Aalto-yliopisto

Gripenberg, Kiiski

MS-A0502 Todennäköisyyslaskennan ja tilastotieteen peruskurssi
1. välikoe 20.11.2014

Kirjoita jokaiseen koepaperiin nimesi, opiskelijanumerosi ym. tiedot!
Laskin, I. Mellinin tilastolliset taulukot, Matlab/Octave-funktiolista ja muistiinpanolappu ovat sallittuja apuvälineitä!
Kirjoita välivaiheet näkyviin!

1. Virheetöntä noppaa heitetään kaksi kertaa.

(a) Määritä $\Pr(A|B)$ kun A on tapahtuma *"Silmälukujen summa on* 8*"* ja B on tapahtuma *"Toisella heitolla tulos on* 1, 2 *tai* 3*"*.

(b) Ovatko tapahtumat *"Toisella heitolla tulos on vähintään* 5*"* ja *"Silmälukujen summa on* 7*"* riippumattomia? Perustele!

Ratkaisu: Kaikki mahdolliset tulokset, eli alkeistapahtumat ovat muotoa (x,y) missä $x,y \in \{1,2,3,4,5,6\}$. Jokaisen alkeistapahtuman tai alkeistapahtumasta muodostetun tapahtuman todennäköisyys on $\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}$.
(a) 1. tulkinta, "toinen"="jälkimmäinen": Tapahtuma A on joukko $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. Tapahtuma B on $\{\,(x,y) : x \in \{1,2,3,4,5,6\},\ y \in \{1,2,3\}\,\}$ jossa on 18 alkiota, josta seuraa että sen todennäköisyys on $\Pr(B) = \dfrac{1}{2}$. Tapahtuma $A \cap B$ on joukko $\{(5,3),(6,2)\}$ joten $\Pr(A\cap B) = \dfrac{2}{36} = \dfrac{1}{18}$. Näin ollen

$$\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)} = \frac{\frac{1}{18}}{\frac{1}{2}} = \frac{1}{9}.$$

2. tulkinta "toinen"="edellinen tai jälkimmäinen": Tapahtuma A on joukko $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. Tapahtuma B on joukko $\{\,(x,y) : x \in \{1,2,3\},\ y \in \{1,2,3,4,5,6\}\,\} \cup \{\,(x,y) : x \in \{1,2,3,4,5,6\},\ y \in \{1,2,3\}\,\}$ jossa on 27 alkiota, josta seuraa että sen todennäköisyys on $\Pr(B) = \dfrac{27}{36} = \dfrac{3}{4}$. Tapahtuma $A \cap B$ on joukko $\{(2,6),(3,5),(5,3),(6,2)\}$ joten $\Pr(A\cap B) = \dfrac{4}{36} = \dfrac{1}{9}$. Näin ollen

$$\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)} = \frac{\frac{1}{9}}{\frac{3}{4}} = \frac{4}{27}.$$

(b) 1. tulkinta, "toinen"="jälkimmäinen": Jos tässä tapauksessa A on tapahtuma "Toisella heitolla tulos on vähintään 5" ja ja B on tapahtuma "Silmälukujen summa on 7" niin $A = \{\,(x,y) : x \in \{1,2,3,4,5,6\},\ y \in \{5,6\}\,\}$ josta seuraa, että $\Pr(A) = \dfrac{12}{36} = \dfrac{1}{3}$ ja B on tapahtuma $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ josta seuraa, että $\Pr(B) = \dfrac{6}{36} = \frac{1}{6}$. Lisäksi

$A \cap B = \{(1,6),(2,5)\}$ joten $\Pr(A \cap B) = \dfrac{2}{36} = \dfrac{1}{18}$. Näin ollen

$$\Pr(A) \cdot \Pr(B) = \frac{1}{3} \cdot \frac{1}{6} = \frac{1}{18} = \Pr(A \cap B),$$

joten A ja B ovat riippumattomia.

