Gripenberg

MS-A0502 Todennäköisyyslaskennan ja tilastotieteen peruskurssi
Tentti ja välikoeuusinta 7.1.2015

Kirjoita jokaiseen koepaperiin nimesi, opiskelijanumerosi ym. tiedot ja minkä kokeen suoritat!
Laskin, I. Mellinin tilastolliset taulukot, Matlab/Octave-funktiolista ja muistiinpanolappu ovat sallittuja apuvälineitä!
Kirjoita välivaiheet näkyviin!
Käytä joko taulukoita tai kirjoita millä Matlab/Octave-komennoilla voisit laskea tarvittavat kertymäfunktioiden tai niiden käänteisfunktioiden arvot ja minkälaisia päätelmiä, esimerkiksi merkitsevyystasojen suhteen, näiden arvojen perusteella voisit tehdä.

Tentti: Tehtävät 1, 4, 6, 7 ja 8.
1. välikoe: Tehtävät 1, 2, 3 ja 4.
2. välikoe: Tehtävät 5, 6, 7 ja 8.

1.

(a) Heitetään virheetöntä noppaa kaksi kertaa. Mikä on todennäköisyys, että silmälukujen summa on 9?

(b) Heitetään virheetöntä noppaa 900 kertaa. Mikä on todennäköisyys, että saadaan 6 korkeintaan 130 kertaa. Käytä normaaliapproksimaatiota.

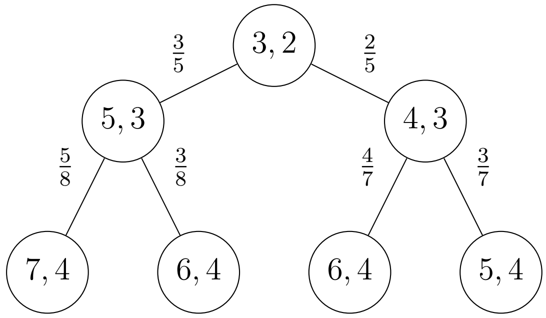
Ratkaisu: (a) Koska jokaisella heitolla saadaan 1, 2, 3, 4, 5 tai 6 todennäköisyydellä $\frac{1}{6}$ ja koska $9 = 3+6 = 4+5 = 5+4 = 6+3$ niin todennäköisyys saada 9 on $4 \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{9}$.
(b) Todennäköisyys saada on 6 on $\frac{1}{6}$ joten jos X on kuutosten luukumäärä 900:ssa heitossa niin $X \sim \mathrm{Bin}(900, \frac{1}{6})$. Näin ollen $\mathrm{E}(X) = 900 \cdot \frac{1}{6} = 150$ ja $\mathrm{Var}(X) = 900 \cdot \frac{1}{6} \cdot \frac{5}{6} = 125$. Koska tämä luku on > 10 voimme käyttää normaaliapproksimaatiota ja saamme

$$\begin{aligned}\Pr(X \le 130) &= \Pr\left(\frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le \frac{130 - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}}\right) \\ &= \Pr\left(\frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} \le -1.7889\right) \approx 0.036819.\end{aligned}$$

---

2. Uurnassa on 3 valkoista ja 2 mustaa palloa. Poimitaan uurnasta satunnaisesti pallo. Jos se on valkoinen niin uurnaan laitetaan takaisin 3 valkoista palloa ja 1 musta pallo ja jos se on musta niin uurnaan laitetaan takaisin 2 mustaa palloa ja 1 valkoinen. Tämän jälkeen poimitaan uurnasta taas satunnaisesti uusi pallo ja menetellään samalla tavalla kuin ensimmäisellä kerralla. Määritä tämän jälkeen uurnassa olevien valkoisten pallojen lukumäärän pistetodennäköisyysfunktio. Käytä puuverkkoa.

Ratkaisu: Puuverkko näyttää seuraavanlaiselta missä solmuissa olevat numerot m, n tarkoittavat, että uurnassa on m valkoista ja n mustaa palloa ja vasemmanpuolinen kaari valitaan jos poimitaan valkioinen pallo ja oikeanpuolinen jos poimitaan musta pallo:



Uurnassa on poimintojen ja takaisinpanojen jälkeen 5, 6 tai 7 valkoista palloa ja jos X on valkoisten pallojen lukumäärä niin

$$\begin{aligned}
\Pr(X = 5) &= \frac{2}{5} \cdot \frac{3}{7} = \frac{6}{35}, \\
\Pr(X = 6) &= \frac{3}{5} \cdot \frac{3}{8} + \frac{2}{5} \cdot \frac{4}{7} = \frac{127}{280}, \\
\Pr(X = 7) &= \frac{3}{5} \cdot \frac{5}{8} = \frac{3}{8}.
\end{aligned}$$

3.

