Tilastollinen päättely, syksy 2013 - kevät 2014

Harjoitus 1

Esimerkkiratkaisut

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

   (a) $Y = X_1 + \cdots + X_n$, kun $X_1, \cdots, X_n \sim B(\theta) \quad \perp\!\!\!\perp$ (otos Bernoulli-jakaumasta)

   (b) $Y = U + V$, kun $U \sim N(\mu_1, \sigma_1^2), V \sim N(\mu_2, \sigma_2^2)$ ja $U \perp\!\!\!\perp V$.

   (c) $Y = U^2$, kun $U \sim N(0, 1)$.

   (d) $Y = aX + b$, kun $X \sim Tas(0, 1)$ (välin (0, 1) tasajakauma) ja $a > 0, b \in \mathbb{R}$.

   (e) $Y = X_1 + \cdots + X_n$, kun $X_1, \cdots, X_n \sim EXP(\lambda) \perp\!\!\!\perp$.

   (f) $Y = X_1 + \cdots + X_n$, kun $X_1 \sim P(\mu_1), \cdots, X_n \sim P(\mu_n)$ ja $X_1, \cdots, X_n$ ovat riippumattomia.

   Ilmoita myös Y:n odotusarvo ja varianssi kussakin tapauksessa.

   [Perusteluja ei tarvitse esittää. Käytä tarvittaessa todennäköisyyslaskennan materiaalia ja taulukkokirjoja apuna.]

   Ratk.

   (a) Kun $X_1, \cdots, X_n$ ovat riippumattomia $B(\theta)$-jakautuneita satunnaismuuttujia eli Bernoulli-jakautuneita satunnaismuuttujia parametrilla θ, niin niiden summa $Y = X_1 + \cdots + X_n$ noudattaa $Bin(n, \theta)$-jakaumaa eli Binomijakaumaa parametreilla n ja θ. Koska $Y \sim Bin(n, \theta)$, niin $\mathrm{E}(Y) = n\theta$ ja $\mathrm{Var}(Y) = n\theta(1 - \theta)$. Jakauman $Bin(n, \theta)$ ptf on

   $$f_Y(y; n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad y = 0, 1, \ldots, n, \ \theta \in [0, 1].$$

   (b) U noudattaa normaalijakaumaa odotusarvolla μ_1 ja varianssilla σ_1^2 ja V noudattaa normaalijakaumaa odotusarvolla μ_2 ja varianssilla σ_2^2. Satunnaismuuttujat U ja V oletetaan lisäksi riippumattomiksi.

   Kahden normaalijakautuneen satunnaismuuttujan summa noudattaa myös normaalijakaumaa. Koska

   $$\mathrm{E}(Y) = \mathrm{E}(U + V) = \mathrm{E}(U) + \mathrm{E}(V) = \mu_1 + \mu_2$$

   ja

   $$\mathrm{Var}(Y) = \mathrm{Var}(U + V) \overset{\perp\!\!\!\perp}{=} \mathrm{Var}(U) + \mathrm{Var}(V) = \sigma_1^2 + \sigma_2^2,$$

niin $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ eli Y noudattaa normaalijakaumaa odotusarvolla $\mu_1 + \mu_2$ ja varianssilla $\sigma_1^2 + \sigma_2^2$. Jakauman $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ tf on

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}, \quad y \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma^2 > 0,$$

jossa $\mu = \mu_1 + \mu_2$ ja $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

(c) Kun U noudattaa standardoitua normaalijakaumaa $N(0,1)$, niin $U^2 \sim \chi_1^2$ eli U^2 noudattaa khii-toiseen jakaumaa vapausasteella 1. Khii-toiseen jakauman odotusarvo on sama kuin vapausasteiden määrä ja varianssi on sama kuin kaksi kertaa vapausasteiden määrä, joten

$$\mathrm{E}(Y) = 1 \text{ ja } \mathrm{Var}(Y) = 2 \cdot 1 = 2.$$

Jakauman χ_n^2 tf on

$$f_Y(y; n) \quad = \quad \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \quad y > 0,\ n = 1, 2, \ldots,$$

joten jakauman χ_1^2 tiheysfunktioksi saadaan

$$\begin{aligned} f_Y(y) &= \frac{1}{2^{1/2}\Gamma(1/2)} y^{1/2-1} e^{-y/2} = \frac{1}{2^{1/2}\sqrt{\pi}} y^{-1/2} e^{-y/2} \\ &= \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, \quad y > 0. \end{aligned}$$

