**Johdatus tilastolliseen päättelyyn, kesä 2015**
**Harjoitus 1 (21.5.)**

**Opastus:** Tehtävät liittyvät tilastollisen päättelyn perusasetelmaan ja tilastollisen mallin käsitteeseen (monisteen luku 2). Tehtävissä 5 ja 6 perehdytään myös uskottavuusfunktioon (jakso 4.1).

**1.** Erästä kylää piinaava leijona viettää yönsä jossakin kolmesta tilasta: se on koko yön joko unelias ($\theta = 1$), aktiivinen ($\theta = 2$) tai ärtynyt ($\theta = 3$). Leijona syö yön aikana $Y$ ihmistä. Satunnaismuuttujan $Y$ pistetodennäköisyydet $f(y;\theta) = P(Y = y)$ riippuvat leijonan tilasta ja käyvät ilmi taulukosta alla.

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

Varmista, että funktiot $f(y;1)$, $f(y;2)$ ja $f(y;3)$ todella kelpaavat pistetodennäköisyysfunktioiksi. Esitä funktiot graafisesti ja laske kunkin jakauman odotusarvo.

**2.** Jatkoa edelliseen tehtävään. Ei ole mahdollista suoraan selvittää, missä tilassa leijona on yöllä ollut. Aamulla kuitenkin havaitaan helposti, kuinka monta ihmistä leijona on yön aikana syönyt. Olkoon syötyjen ihmisten lukumäärä $y$.

a) Havainto on $y = 4$. Nyt voimme varmuudella päätellä leijonan tilan. Mikä se on? Miksi?
b) Havainto on $y = 3$. Mitä nyt voidaan sanoa leijonan tilasta? Miksi?

**3.** Metrojunat kulkevat tasaisin $\theta$ 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 kevään 2015 JTN-kurssin harjoituksen 4 tehtävässä 11.]

**4.** Jatkoa edelliseen tehtävään. Opiskelija on tullut maalta ja hän ei tiedä $\theta$: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 $y_1 = 4.8$ hänen odotusaikansa ensimmäisellä kerralla. Mitä tämän havainnon perusteella voi päätellä $\theta$:sta?

b) Ensimmäiset viisi odotusaikaa $y_1, \ldots, y_5$ ovat 4.8, 1.1, 4.2, 0.8 ja 2.0. Mitä nyt voi päätellä $\theta$:sta?

c) Opiskelija tekee kaikkiaan 50 odotusajan mittausta, ja suurin havaituista odotusajoista on 4.8. Mitä nyt voi päätellä $\theta$: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 $\leq 4.8$?

**5.** Tarkastellaan tehtävien 1 ja 2 pistetodennäköisyysfunktiota (eli tilastollista mallia) $f(y;\theta)$. Esitä havaintoa $y$ vastaava uskottavuusfunktio $L(\theta;y)$ graafisesti ja määritä parametrin $\theta$ suurimman uskottavuuden estimaatti $\hat{\theta}$, kun a) $y = 4$, b) $y = 3$.

**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.