Aalto-universitetet

Gripenberg

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik
Mellanförhör 1, 27.1.2015

Skriv ditt namn, nummer och övriga uppgifter på varje papper!
En kalkylator, I. Mellins statistiska tabeller, Matlab/Octave-funktionslistan och en anteckningslapp är tillåtna hjälpmedel!
Skriv ut alla steg i dina räkningar!
Använd antingen tabeller (eller räknedosa) eller skriv med vilka Matlab/Octave-kommandon du kan räkna de värden av någon fördelningsfunktion eller dess inverser som du behöver och vilka slutsatser, tex. beträffande signifikansnivåer du kunde göra med stöd av dessa värden.

1.

(a) En kommitté består av 6 personer. Sannolikheten för att en medlem försenar sig till ett möte är 20% (oberoende av hur många andra som också försenar sig). Beräkna sannolikheten för att minst 4 medlemmar kommer i tid till mötet.

(b) En förening har 60 medlemmar av vilka tre har samma efternamn (och alla andra olika efternamn). För att ordna föreningens 50-årsjubileum väljs slumpmässigt en arbetsgrupp med 10 medlemmar bland föreningens alla medlemmar. Vad är sannolikheten att två av dem som har samma efternamn väljs till arbetsgruppen?

Dina svar får innehålla $+, -, \cdot, /, !$ och potenser.

Lösning: (a) Vi räknar sannolikheten för att högst 4, 5 eller 6 medlemmar kommer i tid och den är

$$\binom{6}{4} \cdot 0.8^4 \cdot 0.2^2 + \binom{6}{5} \cdot 0.8^5 \cdot 0.2 + 0.8^6 = 15 \cdot 0.8^4 \cdot 0.2^2 + 6 \cdot 0.8^5 \cdot 0.2 + 0.8^6$$
$$= \frac{2816}{3125} \approx= 0.90112.$$

(b) Arbetsgruppen kan väljas på $\binom{60}{10}$ olika sätt. Vi kan välja två av de tre som har samma efternamn på $\binom{3}{2}$ olika sätt och de återstående 8 som har något annat efternamn på $\binom{57}{8}$ olika sätt. Sannolikheten blir därför

$$\frac{\binom{3}{2}\binom{57}{8}}{\binom{60}{10}} = \frac{3 \cdot 57! \cdot 10! \cdot 50!}{60! \cdot 8! \cdot 49!} = \frac{3 \cdot 10 \cdot 9 \cdot 50}{60 \cdot 59 \cdot 58} = \frac{225}{3422} \approx 0.066.$$

---

2. I en skog bor tre olika slags heffaklumpar, stora, medelstora och små så att 30% är stora, 30% är medelstora och 40% små. Sannolikheten för att en stor heffaklump klättrat upp i ett träd är 5%, medan motsvarande sannolikheter för medelstora och små heffaklumpar är 10% och 15%. Om du nu ser en heffaklump som klättrat upp i ett träd, vad är sannolikheten för att det är en stor heffaklump du ser?

Lösning: Låt A, AA och AAA vara händelserna att en heffaklump är liten, medelstor och stor så att $\Pr(A) = 0.4$, $\Pr(AA) = 0.3$ och $\Pr(AAA) = 0.3$.

Låt T vara händelsen att en heffaklump sitter i ett träd. Enligt antagandena vet vi att $\Pr(T|AAA) = 0.05$, $\Pr(T|AA) = 0.1$ och $\Pr(T|A) = 0.15$).

Det vi skall räkna ut är $\Pr(AAA|T)$ och med hjälp av Bayes formel får vi

$$\begin{aligned}\Pr(AAA|T) &= \frac{\Pr(T|AAA)\cdot\Pr(AAA)}{\Pr(T|A)\cdot\Pr(A)+\Pr(T|AA)\cdot\Pr(AA)+\Pr(T|AAA)\cdot\Pr(AAA)}\\ &= \frac{0.05\cdot 0.3}{0.15\cdot 0.4+0.1\cdot 0.3+0.05\cdot 0.3} = \frac{1}{7}\approx 0.14.\end{aligned}$$

---

3. I en stad inträffar det elavbrott med exponentialfördelade (och oberoende) intervall så att det under ett år förekommer i genomsnitt 240 elavbrott.

(a) (4p) Vad är sannolikheten att det första elavbrottet ett år inträffar senast den 4 januari?

(b) (2p) Vad är sannolikheten att det nittionde elavbrottet ett år inträffar senast i april? Använd normalapproximation!

Räkna med att alla månader har 30 dagar.