Pistetodennäköisyysfunktion f_{XY} arvot on annettu seuraavassa taulukossa:

$f_{XY}(x,y)$		Y		
		1	2	3
X	0	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{1}{20}$
	1	0	$\frac{3}{20}$	$\frac{2}{20}$
	2	$\frac{1}{20}$	$\frac{1}{20}$	0
	3	$\frac{1}{20}$	$\frac{4}{20}$	$\frac{3}{20}$

(a) Mistä nähdään, että kyseessä todella on pistetodennäköisyysfunktio.
(b) Määritä satunnaismuuttujan X pistetodennäköisyysfunktio (eli määritä sen reunajakauma).
(c) Määritä satunnaismuuttujan $(X|Y = 2)$ pistetodennäköisyysfunktio eli X:n jakauma ehdolla $Y = 2$.

Ratkaisu: (a) f_{XY} on pistetodennäköisyysfunktio koska $f_{XY}(x,y) \geq 0$ kaikilla x ja y ja $\sum_{x=0}^{3} \sum_{y=1}^{3} f_{XY}(x,y) = 1$.

(b) Satunnaismuuttujan X pistetodennäköisyysfunktio on $f_X(x) = \sum_{y=1}^{3} f_{XY}(x,y)$ jonka arvot ovat

x	0	1	2	3
$f_X(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{2}{5}$

(c) Koska satunnaismuuttujan $(X|Y=2)$ pistetodennäköisyysfunktio on

$$f_{X|Y}(x|2) = \frac{f_{XY}(x,2)}{f_Y(2)}$$

ja $f_Y(2) = \Pr(Y=2) = \frac{2}{20} + \frac{3}{20} + \frac{1}{20} + \frac{4}{20} = \frac{1}{2}$ niin tämän funktion arvot ovat

x	0	1	2	3
$f_{X\|Y}(x\|2)$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{5}$

---

4. Alueen sorasta on 70% hyvää ja 30% huonoa tiettyyn rakennustarkoitukseen. Yhdestä alueen sorakuopasta otettu näyte testattiin testillä, joka todennäköisyydellä 0.8 osoittaa soran hyväksi (ja todennäkösyydellä 0.2 huonoksi), jos sora on hyvää ja todennäköisyydellä 0.9 huonoksi (ja todennäköisyydellä 0.1 hyväksi), jos sora on huonoa. Jos testi näytti että sora oli hyvää, mikä on todennäköisyys että kuopan sora todella on hyvää?

Ratkaisu: Olkoon A tapahtuma, että kuopan sora on hyvää jolloin siis $\Pr(A) = 0.7$ ja $\Pr(A^c) = 0.3$. Olkoon T tapahtuma, että testi osoittaa, että sora on hyvää jolloin $\Pr(T|A) = 0.8$ ja $\Pr(T|A^c) = 0.1$. Bayesin kaavan mukaa saadaan nyt

$$\Pr(A|T) = \frac{\Pr(T|A)\cdot\Pr(A)}{\Pr(T|A)\cdot\Pr(A) + \Pr(T|A^c)\cdot\Pr(A^c)} = \frac{0.8\cdot 0.7}{0.8\cdot 0.7 + 0.1\cdot 0.3} = \frac{0.56}{0.59} \approx 0.95.$$

Toisella tavalla: Jos alueella olisi 1000 sorakuoppaa niin 700 olisivat sellaisia missä sora on hyvää ja 300 sellaisia missä sora on huonoa. Sorakuoppien lukumäärät, joissä testi osoittaisi, että sora on hyvää olisivat $700\cdot 0.8 = 560$ ja $300\cdot 0,1 = 30$ eli yhteensä 590. Näin ollen todennäköisyys, että kuopan sora on hyvää jos testi osoittaa, että näin on asian laita on

$$\frac{560}{590} \approx 0.95.$$

5.

(a) Satunnaismuuttujasta X, joka on Poisson(λ)-jakautunut on saatu havainnot 5, 6, 4, 7 ja 6. Laske momenttimenetelmällä λ:n estimaatti.

(b) Diskreetin satunnaismuuttujan Y jakauma riippuu parametristä $\theta \in [0, \frac{1}{2}]$ siten, että

$$\Pr(Y=j) = \begin{cases} 1-2\theta, & j=0, \\ \theta, & j \in \{-1,1\}, \\ 0, & \text{muuten}. \end{cases}$$

Satunnaismuuttujasta Y on saatu seuraavat havaintoarvot: 0, 1, -1, 0, 1, 0, -1, 1, -1 ja 0. Määritä θ:n estimaatti suurimman uskottavuuden menetelmällä.

