Aalto-universitetet

Gripenberg

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik
Mellanförhörsomtagning och tentamen 12.3.2014

Skriv ditt namn, nummer och övriga uppgifter på varje papper!
Du får använda en räknare och Ilkka Mellins tabeller och ”formelsamling”.

Mellanförhör 1: Uppgifterna 1,2,3,4
Mellanförhör 2: Uppgifterna 5,6,7,8
Tentamen: Välj 5 av uppgifterna 1,3,4,5,7,8

1. Antag att A och B är händelser så att $0 < \Pr(A) < 1$ och $0 < \Pr(B) < 1$. Följer det av antagandet att $\Pr(A|B) = \Pr(A|B^c)$ att A och B är oberoende? Motivera ditt svar!

Lösning: Enligt definitionen är

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{och} \quad \Pr(A|B^c) = \frac{\Pr(A \cap B^c)}{\Pr(B^c)}.$$

Nu är $\Pr(B^c) = 1 - \Pr(B)$ så att antagandet betyder att

$$\Pr(A \cap B)(1 - \Pr(B)) = \Pr(A \cap B^c)\Pr(B) \quad \text{så att} \quad \Pr(A \cap B) = (\Pr(A \cap B) + \Pr(A \cap B^c))\Pr(B).$$

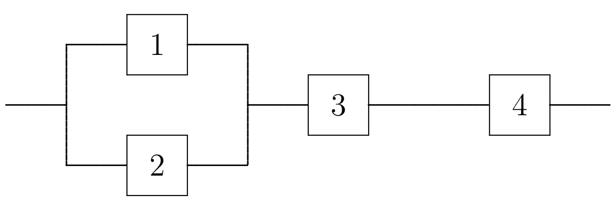
Dessutom gäller $A = (A \cap B) \cup (A \cap B^c)$ så att $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c)$ vilket betyder att

$$\Pr(A \cap B) = \Pr(A)\Pr(B),$$

dvs. A och B är oberoende.

---

2. Bilden visar ett system som består av 4 komponenter som alla fungerar oberoende av varandra med sannolikheten p. Vad är sannolikheten att systemet fungerar, dvs. att det finns en ”fungerande väg” genom systemet då $p = 0.9$?



Lösning: Systemet fungerar om antingen komponent 1 eller komponent 2 fungerar och komponenterna 3 och 4 fungerar. Sannolikheten för att 1 eller 2 fungerar är 1 minus sannolukheten för att båda inte fungerar, dvs. $1 - (1-p)^2$ och därför blir sannolikheten för att hela systremet fungerar

$$\left(1 - (1-p)^2\right) \cdot p \cdot p = (2p - p^2) \cdot p^2 = 2p^3 - p^4.$$

Då $p = 0.9$ blir denna sannolikhet 0.802

---

3. Antag att vi vet att 2% av exemplaren av typ A av en produkt har något fel medan 10% av typ B av samma produkt har något fel. Dessutom vet vi att 30% av de tilverkade exemplaren är av typ A och resten av typ B. Om vi nu har ett felfritt exemplar av produkten, vad är sannolikheten att den är av typ A?

Lösning: Låt A vara händelsen att exemplaret är av typ A och B vara händelsen att den är av typ B. Låt dessutom X vara händelsen att ett exemplar av produkten är felfritt. Detta betyder att

$$\Pr(A) = 0.3, \quad \Pr(B) = 0.7, \quad \Pr(X|A) = 0.98 \quad \text{och} \quad \Pr(X|B) = 0.9.$$

Enligt Bayes formel får vi nu

$$\Pr(A|X) = \frac{\Pr(X|A)\Pr(A)}{\Pr(X|A)\Pr(A) + \Pr(X|B)\Pr(B)}$$
$$= \frac{0.98 \cdot 0.3}{0.98 \cdot 0.3 + 0.9 \cdot 0.7} = \frac{0.294}{0.294 + 0.63} = 0.318,$$

---

4. Frekvensfunktionen för den tvådimensionlla diskreta slumpvariabeln (X, Y) är

$$f_{XY}(x,y) = \begin{cases} \frac{1}{48}(x+y), & 0 \le y \le x \le 3, \quad x \text{ och } y \text{ heltal,} \\ 0 & \text{annars.} \end{cases}$$

(a) Kontrollera att f_{XY} är en frekvensfunktion.
(b) Bestäm $\mathrm{E}(X|Y=2)$.

