Johdatus tilastolliseen päättelyyn, kevät 2014
Harjoitus 2 (24.–28. 3.)

1. Palataan edellisten harjoitusten tehtävissä 1 ja 2 tarkasteltuun malliin $f(y;\theta)$, jossa aineistona oli yksi havainto y ja pistetodennäköisyydet olivat seuraavat:

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

   Esitä havaintoa y vastaava uskottavuusfunktio $L(\theta;y)$ graafisesti ja määritä suurimman uskottavuuden estimaatti $\hat{\theta}$, kun i) $y=5$, ii) $y=4$.

2. Jatkoa edelliseen tehtävään. Yhden havainnon sijasta tehdäänkin ko. jakaumasta kaksi riippumatonta havaintoa, satunnaismuuttujina siis $Y_1 \perp\!\!\!\perp Y_2$, joilla kummallakin on yo. pistetodennäköisyysfunktio.

   Oletetaan, että havainnot ovat $y_1 = 4$ ja $y_2 = 3$. Laske vastaavan uskottavuusfunktion $L(\theta;y_1,y_2)$ arvot ja määritä suurimman uskottavuuden estimaatti $\hat{\theta}$. Miten kommentoisit estimointiin liittyvää epävarmuutta?

3. Palataan edellisten harjoitusten tehtävään 4, jossa havaintoja vastaavat muuttujat $Y_1,\ldots,Y_n$ olivat satunnaisotos jakaumasta $\text{Tas}(0,\theta)$; lyhyesti kirjoitettuna $Y_1,\ldots,Y_n \sim \text{Tas}(0,\theta) \perp\!\!\!\perp$, jossa $\theta > 0$ on parametri.

   a) Palauta mieleen tilastollisen mallin lauseke eli yhteistiheysfunktio ja ilmoita sitten aineistoa $\boldsymbol{y} = (y_1,\ldots,y_n)$ vastaava uskottavuusfunktio $L(\theta;\boldsymbol{y})$. Kiinnitä huomiota siihen, missä 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.
   b) Määritä parametrin θ suurimman uskottavuuden estimaatti $\hat{\theta}$. (Huomaat törmääväsi ongelmaan, mikäli käytit $\text{Tas}(0,\theta)$-jakauman tiheysfunktiona funktiota, joka on $= 1/\theta$ vain avoimella välillä $]0,\theta[$. Kannattaakin valita tiheysfunktio, joka on $= 1/\theta$ vastaavalla suljetulla välillä. Se määrittelee saman jakauman.)
   c) Mitkä ovat $\hat{\theta}$:n arvot edellisten harjoitusten tehtävän 4 kohdissa a–c?

4. Palautetaan todennäköisyyslaskennasta mieleen, että Poissonin jakauman $\text{Poisson}(\lambda)$ pistetodennäköisyysfunktio on $g(x;\lambda) = e^{-\lambda}\lambda^x/x!$, kun $x = 0,1,2,\ldots$, jossa $\lambda > 0$. Tilastollinen malli on satunnaisotos tästä jakaumasta eli $Y_1,\ldots,Y_n \sim \text{Poisson}(\lambda) \perp\!\!\!\perp$.

   Muodosta tilastollisen mallin ja uskottavuusfunktion lausekkeet ja osoita uskottavuusfunktion logaritmia tutkimalla, että $\hat{\lambda} = \overline{y} = (y_1 + \cdots + y_n)/n$.

5. Jatkoa edelliseen tehtävään. Tutkitaan vastaavaa satunnaismuuttujaa eli *estimaattoria* $\hat{\lambda} = \overline{Y} = (Y_1 + \cdots + Y_n)/n$ (merkiten sitä siis samalla symbolilla kuin estimaattia).

   a) Totea, että $\hat{\lambda}$ on *harhaton* eli $E(\hat{\lambda}) = \lambda$. Mitä tämä tulkinnallisesti merkitsee ”toistetun aineistonkeruun” ajatuksen näkökulmasta?
   b) Laske varianssi $\text{var}(\hat{\lambda})$ ja totea, että se $\to 0$, kun havaintojen lukumäärä $n \to \infty$. Mitä tämä tulos voisi merkitä?

   Huom. Tämän tehtävän aihepiireistä puhutaan muistiinpanojen luvussa 3 (ks. erityisesti jaksot 3.4 ja 3.5). Emme kuitenkaan syvenny siihen tällä kurssilla laajemmin.