(d) Olkoon $Y \sim Tas(\alpha, \beta)$ eli Y noudattaa tasaista jakaumaa välillä (α, β). Tällöin satunnaismuuttujan Y jakauman tiheysfunktio on

$$f_Y(y; \alpha, \beta) = \begin{cases} \frac{1}{\beta-\alpha}, & \text{jos } y \in (\alpha, \beta), \\ 0, & \text{jos } y \leq \alpha \text{ tai } y \geq \beta. \end{cases}$$

ja kertymäfunktio on

$$F_Y(y; \alpha, \beta) = \begin{cases} 0, & \text{jos } y \leq \alpha, \\ \frac{y-\alpha}{\beta-\alpha}, & \text{jos } y \in (\alpha, \beta), \\ 1, & \text{jos } y \geq \beta. \end{cases}$$

Jos $X \sim Tas(0,1)$, niin $Y = aX + b \sim Tas(b, a+b)$, sillä

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right) = \frac{y-b}{a} = \frac{y-b}{(a+b)-b} \end{aligned}$$

on jakauman $Tas(b, a+b)$ kertymäfunktio. Satunnaismuuttuja $Y = aX + b$ noudattaa siis tasaista jakaumaa välillä $(b, a+b)$. Tasaisen jakauman $Tas(\alpha, \beta)$ odotusarvo on $(\alpha+\beta)/2$ ja varianssi $(\beta-\alpha)^2/12$, joten

$$\mathrm{E}(Y) = \frac{b + (a+b)}{2} = \frac{a+2b}{2} \quad \text{ja} \quad \mathrm{Var}(Y) = \frac{((a+b)-b)^2}{12} = \frac{a^2}{12}.$$

(e) Jos $X_i \sim Exp(\lambda)$ eli X_i noudattaa eksponenttijakaumaa parametrilla λ, niin silloin myös $X_i \sim G(1, \lambda)$ eli X_i noudattaa gamma-jakaumaa parametrein 1 ja λ. Eksponenttijakauma on siis gamma-jakauman erikoistapaus. Olkoon $X_i \sim G(\kappa_i, \lambda)$, $i = 1, \ldots, n$, riippumattomia satunnaismuuttujia. Tällöin $\sum_{i=1}^n X_i \sim G(\sum_{i=1}^n \kappa_i, \lambda)$. Jos $\kappa_i = 1$, $i = 1, \ldots, n$, niin

$$Y = X_1 + \cdots + X_n \sim G\left(\sum_{i=1}^n 1, \lambda\right) = G(n, \lambda).$$

Satunnaismuuttujan Y jakauman $G(n, \lambda)$ odotusarvo ja varianssi ovat seuraavat:

$$\mathrm{E}(Y) = \frac{n}{\lambda} \text{ ja } \mathrm{Var}(Y) = \frac{n}{\lambda^2}.$$

Jakauman $G(n, \lambda)$ tiheysfunktio on

$$f_Y(y; n, \lambda) \quad = \quad \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}, \quad y > 0, \quad n > 0, \quad \lambda > 0.$$

(f) Olkoon satunnaismuuttujat $X_i \sim P(\mu_i)$, $i = 1, \ldots, n$, riippumattomia. Tällöin $\sum_{i=1}^n X_i \sim P\left(\sum_{i=1}^n \mu_i\right)$. Jakauman $P\left(\sum_{i=1}^n \mu_i\right)$ odotusarvo ja varianssi ovat seuraavat:

$$\mathrm{E}(Y) = \sum_{i=1}^n \mu_i \text{ ja } \mathrm{Var}(Y) = \sum_{i=1}^n \mu_i.$$

Jakauman $P(\sum_{i=1}^n \mu_i)$ ptf on

$$f_Y(y; \mu) \quad = \quad e^{-\mu} \frac{\mu^y}{y!}, \quad y = 0, 1, 2, \ldots, \quad \mu \geq 0,$$

jossa $\mu = \sum_{i=1}^n \mu_i$.