Lösning: Vi skriver först frekvensfunktionen i tabellform där vi också räknat ut rad- ock kolumnsummorna, dvs. marginalfördelningarna:

$f_{XY}(x,z)$		Y				
		0	1	2	3	$f_X(x)$
X	0	0	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{6}{48}$
	1	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{10}{48}$
	2	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{14}{48}$
	3	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{6}{48}$	$\frac{18}{48}$
$f_Y(y)$		$\frac{6}{48}$	$\frac{10}{48}$	$\frac{14}{48}$	$\frac{18}{48}$	1

(a) Av denhär tabellen ser vi att $f_{XY}(x,y) \ge 0$ för alla x och y och $\sum_x \sum_y f_{XY}(x,y) = 1$ så att det är verkligen frågan om en frekvensfunktion.
(b) Enligt definitionen är slumpvariablens X betingade frekvensfunktion under villkoret $Y = 2$ funktionen $f_{X|Y}(x|2) = \frac{f_{XY}(x,2)}{f_Y(2)}$ så vi får följande tabell:

X	0	1	2	3
$f_{X\|Y}(x\|2)$	$\frac{2}{14}$	$\frac{3}{14}$	$\frac{4}{14}$	$\frac{5}{14}$

Det betingade väntevärdet är således

$$\mathrm{E}(X|Y=2) = 0 \cdot \frac{2}{14} + 1 \cdot \frac{3}{14} + 2 \cdot \frac{4}{14} + 3 \cdot \frac{5}{14} = \frac{26}{14}.$$

---

5. Vi antar att en slumpvariabel X är exponentialfördelad med parametern λ så att täthetsfunktionen är $f(x) = \lambda \mathrm{e}^{-\lambda x}$ och väntevärdet $\mathrm{E}(X) = \frac{1}{\lambda}$. Av denna slumpvariabel har vi fått ett observerat stickprov med värdena 0.05, 0.15, 6.4, 0.6 och 0.8.

(a) Använd momentmetoden för att räkna ut ett estimat för λ.

(b) Förklara varför ”maximum likelihood”-metoden alltid ger samma svar som momentmetoden för exponentialfördelningen.

Lösning: (a) Medelvärdet av stickprovet blir

$$\frac{1}{5}(0.05 + 0.15 + 6.4 + 0.6 + 0.8) = 1.6,$$

vilket betyder att estimatet av λ med momentmetoden blir $\frac{1}{1.6} = 0.625$.
(b) Om vi har ett observerat stickprov $x_1, x_2, \ldots, x_n$ så är likelihoodfunktionen för exponentialfördelningen

$$L(\lambda) = \lambda \mathrm{e}^{-\lambda x_1} \lambda \mathrm{e}^{-\lambda x_1} \ldots \lambda \mathrm{e}^{-\lambda x_n} = \lambda^n \mathrm{e}^{-\lambda \sum_{j=1}^{n} x_j}.$$

För att bestämma maximivärdet av denhär funktionen bestämmer vi dess derivatas nollställen och får då

$$0 = n\lambda^{n-1}\mathrm{e}^{-\lambda \sum_{j=1}^{n} x_j} - \lambda^n \mathrm{e}^{-\lambda \sum_{j=1}^{n} x_j} \sum_{j=1]^{n} x_j,}$$

vilket ger som enda lösning

$$\lambda = \frac{1}{\frac{1}{n}\sum_{j=1}^{n} x_j},$$

vilket är samma resultat som man får med momentmetoden.

---

6. En torghandlare hävdar att det i varje morotspåse finns minst 500 g morötter. Mängden morötter i en påse varierar förstås slumpmässigt och vi kan anta vikten är normalfördelad. Om vi skall ta ett stickprov och genomföra en hypotespröving, så vad är nollhypotsesen? När vi tagit ett stickprov med 36 påsar blev stickprovsmedelvärdet 498 och stickprovsvariansen 27 g^2. Testa nollhypotesen på 5% och 1% signifikansnivå.

Lösning: Nollhypotsen är $\mathrm{H}_0 : \mathrm{E}(X) \geq 500$g så att alternativet är ensidigt. När vi räknar ut testvariablen antar vi att vikten av en morotspåse har fördelningen $\mathrm{N}(500, \sigma^2)$ (även om nollhypotesen innehåller möjligheten att väntevärdet är större än 500). Som testvariabel kan vi använda

$$\frac{\overline{X} - 500}{\frac{1}{\sqrt{n}} S}$$

som har t-fördelningen med $n - 1 = 35$ frihetsgrader.