Ratkaisu: (a) Koska $\mathrm{E}(X) = \lambda$ niin momenttimenetelmän mukaisesti λ:n estimaatiksi otetaan havaintoarvojen kesiarvo, eli tässä tapauksessa $\hat{\lambda} = \frac{1}{5}(5+6+4+7+6) = \frac{28}{5} = 5.6$.

(b) Koska satunnaismuuttuja on diskreetti niin uskottavuusfunktio on $\mathcal{L}(\theta)$ on todennäköisyys, että saadaan havaitut arvot eli

$$\mathcal{L}(\theta) = \Pr(Y=0)\cdot\Pr(Y=1)\cdot\Pr(Y=-1)\cdot\Pr(Y=0)\cdot\Pr(Y=1)\cdot\Pr(Y=0) \cdot\Pr(Y=-1)\cdot\Pr(Y=0)\cdot\Pr(Y=0) = (1-2\theta)^4\cdot\theta^6.$$

Maksimiarvokohdan löytämiseksi määritämme derivaatan nollakohdan:

$$\mathcal{L}'(\theta) = 4\cdot(1-2\theta)^3\cdot(-2)\cdot\theta^6 + (1-2\theta)^4\cdot 6\cdot\theta^5 = (1-2\theta)^4\cdot\theta^6\cdot\left(\frac{-8}{1-2\theta}+\frac{6}{\theta}\right) = 0,$$

josta saamme (tapausten $\theta = 0$ tai $\theta = \frac{1}{2}$) lisäksi

$$-8\cdot\theta + 6 - 12\cdot\theta = 0,$$

josta saamme ratkaisuksi $\theta = \dfrac{3}{10}$.

---

6. Sairaalan poliklinikalle tulee kokemuksen perusteella keskimäärin 9 potilasta tunnissa. Eräänä päivänä, kun keli on ollut erityisen liukas tulee 12 tunnin aikana 130 potilasta. Oletetaan, että potilaiden lukumäärä aikayksikössä noudattaa Poisson-jakaumaa. Testaa merkitsevyystasolla 0.05 nollahypoteesia, jonka mukaan potilaiden lukumäärän odotusarvo kyseisenä päivänä ei ole ainakaan tavanomaista suurempi. Käytä normaaliapproksimaatiota.

Ratkaisu: Olkoon X kyseisenä päivänä 12 tunnin aikana tulevien potilaiden lukumäärä. Koska normaalipäivänä tulee tunnissa keskimäärin 9 potilasta niin nollahypoteesi tarkoittaa, että $\mathrm{E}(X) \le 9\cdot 12 = 108$. Koska $\mathrm{Poisson}(\lambda)$ jakautuneen satunnaissuureen odotusarvo on λ niin voimme nollahypoteesin mukaisesti olettaa, että $X \sim \mathrm{Poisson}(\lambda)$ missä $\lambda \le 108$. Koska Poisson jakauma lähestyy normaalijakumaa kun parametri kasvaa äärettömyyteen ja koska $\mathrm{Poisson}(\lambda)$ jakautuneen satunnaissuureen varianssi on λ niin voimme käyttää testimuuttujana

$$\frac{X-108}{\sqrt{108}} \sim_a \mathrm{N}(0,1).$$

Tässä tapauksessa testisuureen arvo on $\frac{22}{\sqrt{108}} = 2.117$ ja koska vaihtoehto on yksisuuntainen niin p-arvoksi tulee

$$p = 1 - F_{\mathrm{N}(0,1)}(2.117) = 0.01713,$$

ja koska $p < 0.05$ niin nollahypoteesi hylätään.

---

7. Havainnot X_i, $i = 1,\ldots,21$ ja Y_i, $i = 1,\ldots,31$ ovat satunnaisotoksia jakaumista $\mathrm{N}(\mu_X,\sigma^2)$ ja $\mathrm{N}(\mu_Y,\sigma^2)$. Havaintoarvoista on laskettu $\overline{x} = 2.8$, $\sum_{i=1}^{21}|x_i-\overline{x}| = 26$, $\sum_{i=1}^{21}(x_i-\overline{x})^2 = 46$, $\overline{y} = 2.0$, $\sum_{i=1}^{31}|y_i-\overline{y}| = 35$ ja $\sum_{i=1}^{31}(y_i-\overline{y})^2 = 81$. Testaa merkitsevyystasolla 0.05 hypoteeseja

(a) $\mathrm{H}_0 : \mu_X = 3.5$, (jolloin sinun ei pidä ottaa huomioon y_i-havaintoarvoja),

(b) $\mathrm{H}_0 : \mu_X = \mu_Y$.