2. Erään sähkölaitteen eliniän (päivinä) oletetaan noudattavan eksponenttijakaumaa. Olkoon μ keskimääräinen elinikä (eliniän odotusarvo). Poimitaan n laitetta yksinkertaisella satunnaisotannalla ja mitataan niiden eliniät $y_1, \ldots, y_n$, jolloin vastaavat satunnaismuuttujat ovat $Y_1, \ldots, Y_n \sim EXP(1/\mu)$. Muodosta syntyvän tilastollisen mallin eli satunnaisvektorin $Y = (Y_1, \ldots, Y_n)$ ytf:n lauseke. Ilmoita myös mallin uskottavuusfunktio. Pane merkille, että se riippuu aineistosta vain yhden luvun kautta.

Ratk. Koska havainnot ovat riippumattomia, niin yhteistiheysfunktioksi saadaan

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu) &= \prod_{i=1}^n f_{Y_i}(y_i; \mu) = \prod_{i=1}^n \frac{1}{\mu} \exp\left(-\frac{y_i}{\mu}\right) = \frac{1}{\mu^n} \exp\left(-\frac{1}{\mu} \sum_{i=1}^n y_i\right) \\ &= \frac{1}{\mu^n} \exp\left(-\frac{n\bar{y}}{\mu}\right) \end{aligned}$$

Uskottavuusfunktioksi saadaan

$$L(\mu; \boldsymbol{y}) \quad = \quad f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu) = \frac{1}{\mu^n} \exp\left(-\frac{n\bar{y}}{\mu}\right)$$

Huomataan, että uskottavuusfunktio riippuu aineistosta ainoastaan havaintojen keskiarvon $\bar{y}$ (tai summan $\sum y_i$) kautta.

3. Tilanne on sama kuin edellisessä tehtävässä. Nyt ei kuitenkaan ole mahdollista mitata kunkin laitteen elinikää erikseen, vaan käytetään halvempaa koejärjestelyä, jossa n laitetta pannaan yhtä aikaa käyntiin ja kuukauden (30 päivää) kuluttua tullaan tarkastamaan, mitkä niistä ovat ehjiä ja mitkä rikki. Muodosta tätä asetelmaa kuvaava tilastollinen malli parametrille μ.

   [Vihje. Ajattele koeasetelmaa n-kertaisena toistokokeena, jossa onnistuminen merkitsee, että laite on ehjä ja epäonnistuminen, että se on rikki. Mikä on onnistumistodennäköisyys μ:n avulla lausuttuna.]

   Ratk. Olkoon

$$Y_i = \begin{cases} 1, & \text{jos laite } i \text{ on ehjä 30 päivän kuluttua,} \\ 0, & \text{jos laite } i \text{ on rikki 30 päivän kuluttua.} \end{cases}$$

Eksponenttijakaumaman kertymäfunktio on

$$F_X(x) = \int_0^x \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) dy = \Big|_0^x - \exp\left(-\frac{y}{\mu}\right) = 1 - \exp\left(-\frac{x}{\mu}\right),$$

joten

$$P(Y_i = 1) = P(X_i > 30) = 1 - P(X_i \leq 30) = 1 - F_{X_i}(30) = \exp\left(-\frac{30}{\mu}\right).$$

Riippumattomat satunnaismuuttujat Y_i noudattavat siis Bernoulli-jakaumaa parametrilla $\theta = \exp\left(-\frac{30}{\mu}\right)$. Satunnaisvektorin $Y = (Y_1, \ldots, Y_n)$ yptf on nyt

