroende ty t.ex. *P*(*A* | *B*) ≠ *P*(*A*)

1. Givet i fetstil samt *A* ∩ *C* = ø vilket betyder att *P*(*A* ∩ *C*) = *P*(ø) = 0

|  |  |  |  |
| --- | --- | --- | --- |
|  | *C* | $$\overbar{C}$$ |  |
| *A* | 0 | 0,6 | **0,6** |
| $$\overbar{A}$$ | 0,3 | 0,1 | 0,4 |
|  | **0,3** | 0,7 | 1,0 |

*P*(*A* | *C*) = *P*(*A* ∩ *C*) / *P*(*C*) = [ty *A* och *C* är disjunkta] = 0/0,3 = 0

*P*(*C* | *A*) = *P*(*A* ∩ *C*) / *P*(*A*) = 0/0,3 = 0

*P*($\overbar{A}$ | *C*) = 1 – *P*(*A* | *C*) = 1 – 0 = 1

*P*(*A* | $\overbar{C}$) = *P*(*A* ∩ $\overbar{C}$) / *P*($\overbar{C}$) = *P*(*A*) / [1 - *P*(*C*)] = 0,6/0,7 = 6/7 ≈ 0,857

*A* och *C* är beroende ty t.ex. *P*(*A* | *C*) ≠ *P*(*A*)

1. $P\left(\overbar{A}∪\overbar{B}\right)$ = [additionssatsen] = $P\left(\overbar{A}\right)+P\left(\overbar{B}\right)-P\left(\overbar{A}∩\overbar{B}\right)$ = $P\left(\overbar{A}\right)+P\left(\overbar{B}\right)-$ 0 = = 1 - *P*(*A*) + 1 – *P*(*B*) = 2 – 0,6 – 0,9 = 0,5

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

$$ \overbar{A}$$

$$ A$$

$$ C$$

$$ Ω$$

0,5

0,1

0,3

0,1

$$ \overbar{B}$$

$$ B$$

**Uppgift 2.**

1. *X* ~ *Po*(0,4). *P*(*X* > 1) = 1 – P(X =0) = 1 – $\frac{0,4^{0}e^{-0,4}}{0!}$ = [alt. enl. tabell] = 1 – 0,67032 = = 0,32968 ≈ 0,33
2. *Y* ~ *Bin*(*n*,π) där *n* = 10 och π = 0,33 från a-uppgiften.

*E*(*Y*) = *n*π = 3,3 och *V*(*Y*) = *n*π(1 – π) = 2,21.

1. (i) P(Y = 0) = $\left(\genfrac{}{}{0pt}{}{10}{0}\right)0,33^{0}0,67^{10}$ = 1·1·$0,67^{10}$ ≈ 0,0182

(ii) P(Y = 10) = $\left(\genfrac{}{}{0pt}{}{10}{10}\right)0,33^{10}0,67^{0}$ = 1·$0,33^{10}$·1 ≈ 0,0000153

(iii) P(Y = 10 | Y > 0) = P(Y = 10 ∩ Y > 0) / P(Y > 0) = P(Y = 10) / [1 – P(Y = 0) =

 = $0,33^{10}$/ (1 – 0,0182) ≈ 0,0000156

1. Se kurslitteratur och föreläsningsanteckningar. 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.**

1. Simultana sannolikheter och marginalsannolikheter:

|  |  |  |  |
| --- | --- | --- | --- |
| *f*(*x*,*y*) | *Y* = 0 | 1 | *fY*(y) |
| *X* = 0 | 2/9 | 1/9 | 1/3 |
| 1 | 4/9 | 2/9 | 2/3 |
| *fX*(*x*) | 2/3 | 1/3 | 1 |

1. *P*(*X* = 0 | *Y* = 0) = *f*(0,0) / *fY*(0) = (2/9)/(2/3) = 1/3

*P*(*X* = 1 | *Y* = 0) = *f*(1,0) / *fY*(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) / *fY*(1) = (1/9)/(1/3) = 1/3

*P*(*X* = 1 | *Y* = 1) = *f*(1,1) / *fY*(1) = (2/9)/(1/3) = 2/3

*E*(*X* | *Y* = 0) = 2/3; *V*(*X* | *Y* = 0) = 2/9

1. 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.**

1. *P* (*X* < 9,92 ∩ *Y* < 9,92) = [ty *X* och *Y* oberoende] = *P*(*X* < 9,92)·*P*(*Y* < 9,92) = $=P\left(Z<\frac{9,92-9,94}{0,05}\right)∙P\left(Z<\frac{9,92-9,90}{0,08}\right)$ = *P*(*Z* < -0,4)·*P*(*Z* < 0,25) =

 = [1 – Φ(0,4)]·Φ(0,25) = [enl. tabell] = (1 – 0,65542) · 0,59871 = 0,20630 ≈ 0,21

1. *P* (*X* < *Y*) = *P* (*X* – *Y* < 0) = *P* (*W* < 0) = [där *W* ~ *N*(0,04;0,089)] =$ P\left(Z<\frac{0-0,04}{0,09434}\right)$ = = Φ(-0,423999) ≈ 1 - Φ(0,42) = [enl. tabell] = 1 – 0,66276 ≈ 0,337
2. 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.**

1. Lista samtliga sekvenser och beräkna sannolikheterna.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sekvens | Färg dragning nr | Sannolikhet vid | *P*(Sekvens) | *y* | *P*(*Y* = *y*) |
| nr *i* | 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 |  |
| 3 | Röd | Blå | Röd | 3/5 | 2/4 | 2/3 | 12/60 | 1 | 0,6 |
| 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 |  |
| 6 | Blå | Röd | Blå | 2/5 | 3/4 | 1/3 | 6/60 | 2 | 0,3 |
| 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.

1. *N* = 5; *n* = 3; *p* = 2/5 = 0,4

*E*(*Y*) = 0·0,1 + 1·0,6 + 2·0,3 = 1,2;

*np* = 3·0,4 = 1,2

*V*(*Y*) = (0-1,2)2·0,1 + (1-1,2)2·0,6 + (2-1,2)2·0,3 = 0,36

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

1. 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 identisk 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 utan återläggning som påverkar variansen.