Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 1
Malliratkaisut

1. Kertausta todennäköisyyslaskennasta. Ilmoita satunnaismuuttujan Y jakauman nimi ja pistetodennäköisyys- tai tiheysfunktio seuraavissa tapauksissa:

a) $Y = X_1 + \cdots + X_n$, kun $X_1, \ldots, X_n \sim B(\theta) \perp\!\!\!\perp$ (otos Bernoulli-jakaumasta)
b) $Y = U + V$, kun $U \sim N(1,1)$, $V \sim N(3,4)$ ja $U \perp\!\!\!\perp V$
c) $Y = U^2$, kun $U \sim N(0,1)$
d) $Y = aX + b$, kun $X \sim \text{Tas}(0,1)$ ja $a > 0$, $b \in \mathbb{R}$
e) $Y = X_1 + \cdots + X_n$, kun $X_i \sim P(\mu_i)$ ja $X_1, \ldots, X_n \perp\!\!\!\perp$.

Ilmoita myös Y:n odotusarvo ja varianssi kussakin tapauksessa. Perusteluja ei vaadita.

Käytä tarvittaessa todennäköisyyslaskennan materiaalia ja taulukkokirjoja apuna. Huomaa myös, että tällä kurssilla eri jakaumista käytettävät tunnukset eroavat todennäköisyyslaskennan kurssilla käytetyistä; ks. liitettä muistiinpanojen sivulla 88.

Ratkaisu 1

a) Satunnaismuuttuja Y noudattaa binomijakaumaa parametrein n ja θ, eli $Y \sim \text{Bin}(n, \theta)$ ja sen pistetodennäköisyysfunktio on

$$f_Y(y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \text{ kun } y = 0, 1, \ldots, n$$

Y:n odotusarvo $E(Y) = n\theta$ ja varianssi $var(Y) = n\theta(1-\theta)$

b) Riippumattomien normaalijakautuneiden satunnaismuuttujien summa noudattaa normaalijakaumaa, jonka odotusarvo on summattavien satunnaismuuttujien odotusarvojen summa ja varianssi on niiden varianssien summa. Siis $Y \sim N(4,5)$ ja Y:n odotusarvo $E(Y) = 4$ ja varianssi $var(Y) = 5$. Y:n tiheysfunktio on

$$f_Y(y) = \frac{1}{\sqrt{10\pi}} e^{-\frac{(y-4)^2}{10}}$$

c) Riippumattomien standardinormaalijakaumaa noudattavien satunnaismuuttujien summa noudattaa χ^2-jakaumaa, jonka vapausaste on summattaven satunnaismuuttujien määrä. Satunnaismuuttuja Y siis noudattaa χ^2-jakaumaa vapausasteella 1, eli $Y \sim \chi^2_1$ ja sen tiheysfunktio on

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\ y^{-\frac{1}{2}}\ e^{-\frac{1}{2}y}, \text{ kun } y > 0$$

Y:n odotusarvo $E(Y) = 1$ ja varianssi $var(Y) = 2 \times 1 = 2$.

d) Ratkaistaan X yhtälöstä $Y = aX + b$. Nyt $X = \frac{Y-b}{a}$. Nyt muuttujanvaihtokaavan avulla saadaan selville satunnaismuuttujan Y tiheysfunktio.

$$f_Y(y) = f_X(h(y))\,|h'(y)|, \text{ jossa } h(y) = \frac{y-b}{a} \text{ ja } h'(y) = \frac{1}{a} \text{ ja } f_X(x) = \begin{cases} 1, \text{ kun } 0 \le x \le 1 \\ 0, \text{ muulloin} \end{cases}$$

Siis

$$
\begin{aligned}
f_Y(y) &= 1 \times \frac{1}{a}, \text{ kun } 0 \leq \frac{y-b}{a} \leq 1 \\
&= \frac{1}{a}, \text{ kun } \ b \leq y \leq a+b
\end{aligned}
$$

Y siis noudattaa välin $(b, a+b)$ tasajakaumaa, eli $Y \sim \text{Tas}(b, a+b)$ ja Y:n odotusarvo $E(Y) = \frac{(a+2b)}{2}$ ja varianssi $var(Y) = \frac{a^2}{12}$.

e) Poisson jakauman yhteenlaskuominaisuuden perusteella satunnaismuuttuja Y noudattaa Poisson jakaumaa odotusarvolla $\sum_{i=1}^{n} \mu_i$. Siis $Y \sim P(\sum_{i=1}^{n} \mu_i)$ ja sen odotusarvo $E(Y) = \sum_{i=1}^{n} \mu_i$ ja varianssi $var(Y) = \sum_{i=1}^{n} \mu_i$. Y:n pistetodennäköisyysfunktio on

$$
f_Y(y) = e^{-\sum_{i=1}^{n} \mu_i} \frac{(\sum_{i=1}^{n} \mu_i)^y}{y!}, \ \text{ kun } y = 0, 1, 2, \ldots
$$

2. a) Satunnaismuuttuja Y_1 on diskreetti, arvojoukkonaan $\{1, 2, 3, 4, 5\}$. Sille on kaksi vaihtoehtoista pistetodennäköisyysfunktiota, jotka on taulukoitu alla.

y	1	2	3	4	5
$f(y;1)$	0.2	0.5	0.2	0.1	0
$f(y;2)$	0	0.1	0.2	0.4	0.3

Emme tiedä, kumpaa näistä sattuma käyttää arpoessaan Y_1:n arvon.

Tulkitse tämä parametrisena tilastollisena mallina, jonka parametria merkitään θ:lla. Mikä on parametriavaruus? Jos havaitaan $Y_1 = 4$, mitä voit päätellä θ:sta? Kuinka varmasti?

b) Kyseisestä jakaumasta tehdään myös toinen havainto Y_2, jonka arvoksi saadaan $Y_2 = 1$. Mitä nyt voit päätellä θ:sta?