Vi har följande numeriska värden: $\overline{x} = 498$ g, $s^2 = 27$, $n = 36$ och $\mu_0 = 500$ så att värdet av testvariabeln blir

$$t = \frac{\overline{x} - \mu_0}{\frac{1}{\sqrt{n}} s} = \frac{498 - 500}{\sqrt{\frac{27}{36}}} \approx -2.31.$$

Eftersom den alternativa hypotesten är ensidig och testvariabeln är t(35)-fördelad så blir de kritiska värdena $t_{0.05}(35) = 1.69$ och $t_{0.01}(35) = 2.438$. Eftersom $-2.438 < -2.31 < -1.69$ så förkastas nollhytpotesen på signifikansnivån 0.05 men inte på signifikansnivån 0.01.

---

7. Inför ett val förväntas kandidat X ha ett lika stort stöd inom området A som inom området B. Av 300 tillfrågade personer i området A sade sig 58% stöda kandidat X medan av 200 tillfrågade personer i området B sade sig 48% stöda kandidat X. Vad är nollhypotesen i detta fall? Testa nollhypotesen på signifikansnivån 1%.

Lösning: Nollhypotesen är att sannolikheten att en tillfrågad person stöder kandidat X är densamma i båda områdena. Låt denna sannolikhet vara p_0. Den approximativa testvariabeln blir då

$$\frac{\overline{X} - \overline{Y}}{\sqrt{p_0(1-p_0)(\frac{1}{n_A} + \frac{1}{n_B})}},$$

då X_j, $j = 1, \ldots, n_A$ är stickprovet i området A och Y_j, $j = 1, \ldots, n_B$ är stickprovet i området A. Vi vet inte vad p_0 är men estimerar det med $(n_A\overline{X} + n_B\overline{Y})/(n_A + n_B)$ vilket i detta fall blir 0.54. Värdet av testvariabeln blir därför

$$\frac{0.58 - 0.48}{\sqrt{0.54 \cdot 0.46 \cdot (\frac{1}{300} + \frac{1}{200})}} \approx 2.12.$$

Testvariabeln är approximativt normalfördelad och eftersom alternativet till nollhypotesen är dubbelsidigt så blir de kritiska värdena $z_{0.005} = \pm 2.5758$. Eftersom $-2.5758 < 2.12 < 2.5758$ så förkastar vi inte nollhypotesen.

---

8. Följande data har insamlats:

x	-3	0	1	2
y	6.7	1.3	1.2	-1.2

Antag att talen y_i är observerade värden av slumpvariablerna $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ där $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$ är oberoende av varandra. Bestäm med minsta kvadratmetoden estimat för regressionskoefficienterna β_0 och β_1 och pröva på signifikansnivån 0.05 nollhypotesen $\beta_1 = 0$. Du kan använda dig av att $\sum_{i=1}^4 (y_i - \overline{y})^2 = 33.46$ och $\sum_{i=1}^4 x_i y_i = -21.3$.

Lösning: Först skall vi räkna ut medelvärdena och de blir $\overline{x} = 0$ och $\overline{y} = 2$. Detta innebär att $s_x^2 = \frac{1}{3}\sum_{i=1}^4 (x_i - \overline{x})^2 = \frac{1}{3}(9 + 0 + 1 + 4) = 4.6667$, $s_y^2 = \frac{1}{3}33.46 = 11.153$ och $s_{xy} = \frac{1}{3}\sum_{i=1}^4 (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{3}\sum_{i=1}^4 x_i y_i - \frac{\overline{y}}{3}\sum_{i=1}^4 x_i = -\frac{1}{3} \cdot 21.3 = -7.1$. Nu får vi estimaten

$$b_1 = \frac{s_{xy}}{s_x^2} = -1.521 \quad \text{ja} \quad b_0 = \overline{y} - b_1\overline{x} = 2.$$

Dessutom kan vi konstatera att korrelationskoefficienten är

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = -0.984,$$

så att estimatet av restvariansen blir

$$s^2 = \frac{n-1}{n-2}s_y^2(1 - r_{xy}^2) = 0.531.$$

Eftersom vi skall testa nollhypotesen $\beta_1 = 0$ så blir testvariabeln

$$\frac{B_1}{S\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}},$$

som har en $\mathrm{t}(n-2) = \mathrm{t}(2)$-fördelning. Värdet av testvariablen blir i detta fall

$$\frac{-1.521}{\sqrt{\frac{0.531}{14}}} = -7.81.$$

Eftersom nollhypotesen är $\beta_1 = 0$ så är alternativet tvåsidigit och de kritiska värdena är $\pm t_{0.025}(2) = \pm 4.303$. Eftersom $-7.81 < -4.303$ så kan vi förkasta nollhypotetsen på signifikansnivån 0.05.