Lösning: Det är en följd av antagandet att intervallen mellan elavbrotten är exponentialfördelade att antalet elavbrott under ett år är Poisson fördelade med parametern $\lambda\cdot 360$ om tiden mäts i dagar och λ är parametern i exponentialfördelningen. Nu är väntevärdet av en Poissonfördelad slumpvariabel parametern så att $240 = \lambda\cdot 360$ och $\lambda = \frac{240}{360} = \frac{2}{3}$. Sannolikheten att det först elavbrottet inträffar senast den 4 januari är $\Pr(X\le 4)$ då $X\sim \text{Exp}(\frac{2}{3})$ vilket är

$$\Pr(X\le 4) = 1-\mathrm{e}^{-\frac{2}{3}\cdot 4}\approx 0.93.$$

(b) Här kan vi räkna på två olika sätt som ger något olika resultat: Om X_j är tiden mellan två elavbrott (det första räknat från början av året) så inträffar elavbrott nummer 90 vi tidpunkten $Y = \sum_{j=1}^{90} X_j$. Eftersom väntevärdet av en $\text{Exp}(\lambda)$-fördelad slumpvariabel är $\frac{1}{\lambda}$ och och variansen är $\frac{1}{\lambda^2}$ och eftersom $X_j\sim\text{Exp}(\frac{2}{3})$ så är $\mathrm{E}(Y) = 90\cdot\frac{3}{2} = 135$ och $\text{Var}(Y) = 90\cdot(\frac{3}{2})^2 = 202.5$. Med hjälp av den centrala gränsvärdessatsen får vi nu, eftersom det finns 120 dagar före slutet av april,

$$\begin{aligned}\Pr(Y\le 120) &= \Pr\left(\frac{Y-\mathrm{E}(Y)}{\sqrt{\text{Var}(Y)}}\le\frac{120-\mathrm{E}(Y)}{\sqrt{\text{Var}(Y)}}\right) = \Pr\left(\frac{Y-\mathrm{E}(Y)}{\sqrt{\text{Var}(Y)}}\le\frac{120-135}{\sqrt{202.5}}\right)\\ &= \Pr\left(\frac{Y-\mathrm{E}(Y)}{\sqrt{\text{Var}(Y)}}\le -1.0545\right)\approx 0.146.\end{aligned}$$

Ett annat sätt är att räkna ut sannolikheten för att det inträffar minst 90 elavabrott före slutet av april. Eftersom antalet elabrott U under 120 dagar är Poisson$(\frac{2}{3}\cdot 120)$-fördelad så skall vi

räkna ut $\Pr(U \geq 90)$ och eftersom $\mathrm{E}(U) = \mathrm{Var}(U) = \frac{2}{3} \cdot 120 = 80$ så får vi

$$\Pr(U \geq 90) = \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \geq \frac{90 - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}}\right) = \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \geq \frac{90 - 80}{\sqrt{80}}\right)$$
$$= \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \geq 1.118\right) \approx 0.132.$$

Om man inte använder normalapproximationer blir svaret 0.14448 och orsaken till att den första approximatione ger ett bättre resultat är att man där approximerar en kontinuerlig slumpvariabel. Om man i det senare fallet använder en kontinuitetskorrigering får man som svar 0.144.

---

4. En frekvensfunktion f_{XY} definieras med följande tabell:

$f_{XY}(x,y)$		Y		
		2	3	4
X	1	$\frac{1}{10}$	$\frac{1}{10}$	0
	2	0	$\frac{2}{10}$	$\frac{1}{10}$
	3	$\frac{1}{10}$	0	0
	4	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

(a) Hur ser man att detta faktiskt är en frekvensfunktion?
(b) Bestäm den betingade fördelningen av X under villkoret $Y = 3$.
(c) Bestäm det betingade väntevärdet $\mathrm{E}(Y|X = 1)$.

Lösning: (a) f_{XY} är en frekvensfunktion eftersom $f_{XY}(x, y) \geq 0$ för alla x och y och $\sum_{x=1}^{4} \sum_{y=2}^{4} f_{XY}(x, y) = 1$.

(b) Enligt tabellen är

$$f_Y(3) = \Pr(Y = 3) = \sum_{x=1}^{4} f_{XY}(x, 3) = \frac{1}{10} + \frac{2}{10} + 0 + \frac{2}{10} = \frac{1}{2}.$$

Detta innebär att den betingade fördelningen av X under villkoret $Y = 3$ har frekvensfunktionen $\dfrac{f_{XY}(x, 3)}{f_Y(3)}$ och den är

x	1	2	3	4
$f_{X\|Y}(x\|3)$	$\frac{1}{5}$	$\frac{2}{5}$	0	$\frac{2}{5}$

(c) Eftersom $\Pr(X = 1) = \frac{1}{10} + \frac{1}{10}$ så är den betingade fördelningen av Y under villkoret $X = 1$ följande

y	2	3	4
$f_{Y\|X}(y\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Det betingade väntevärdet blir därför

$$2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} + 4 \cdot 0 = \frac{5}{2}.$$