Ratkaisu 2

a) Tilastollisen mallin parametrin θ mahdolliset arvot ovat 1 ja 2, joten sen parametriavaruus on $\Omega = \{1, 2\}$.

Havaitaan $Y_1 = 4$. Parametrin θ arvolla 1 todennäköisyys saada kyseinen havainto on 0.1, kun taas arvolla 2 todennäköisyys on 0.4. Parametrin θ arvo 2 on siis uskottavampi, mutta molemmat arvot ovat mahdollisia.

b) Tehdään toinen havainto $Y_2 = 1$. Koska parametrin θ arvolla 2 kyseisen havainnon todennäköisyys on 0, voidaan varmasti sanoa, että $\theta = 1$.

3. Opiskele luentomonisteen esimerkki 1.2.2 (kestoiät). Millainen käytännön ongelma saattaa liittyä tällaisen satunnaiskokeen tekemiseen?

Päädytään käyttämään halvempaa koejärjestelyä, jossa n laitetta pannaan yhtä aikaa käyntiin ja 30 päivän kuluttua tullaan tarkastamaan, mitkä yksilöt (tai kuinka monta yksilöä) ovat menneet rikki. Muodosta tätä koeasetelmaa kuvaava tilastollinen malli parametrille μ (kestoikien odotusarvo).

Apu. Ajattele, että kyseessä on n-kertainen toistokoe (binomikoe), jossa ”onnistuminen” merkitsee, että laite on mennyt rikki. Tällöin ”onnitumistodennäköisyys” (so. todennäköisyys, että yksittäinen laite rikkoutuu 30 päivän kuluessa) voidaan lausua μ:n avulla eksponenttijakaumasta laskemalla.

Ratkaisu 3

Kestoikiä mitattaessa voi tulla vastaan tilanne, jossa koe joudutaan jostain syystä keskeyttämään ennen tulosten saamista, eli laitteiden hajoamista. Tällöin tiedetään vain laitteiden kestoiän olevan suurempaa kuin kokeeseen käytetty aika.

Halvempi koejäsjestely:
Yhden laitteen kestoikä päivinä noudattaa eksponenttijakaumaa parametrilla $\frac{1}{\mu}$, jossa $\mu = \frac{1}{\lambda}$ on laitteiden kestoiän odotusarvo. Jos muuttujalla X_i kuvataan i:nnen laitteen kestoikää, niin $X_1, \ldots, X_n \sim \text{Exp}\left(\frac{1}{\mu}\right) \perp\!\!\!\perp$.

Selvitetään ensin X_i:n kertymäfunktio:

$$\begin{aligned}F\left(x\,;\frac{1}{\mu}\right) &= \int_{-\infty}^{x} f\left(t;\frac{1}{\mu}\right)dt \\ &= \int_{0}^{x} \frac{1}{\mu}e^{-\frac{t}{\mu}}dt \\ &= \Big/_{0}^{x} -e^{-\frac{t}{\mu}} \\ &= 1-e^{-\frac{x}{\mu}},\ \text{ kun } x>0\end{aligned}$$

Kertymäfunkiton avulla voidaan selvittää todennäköisyys yhden laitteen hajoamiselle 30 päivän aikana.

$$P(X_i \leq 30) = F\left(30\,;\frac{1}{\mu}\right) = 1-e^{-\frac{30}{\mu}}$$

Kuvataan satunnaismuuttujalla Y 30 päivän aikana hajonneiden laitteiden määrää. Y noudattaa binomijakaumaa, jonka parametrit ovat laitteiden määrä (n) ja yhden laitteen hajoamistodennäköisyys. Siis $Y \sim \text{Bin}\left(n\,,\,1-e^{-\frac{30}{\mu}}\right)$ ja sen pistetodennäköisyysfunktio on

$$f_Y(y\,;\mu) = \binom{n}{y}\left(1-e^{-\frac{30}{\mu}}\right)^{y}\left(e^{-\frac{30}{\mu}}\right)^{n-y},\ \text{ kun } y=0,1,\ldots,n$$

4. Palauta mieleen JTP-kurssilta (nyk. Tilastollinen päättely I) tai opiskele monisteen jaksosta 2.1 uskottavuusfunktion määritelmä. Muodosta uskottavuusfunktiot tehtävän 2 a- ja b-kohdan tilanteessa.

Ratkaisu 4

Muodostetaan ensin uskottavuusfunktio tilanteelle, jossa on havaittu $Y_1 = 4$. Luentomateriaalin perusteella tiedetään, että $L(\theta) = P_\theta(Y_i = y)$, Siis

$$L(\theta) = P_\theta(Y_1 = 4) = \begin{cases} 0.1, \text{ kun } \ \theta = 1 \\ 0.4, \text{ kun } \ \theta = 2 \end{cases}$$

Muodostetaan uskottavuusfunktio tilanteelle, jossa on havaittu $Y_1 = 4$ ja $Y_2 = 1$. Nyt (jos $Y_1 \perp\!\!\!\perp Y_2$)

$$L(\theta) = P_\theta(Y_1 = 4, Y_2 = 1) = P_\theta(Y_1 = 4)P_\theta(Y_2 = 1) = \begin{cases} 0.02, \text{ kun } \ \theta = 1 \\ 0, \text{ kun } \ \theta = 2 \end{cases}$$

Huom! Jos $Y_1 \perp\!\!\!\perp Y_2$ *ei päde*, niin $L(2) = P_{\theta=2}(Y_1 = 4, Y_2 = 1) = 0$, koska $P_{\theta=2}(Y_2 = 1) = 0$, mutta $L(1)$:n arvoa ei voi laskea.