2. tulkinta ”toinen”=”edellinen tai jälkimmäinen”: Jos tässä tapauksessa A on tapahtuma ”Toisella heitolla tulos on vähintään 5” ja ja B on tapahtuma ”Silmälukujen summa on 7” niin $A = \{\, (x,y) : x \in \{5,6\},\, y \in \{1,2,3,4,5,6\}\,\} \cup \{\, (x,y) : x \in \{1,2,3,4,5,6\},\, y \in \{5,6\}\,\}$ josta seuraa, että $\Pr(A) = \dfrac{20}{36} = \dfrac{5}{9}$ ja B on tapahtuma $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ josta seuraa, että $\Pr(B) = \dfrac{6}{36} = \frac{1}{6}$. Lisäksi $A \cap B = \{(6,1), 5,2), (1,6), (2,5)\}$ joten $\Pr(A \cap B) = \dfrac{4}{36} = \dfrac{1}{9}$. Näin ollen

$$\Pr(A) \cdot \Pr(B) = \frac{5}{9} \cdot \frac{1}{6} = \frac{5}{54} \neq \Pr(A \cap B),$$

joten A ja B eivät ole riippumattomia.

---

2. Laiva haaksirikkoutuu ja suuri joukko muovileluja joutuu mereen. Näistä 40% ovat punaisia ja 60% sinisiä. Sinisistä 60% ja punaisista 80% uppoavat ja muut huuhtoutuvat rannalle. Löydat rannalta muovilelun, joka on peräisin tältä laivalta. Mikä on todennäköisyys, että se on punainen.

Ratkaisu: Olkoon P tapahtuma ”Lelu on punainen”, S tapahtuma ”Lelu on sininen” ja R tapahtuma ”Lelu huuhtoutuu rannalle”. Annettujen tietojen perusteella $\Pr(P) = 0.4$, $\Pr(S) = 0.6$, $\Pr(R|P) = 0.2$ ja $\Pr(R|S) = 0.4$. Bayesin kaavan nojalla saadaan nyt

$$\Pr(P|R) = \frac{\Pr(R|P) \cdot \Pr(P)}{\Pr(R|P) \cdot \Pr(P) + \Pr(R|S) \cdot \Pr(S)} = \frac{0.2 \cdot 0.4}{0.2 \cdot 0.4 + 0.4 \cdot 0.6} = \frac{0.08}{0.32} = 0.25.$$

Toisella tavalla: Jos laivassa oli 1000 lelua niin näistä 400 ovat punaisia ja 600 sinisiä. Punaisista 20% eli 80 huuhtoutuvat rannalle ja samoin sinisistä 40% eli 240 kappaletta. Rannalle on siten tullut 320 lelua, joista 80 ovat punaisia joten todennäköisyys että löytämäsi lelu on punainen on $\frac{80}{320} = 0.25$.

3.

(a) Eräällä kurssilla on 40 opiskelijaa koulutusohjelmasta A ja 60 opiskelijaa koultusohjelmasta B. Jos harjoitusryhmään valitaan näistä opiskelijoista satunnaisesti 15, niin mikä on todennäköisyys, että tähän ryhmään kuuluu 10 opiskelijaa koulutusohjelmasta A ja 5 kolutusohjelmasta B.

(b) Kuuden viikon aikana valitaan joka viikko satunnaisesti yksi opiskelija harjoitusryhmästä, johon kuuluu 10 opiskelijaa koulutusohjelmasta A ja 5 kolutusohjelmasta B, ja tämän opiskelijan harjoitustehtävien ratkaisuja tarkistetaan erittäin huolellisesti. Mikä on todennäköisyys, että 4 kertaa valitaan opiskelija koulutsohjelmasta A ja 2 kertaa koulutusohjelmasta B?

Anna tarkkoja vastauksia, mutta ne saavat sisältää $+$, $-$, $\cdot$, $/$ ja ! mutta ei esim. binomikertoimia.

