Johdatus tilastolliseen päättelyyn, 1. harjoitus (19.–23.3.2012)

1. Parametrilla θ on nyt kaksi vaihtoehtoista arvoa: 1 tai 2. Diskreetin satunnaismuuttujan Y pistetodennäköisyysfunktion $f(y;\theta)$ arvot on taulukoitu alla. (Tavalliseen tapaan sovimme, että $f(y;\theta) = 0$ kaikilla niillä arvoilla y, joita ei ole mainittu.)

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

Tarkista, että kumpikin funktioista $f(y;1)$ ja $f(y;2)$ on pistetodennäköisyysfunktio. Esitä kumpikin funktio graafisesti.

2. (Jatkoa edelliseen tehtävään.) Emme tiedä, kumpi parametrinarvo on oikea. Suoritamme satunnaiskokeen, jonka tuloksena havaitaan satunnaismuuttujan Y arvo y.

a) Tulos oli $y = 1$. Tässä tilanteessa voimme sanoa varmuudella, kumpi parametrinarvo on oikea. Miksi?

b) Tulos oli $y = 2$. Kummalla parametrinarvolla tämä havainto on todennäköisempi?

3. Olkoot $y_1, \dots, y_n$ reaalilukuja, ja olkoon $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ niiden aritmeettinen keskiarvo. Näytä, että mille tahansa $a \in \mathbb{R}$ pätee kaava

$$\sum_{i=1}^{n} (y_i - a)^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - a)^2.$$

Opastus: lähde liikkeelle esityksestä $(y_i - a)^2 = (y_i - \bar{y} + \bar{y} - a)^2$.

4. Havainnoista $y_1, \dots, y_n$ lasketaan otosvarianssi kaavalla

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} (y_i - a)^2 - n(\bar{y} - a)^2 \right),$$

jossa ensimmäinen kaavan muoto on otosvarianssin määritelmä, ja toinen seuraa edellisen tehtävän tuloksesta. Jos otosvarianssin joutuu laskemaan käsin (mikä on nykyaikana epärealistinen oletus muualla paitsi harjoitustehtävissä ja tenteissä), niin jälkimmäinen kaavan muoto saattaa olla hyödyllinen sopivalla arvon a valinnalla. Havainnot $y_i, i = 1, \dots, 9$ ovat

2011.6, 2016.9, 2010.4, 2010.9, 2013.4, 2010.1, 2011.5, 2011.3, 2012.3

a) Laske havaintojen keskiarvo $\bar{y}$.

b) Laske otosvarianssi valitsemalla (kaavan jälkimmäisessä muodossa) $a = 0$. (Vastaavaa laskentakaavaa suositellaan usein oppikirjoissa!)

c) Laske otosvarianssi valitsemalla (kaavan jälkimmäisessä muodossa) $a = 2012$.

d) Jos otosvarianssi laskettaisiin alkuperäisten lukujen y_i sijasta luvuista $x_i = y_i - 2012$, niin saataisiin sama arvo kuin b- ja c-kohdassa. Miksi?

5. Olkoon X satunnaismuuttuja ja a reaaliluku. Näytä, että

$$E(X - a)^2 = \operatorname{var} X + (\mu - a)^2.$$

Tässä $\mu = EX$ on X:n odotusarvo ja $\operatorname{var} X = E(X - \mu)^2$ on X:n varianssi.

Opastus: lähde liikkeelle esityksestä $(X - a)^2 = (X - \mu + \mu - a)^2$.