Tilastollinen päättely, syksy 2014 – kevät 2015
Harjoitus 1 (4. ja 6. 11. 2014)

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

   a) $Y = X_1 + \cdots + X_n$, kun $X_1, \ldots, X_n \sim B(\theta) \perp\!\!\!\perp$ (otos Bernoulli-jakaumasta)
   b) $Y = U + V$, kun $U \sim N(1, 2)$, $V \sim N(3, 4)$ ja $U \perp\!\!\!\perp V$
   c) $Y = U^2$, kun $U \sim N(0, 1)$
   d) $Y = aX + b$, kun $X \sim Tas(0, 1)$ ja $a > 0$, $b \in \mathbb{R}$
   e) $Y = X_1 + \cdots + X_n$, kun $X_i \sim P(\mu_i)$ ja $X_1, \ldots, X_n \perp\!\!\!\perp$.

   Ilmoita myös Y:n odotusarvo ja varianssi kussakin tapauksessa. Perusteluja ei vaadita.

   Käytä tarvittaessa todennäköisyyslaskennan materiaalia ja taulukkokirjoja apuna. Huomaa myös, että tällä kurssilla eri jakaumista käytettävät tunnukset eroavat todennäköisyyslaskennan kurssilla käytetyistä; ks. liitettä muistiinpanojen sivulla 88.

2. Satunnaismuuttuja Y_1 on diskreetti, arvojoukkonaan $\{1, 2, 3, 4, 5\}$. Sille on kaksi vaihtoehtoista pistetodennäköisyysfunktiota, jotka on taulukoitu alla.

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

   Emme tiedä, kumpaa näistä sattuma käyttää arpoessaan Y_1:n arvon.

   Tulkitse tämä parametrisena tilastollisena mallina, jonka parametria merkitään θ:lla. Mikä on parametriavaruus? Jos havaitaan $Y_1 = 4$, mitä voit päätellä θ:sta? Kuinka varmasti?

   Kyseisestä jakaumasta tehdään myös toinen havainto Y_2, jonka arvoksi saadaan $Y_2 = 1$. Mitä nyt voit päätellä θ:sta?

3. Opiskele luentomonisteen esimerkki 1.2.2 (kestoiät). Millainen käytännön ongelma saattaa liittyä tällaisen satunnaiskokeen tekemiseen?

   Päädytään käyttämään halvempaa koejärjestelyä, jossa n laitetta pannaan yhtä aikaa käyntiin ja 30 päivän kuluttua tullaan tarkastamaan, mitkä yksilöt (tai kuinka monta yksilöä) ovat menneet rikki. Muodosta tätä koeasetelmaa kuvaava tilastollinen malli parametrille μ (kestoikien odotusarvo).

   Apu. Ajattele, että kyseessä on n-kertainen toistokoe (binomikoe), jossa ”onnistuminen” merkitsee, että laite on mennyt rikki. Tällöin ”onnitumistodennäköisyys” (so. todennäköisyys, että yksittäinen laite rikkoutuu 30 päivän kuluessa) voidaan lausua μ:n avulla eksponenttijakaumasta laskemalla.

KÄÄNNÄ!

4. Suomenkielisen Wikipedian artikkelissa *Tilastollinen malli* lukee:

   Tilastollinen malli on pyrkimys yleistää tietyssä satunnaisotoksessa tai satunnaisesti valitussa osapopulaatiossa havaittu tapahtuma koskemaan koko populaatiota. Tilastollinen malli esittää tämän satunnaismuuttujien ja niihin liittyvien todennäköisyysjakaumien suhteen matemaattisesti erilaisten yhtälöiden avulla. Tyypillisesti sillä kuvataan, kuinka yksi tai useampi satunnaismuuttuja (selittävä/t muuttujat) selittää tarkasteltavan satunnaismuuttujan (selitettävä muuttuja) vaihtelua. Tilastollinen malli ei määräydy deterministisesti vaan se sisältää mallin parametreihin liittyvää satunnaisvaihtelua.

   Tämä ei aivan vastaa sitä tilastollisen mallin määrittelyä, joka kurssillamme on esitetty. Mitä ongelmia tai puutteita näet?

   Teksti jatkuu:

   Matemaattisesti tilastollinen malli voidaan esittää parina $(\mathcal{Y}, \mathcal{P})$*, jossa* $\mathcal{Y}$ *on mahdollisten havaintojen joukko ja* $\mathcal{P}$ *on* $\mathcal{Y}$*:hyn liittyvien todennäköisyysjakaumien joukko. Tilastollisessa analyysissa oletetaan, että havaittu aineisto on generoitunut tietyistä joukon* $\mathcal{P}$ *jakaumista. Tilastollinen malli mahdollistaa tilastollisen päättelyn, jonka avulla voidaan tehdään päätelmiä mallin hyvyydestä kuvaamaan tiettyä satunnaisilmiötä.*

   Tarkastellaan vaikkapa esimerkkiä 1.2.1 tässä valossa. Mikä on $\mathcal{Y}$? Entä $\mathcal{P}$?

   Tämä jälkimmäinen katkelmakaan ei taida olla täysin korrekti. Mitä ongelmia siinä on?

Lisäpisteitä saa tehtävien ratkaisemisesta 1, 2, 3, 4, 5, 6 tai 7, jos ratkaisee vastaavasti 20, 30, 40, 50, 60, 70 tai 80 prosenttia annetuista tehtävistä.