Ratkaisu: (b) Koulutusohjelmasta A on valittava 10 opiskelijaa ja tämä on mahdollista tehdä $\binom{40}{10}$ eri tavalla ja koulutusohjelmasta B on valittava 5 opiskelijaa ja tämä on mahdollista tehdä $\binom{60}{5}$ eri tavalla. Tuloperiaatteen nojalla voidaan $\binom{40}{10} \cdot \binom{60}{5}$ eri tavalla valita 10 opiskelijaa ohjelmasta A ja 5 ohjelmasta B. Koska kurssilla on yhteensä 100 opiskelijaa niin näiden joukosta voidaan valita 15 opiskelijaa $\binom{100}{15}$ eri tavalla. Todennäköisyys, että ryhmään kuuluu 10 opiskelijaa koulutusohjelmasta A ja 5 kolutusohjelmasta B on siten

$$\frac{\binom{40}{10} \cdot \binom{60}{5}}{\binom{100}{15}} = \frac{40! \cdot 60! \cdot 15! \cdot 85!}{10! \cdot 30! \cdot 5! \cdot 55! \cdot 100!} = \frac{36665668}{2006438695}.$$

(b) Tämä on valinta takaisinpanolla koska on mahdollista että sama opiskelija valitaan monta kertaa. Näin ollen todennäköisyys, että opiskelija koultusohjelmasta A tulee valituksi on $\frac{2}{3}$ ja koulutusohjelmasta B vastaavasti $\frac{1}{3}$. Koska tämä valinta toistetaan 6 kertaa niin binomijakauman mukaisesti todennäköisyys että valitaan 4 kertaa opiskelija koulutusohjelmasta A ja 2 kertaa koulutusohjelmasta B on

$$\binom{6}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2 = \frac{6! \cdot 2^4}{4! \cdot 2! \cdot 3^6} = \frac{80}{243}.$$

4.

(a) Miksi ei ole järkevää olettaa, että tiettynä ajanjaksona tapahtuvien liikenneonnettomuuksien lukumäärä noudattaisi eksponenttijakaumaa? (Vastaukseksi ei kelpaa että (b)-kohdassa puhutaan Poisson-jakaumasta.)

(b) Oletetaan, että eräässä maassa vuoden aikana tapahtuvien liikenneonnettomuuksien lukumäärä on Poisson-jakautunut satunnaismuuttuja odotusarvolla $14\,400$. Mikä on todennäköisyys, että vuoden aikana tapahtuvien liikenneonnettomuuksien lukumäärä on vähintään $14\,100$ ja enintään $14\,700$. Käytä normaaliapproksimaatiota ja käytä joko taulukoita tai kirjoita millä Matlab/Octave-komennoilla saisit lopullisen vastauksen.

Ratkaisu: (a) Exponenttijakauma on jatkuva jakauma ja tästä seuraa mm. että todennäköisyys, että eksponenttijakautunut satunnaismuuttuja saa kokonaislukuarvoja on 0, joten ei ole järkevää olettaa, että mikään lukumäärä olisi eksponenttijakautunut. Jos kuitenkin ajatellaan, että kyse olisi vain approksimaatiosta niin silloin normaalijakauma olis paljon parempi valinta.

(b) Jos X on liikenneonnettomuuksien lukumäärä vuoden aikana ja X on Poisson-jakautunut odotusarvolla $14\,400$ niin $X \sim \text{Poisson}(14\,400)$ jolloin X:n varianssi on myös $14\,400$. Näin ollen

$$\frac{X - 14\,400}{\sqrt{14\,400}} \sim_a \mathrm{N}(0,1),$$

ja

$$\Pr(14\,100 \le X \le 14\,700) = \Pr\left(\frac{14\,100 - 14\,400}{\sqrt{14\,400}} \le \frac{X - 14\,400}{\sqrt{14\,400}} \le \frac{14\,700 - 14\,400}{\sqrt{14\,400}}\right)$$

$$= \Pr\left(-2.5 \le \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le 2.5\right) \approx F_{\mathrm{N}(0,1)}(2.5) - F_{\mathrm{N}(0,1)}(-2.5) = 2(1 - F_{\mathrm{N}(0,1)}(-2.5)).$$

Taulukosta saadaan $F_{\mathrm{N}(0,1)}(2.5) = 0.9938$ ja $F_{\mathrm{N}(0,1)}(-2.5) = 0.0062$ ja ja näiden lukujen erotus voidaan Matlab/Octavessa laskea komennolla `normcdf(2.5)- normcdf(-2.5)` jolloin vastaukseksi tulee 0.9876.
