Tilastollinen päättely I, kevät 2017
Harjoitus 1 – palautuspäivämäärä 21. 3.

Opastus: Tehtävät 1–4 johdattelevat tilastollisen päättelyn kysymyksenasetteluun ja liittyvät ensimmäisen viikon luentoihin. Tehtävää 5 varten opiskele jo etukäteen monisteen jaksoja 2.1, 4.1 ja 4.2. Tehtävä 6 on ”yleissivistystä”.

1. Diskreetin satunnaismuuttujan Y pistetodennäköisyydet $f(y;\theta)$ riippuvat parametrista θ, jolla on kaksi vaihtoehtoista arvoa: 1 tai 2. Pistetodennäköisyydet on taulukoitu alla. (Tavalliseen tapaan sovimme, että $f(y;\theta) = 0$ kaikilla niillä y:n arvoilla, joita ei ole mainittu.)

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

Varmista, että kumpikin funktioista $f(y;1)$ ja $f(y;2)$ kelpaa pistetodennäköisyysfunktioksi. Esitä kumpikin funktio graafisesti ja laske vielä kummankin jakauman odotusarvo.

2. Jatkoa edelliseen tehtävään. Emme tiedä, kumpi θ:n arvoista on oikea, joten suoritamme satunnaiskokeen, jonka tuloksena havaitaan Y:n arvo y.

a) Havainto on $y = 5$. Nyt voimme varmuudella päätellä oikean parametriarvon. Kumpi se on? Miksi?

b) Havainto on $y = 4$. Mitä nyt sanoisit θ:n arvosta? Kummalla θ:n arvolla tämä havainto on todennäköisempi?

3. Metrojunat kulkevat tasaisin θ minuutin väliajoin, jossa $\theta > 0$ on reaaliluku. Opiskelija menee laiturille sattumanvaraisesti katsomatta kelloa ja tuntematta aikataulua. Olkoon Y (minuuttia) se aika, jonka hän joutuu odottamaan junaa. Mitä jakaumaa satunnaismuuttuja Y noudattaa? Kerro jakauman nimi ja lausu sen tiheysfunktio sekä kertymäfunktio. Kiinnitä huomiota siihen, mikä on jakauman alusta eli missä joukossa esimerkiksi tiheysfunktio poikkeaa nollasta.

4. Jatkoa edelliseen tehtävään. Opiskelija on tullut maalta ja hän ei tiedä θ:n arvoa. Hän haluaa tehdä päätelmiä siitä tilastollisen päättelyn keinoin menemällä toistuvasti laiturille ja mittaamalla odotusaikansa. Oletamme, että eri odotuskerrat ovat toisistaan täysin riippumattomia (esim. eri päivinä sattumanvaraiseen aikaan toteutettuja).

a) Olkoon $y_1 = 4.5$ hänen odotusaikansa ensimmäisellä kerralla. Mitä tämän havainnon perusteella voi päätellä θ:sta?

b) Ensimmäiset viisi odotusaikaa $y_1, \ldots, y_5$ ovat 4.5, 1.1, 4.8, 0.8 ja 2.0. Mitä nyt voi päätellä θ:sta?

c) Opiskelija tekee kaikkiaan 50 odotusajan mittausta, ja suurin havaituista odotusajoista on 4.8. Mitä nyt voi päätellä θ:sta?

d) Opiskelija on kuullut väitettävän, että $\theta = 8$. Pohdi, miten hänen suhtautumisensa tämän väitteen (eli hypoteesin) todenperäisyyteen muuttuu a–c-tilanteiden myötä. Jos todella pätisi $\theta = 8$, kuinka todennäköistä olisi, että 50 riippumatonta odotusaikaa olisivat kaikki ≤ 4.8?

5. Tarkastellaan tehtävän 1 pistetodennäköisyysfunktiota (eli mallia) $f(y;\theta)$. Muodosta havaintoa y vastaava uskottavuusfunktio $L(\theta;y)$ (luettele sen arvot) ja esitä ko. funktio myös graafisesti, kun i) $y=5$, ii) $y=4$. Määritä kummassakin tapauksessa myös parametrin θ suurimman uskottavuuden estimaatti $\hat{\theta}$,

6. Käytämme tällä kurssilla usein merkintää $\prod_{i=1}^n x_i$ tarkoittamaan reaalilukujen $x_1,\dots,x_n$ tuloa, ts.

$$\prod_{i=1}^n x_i = x_1x_2\cdots x_n,$$

ja merkintä $\sum_{i=1}^n x_i$ tarkoittaa tietenkin summaa $x_1+\cdots+x_n$.

Osa seuraavista kaavoista on identiteettejä (tosia kaikilla mahdollisilla lukujen x_i, y_i ja a arvoilla) ja osa taas ei. Selvitä kustakin kaavasta, onko se identiteetti vai ei. Myönteisessä tapauksessa ei tarvitse esittää varsinaista todistusta. Korjaa epäidenttisten kaavojen *oikeaa* puolta niin, että saat identiteetin, mikäli se on helposti tehtävissä; vaikeammissa tapauksissa voit antaa vastaesimerkin.

$$\left(\prod_{i=1}^n x_i\right)\left(\prod_{i=1}^n y_i\right) = \prod_{i=1}^n x_iy_i \tag{1}$$

$$\prod_{i=1}^n x_i + \prod_{i=1}^n y_i = \prod_{i=1}^n (x_i+y_i) \tag{2}$$

$$\prod_{i=1}^n ax_i = a\prod_{i=1}^n x_i \tag{3}$$

$$\log\left(\prod_{i=1}^n x_i\right) = \sum_{i=1}^n \log x_i \quad (x_i>0 \text{ kaikilla } i) \tag{4}$$

$$\exp\left(\sum_{i=1}^n x_i\right) = \sum_{i=1}^n \exp x_i. \tag{5}$$

Tässä ja yleensäkin tällä kurssilla ”log” tarkoittaa luonnollista eli e-kantaista logaritmia (laskimissa usein ”ln”). Lisäksi $\exp x = e^x$.