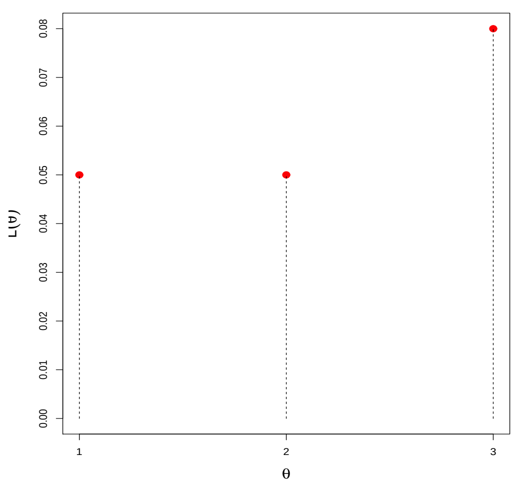
$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}; \mu) \quad &= \quad \prod_{i=1}^n f_{Y_i}(y_i; \mu) = \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \\ &= \quad \theta^{n\bar{y}}(1-\theta)^{n(1-\bar{y})} = \left[\exp\left(-\frac{30}{\mu}\right)\right]^{n\bar{y}} \left[1 - \exp\left(-\frac{30}{\mu}\right)\right]^{n(1-\bar{y})} \end{aligned}$$

4. Eräs leijona viettää yönsä jossakin kolmesta tilasta: se on koko yön joko hyvin aktiivinen ($\theta = 1$), kohtalaisen aktiivinen ($\theta = 2$) tai unelias ($\theta = 3$). Leijona syö yön aikana Y ihmistä. Satunnaismuuttujan Y pistetodennäköisyydet riippuvat leijonan tilasta ja käyvät ilmi oheisesta taulukosta.

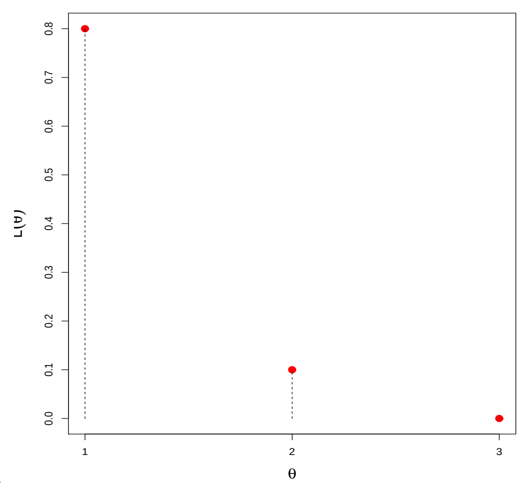
y	0	1	2	3	4
$f_Y(y;1)$	.00	.05	.05	.80	.10
$f_Y(y;2)$	.05	.05	.80	.10	.00
$f_Y(y;3)$	.90	.08	.02	.00	.00

Eräänä aamuna havaittiin, että yön aikana leijona oli syönyt a) $y = 1$, b) $y = 3$ ihmistä. Esitä graafisesti vastaavat uskottavuusfunktiot. Millaisia päätelmiä tekisit leijonan tilasta kyseisenä yönä? Kuinka luotettavina pitäisit päätelmiäsi?

Ratk. Jokaisella $y \in \{1, 2, 3, 4\}$ uskottavuusfunktio saadaan yhtälöstä $L(\theta; y) = f(y; \theta)$. Annetun taulukon arvoilla saadaan kuvat:



(a)



(b)

Kuvista voidaan päätellä, että SU-estimaatit ovat (a) $\hat{\theta} = 3$ ja (b) $\hat{\theta} = 1$. Kohdan (b) estimaatti vaikuttaa luotettavammalta kuin kohdan (a), sillä suhde $L(1;3)/L(2;3) = 8$ on huomattavasti suurempi kuin suhde $L(3;1)/L(2;1) = L(3;1)/L(1;1) = 8/5$.