Johdatus tilastolliseen päättelyyn, kevät 2015
Harjoitus 1 (17.–20. 3.)

Opastus: Alkupään tehtävät (1–4) liittyvät ensimmäisen viikon luennoilla käsiteltyihin aiheisiin. Loppupään tehtävät (5 ja 6) valmistavat seuraavan viikon aiheisiin ja edellyttävät luentomonisteen opiskelua etukäteen.

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.

[Vastaavaa satunnaisilmiötä tarkasteltiin JTN-kurssin harjoituksen 4 tehtävässä 11.]

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?

Opiskele luentomonisteen jaksoja 2.1, 4.1 ja 4.2 ja vastaa sitten seuraaviin tehtäviin.

5. Tarkastellaan tehtävän 1 pistetodennäköisyysfunktiota (eli mallia) $f(y;\theta)$. Esitä havaintoa y vastaava uskottavuusfunktio $L(\theta; y)$ graafisesti ja määritä parametrin θ suurimman uskottavuuden estimaatti $\hat{\theta}$, kun i) $y = 5$, ii) $y = 4$.

6. a) Muodosta sen tilastollisen mallin lauseke (yhteistiheysfunktio), joka kuvaa tehtävän 4c satunnaiskoetta, jossa aineisto $\mathbf{y}$ koostuu n:stä riippumattomasta odotusajan mittauksesta $y_1, \ldots, y_n$.

b) Ilmoita aineistoa $\mathbf{y}$ vastaava uskottavuusfunktio $L(\theta; \mathbf{y})$. Kiinnitä huomiota siihen, missä joukossa uskottavuusfunktio on määritelty ja missä se on nollasta poikkeava. (Havainnoista $y_1, \ldots, y_n$ suurin on tässä avainasemassa; sille käytetään usein merkintää $y_{(n)}$.) Piirrä kuva.

Huom: Harjoitustehtävien tekemisestä saa lisäpisteitä 1, 2, 3 tai 4, jos tekee vastaavasti ainakin 20, 40, 60 tai 80 % kaikista kurssin tehtävistä. Tämä edellyttää läsnäoloa harjoitusryhmässä ja valmiutta esittää oma ratkaisunsa.