Ratkaisu: (a) Ensin lasketaan X-satunnaismuuttujan ostosvarianssi ja se on

$$s_x^2 = \frac{1}{20}\sum_{i=1}^{21}(x-\overline{x})^2 = 2.3.$$

Koska nollahypoteesi on $\mu_X = 3.5$ niin testisuure on

$$w = \frac{\overline{x} - 3.5}{\sqrt{\frac{s_x^2}{21}}} = \frac{2.8 - 3.5}{\sqrt{\frac{2.3}{21}}} = -2.1152.$$

Koska nollahypoteesin vaihtoehto on kaksisuuntainen niin p-arvoksi tulee

$$p = 2F_{\mathrm{t}(20)}(-2.1152) = 2 \cdot 0.0236 = 0.0472.$$

Koska $p < 0.05$ niin nollahypoteesi hylätään.

(b) Ensin lasketaan Y-satunnaismuuttujan otosvarianssi ja saadaan

$$s_y^2 = \frac{1}{30} \sum_{i=1}^{31} (y - \overline{y})^2 = 2.7.$$

Seuraavaksi lasketaan varianssin yhdistetty estimaatti ja saadaan

$$s_p^2 = \frac{(21-1)s_x^2 + (31-1)s_y^2}{21 + 31 - 2} = 2.54.$$

Testisuure on

$$\frac{\overline{x} - \overline{y}}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{0.8}{\sqrt{2.54 \cdot \left(\frac{1}{21} + \frac{1}{31}\right)}} = 1.7761.$$

Tässäkin tapauksessa vaihtoehto on kaksisuuntainen ja p-arvoksi tulee

$$p = 2(1 - F_{\mathrm{t}(50)}(1.7761)) = 0.082.$$

Koska $p > 0.05$ niin nollahypoteesia ei hylätä.

---

8. Lukiokoulutusta antavien oppilaitosten lukumäärät olivat vuosina 2002–2013 seuraavat:

Vuosi	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Lkm.	484	483	479	471	461	449	449	441	439	433	428	421

Oletetaan, että (ainakin approksimatiivisesti) $Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$, missä Y_j on lukiokoulutusta antavien oppilaitosten lukumäärä vuonna x_j ja satunnaismuuttujat ε_j ovat $\mathrm{N}(0, \sigma^2)$-jakautuneita ja riippumattomia.

Määritä parametrin β_1 pienimmän neliösumman estimaatti ja testaa merkitsevyystasolla 0.01 nollahypoteesia, että lukiokoulutusta antavien oppilaitosten lukumäärän odotusarvo vuonna 2016 on korkeintaan 390.

Alla olevassa taulukossa on annettu joitakin tunnuslukuja:

$\overline{y}$	s_x^2	s_y^2	s_{xy}
453.17	13	485.97	-78.636

Ratkaisu: Parametrin β_1 estimaatiksi tulee

$$b_1 = \frac{s_{xy}}{s_x^2} = -6.049.$$

Jos merkitään $z_j = x_j - 2016$ niin $\overline{z} = -8.5$, $s_z^2 = s_x^2$ ja $s_{zy} = s_{xy}$ ja $Y_j = \tilde{\beta}_0 + \beta_1 z_j + \varepsilon_j$ missä $\tilde{\beta}_0$ on lukiokoulutusta antavien oppilaitosten lukumäärän odotusarvo vuonna 2016. Parametrin $\tilde{\beta}_0$ pienimmän neliösumman estimaatiksi tulee

$$\tilde{b}_0 = \overline{y} - b_1 \overline{z} = 401.75.$$

Koska $r_{zy} = r_{xy} = \frac{s_{zy}}{\sqrt{s_z^2 s_y^2}} = -0.98934$ niin jäännösvarianssin estimaatiksi tulee

$$s^2 = \frac{11/10^{2}}{s\quad_y}(1 - r_{zy}^2) = 11.34.$$

Testimuuttujan arvo on tässä tapauksessa

$$w_0 = \frac{\tilde{b}_0 - 490}{\sqrt{s^2(\frac{1}{12} + \frac{\bar{z}^2}{11 \cdot s_z^2})}} = \frac{\tilde{4}01.75 - 490}{\sqrt{11.34(\frac{1}{12} + \frac{(-8.5)^2}{11 \cdot 13})}} = 4.548.$$

Koska nollahypoteesi on $\tilde{\beta}_0 \leq 390$ jolloin siis vaihtoehto on yksisuuntainen niin p-arvoksi tulee

$$p = (1 - F_{\mathrm{t}(10)}(4.548)) = 0.00053.$$

Koska $p < 0.01$ niin nollahypoteesi hylätään.
