TILTA1 Matemaattinen tilastotiede
9. harjoitukset, 46. viikko 2007
Vastauksia

Huom. merkintä (a nCr b) tarkoittaa "a yli b:n" eli $\begin{pmatrix} a \\ b \end{pmatrix}$ = (a nCr b) = (a!)/(b!(a–b)!) = $a^{(b)}$/(b!)

9.1 E(X) = $\sum_{x\in S_x} x\cdot P(X=x) = \sum_{x<0} x\cdot P(X=x) + \sum_{x>0} x\cdot P(X=x) = \sum_{x>0} -x\cdot P(X=-x) + \sum_{x>0} x\cdot P(X=x)$

$$= \sum_{x>0} x\cdot[-P(X=-x)+P(X=x)] = \sum_{x>0} x\cdot[-P(X=x)+P(X=x)] = 0.$$

E(XY) = E(X^3) = $\sum_{x\in S_x} x^3\cdot P(X=x) = \sum_{x<0} x^3\cdot P(X=x) + \sum_{x>0} x^3\cdot P(X=x)$

$$= \sum_{x>0} -x^3\cdot P(X=-x) + \sum_{x>0} x^3\cdot P(X=x) = \sum_{x>0} x^3\cdot[-P(X=-x)+P(X=x)]$$

$$= \sum_{x>0} x^3\cdot[-P(X=x)+P(X=x)] = 0.$$

Siis Cov(X,Y) = E(XY) – E(X)E(Y) = 0.

Ovatko X ja Y riippumattomat?
P(X = x, Y = y) = P(X = x, $X^2 = x^2$) = P(X = x).
P(X = x)·P(Y = y) = P(X = x)·P($X^2 = x^2$) = P(X = x)·[P(X = x)+ P(X = –x)]
= P(X = x)·[P(X = x)+ P(X = x)]=2·$[P(X = x)]^2$.
Siis P(X=x,Y=y) ≠ P(X=x)P(Y=y), eli X ja Y eivät ole riippumattomat.
Huomaa, että kovarianssi (ja täten korrelaatio) voi olla 0, vaikka muuttujien välillä on riippuvuus!

9.2 X ~ Poi(3.5) ja Y|X ~ Bin(X,0.75).
E(Y) = E(E(Y|X)) = E(X·0.75) = 0.75·E(X) = 0.75·3.5 = 2.625.

9.3 S = { (1,2) , (1,3) , (2,2) , (2,3) , (3,2) , (3,3) }
k·[(1/1+1/2) + (1/1+1/3) + (1/2+1/2) + (1/2+1/3) + (1/3+1/2) + (1/3+1/3)] = 1.
Yllä olevasta yhtälöstä voidaan ratkaista k:n arvo. Siis k = 6/37.
P(X ≥ Y) = 6/37·[(1/2+1/2) + (1/3+1/2) + (1/3+1/3)] = 15/37.

9.4 Merk. X = 'naisten lkm komiteassa', Y = 'miesten lkm komiteassa'.

(a) $f(x,y) = P(X=x, Y=y) = \dfrac{(90\ nCr\ x)(30\ nCr\ y)}{(120\ nCr\ 5)}$, kun x, y =0,1,2,3,4,5, ja x + y = 5

(muulloin $f(x,y)=0$), eli kyseessä on (moniulotteinen) hypergeometrinen jakauma.

(b) $f_X(x) = \sum_{y=0}^{5} f(x,y) = f(x,5-x) = \dfrac{(90\ nCr\ x)(30\ nCr\ (5-x))}{(120\ nCr\ 5)}$, kun x = 0,1,2,3,4,5.

$f_Y(y) = \sum_{x=0}^{5} f(x,y) = f(5-y,y) = \dfrac{(90\ nCr\ (5-y))(30\ nCr\ y)}{(120\ nCr\ 5)}$, kun y = 0,1,2,3,4,5.

(molemmissa kyse hypergeometrisesta jakaumasta).

9.5 Nyt X ~ Bin(3, $1/6$) ja Y ~ Bin(3, $1/2$).

(a) E(X) = 3· $1/6$ = $1/2$ ja Var(Y) = 3· $1/2 \cdot (1-1/2) = 3/4$.

(b) $Cov(X,Y)$ = –3· $1/6 \cdot 1/2 = -1/4$. Siis $Cor(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} = \dfrac{-1/4}{\sqrt{5/12\cdot 3/4}} = \dfrac{-1}{\sqrt{5}} \approx -0.447.$

9.6 (a) Moniulotteinen hypergeometrinen jakauma:

$$f(x,y) = P(X = x, Y = y) = \frac{(10\ \ nCr\ x)(20\ \ nCr\ y)(10\ \ nCr\ (8-x-y))}{(40\ \ nCr\ 8)},\ \text{kun } 0 \le x + y \le 8$$

ja lisäksi x ja y ovat ei-negatiivisia kokonaislukuja.

(b) $f(y \mid x = 2) = \dfrac{f(2,y)}{f_X(2)}$. Selvitetään ensin X:n reunajakauma:

$$f_X(x) = \sum_{y \in S_Y} f(x,y) = \sum_{y=0}^{8-x} f(x,y) = \frac{(10\ \ nCr\ x)}{(40\ \ nCr\ 8)} \cdot \sum_{y=0}^{8-y} (20\ \ nCr\ y)(10\ \ nCr\ (8-x-y))$$

$$\overset{\Theta}{=} \frac{(10\ \ nCr\ x)}{(40\ \ nCr\ 8)} \cdot (30\ \ nCr\ (8-x)) = \frac{(10\ \ nCr\ x)(30\ \ nCr\ (8-x))}{(40\ \ nCr\ 8)}.$$

$^{\Theta}$) Hypergeometrinen identiteetti (ks. esimerkiksi s. 64 luentomateriaalissa).

Siis X ~ HGeo(8,40, $^{10}\!/_{40}$).

Edellisten tietojen perusteella saadaan

$$f(y \mid x = 2) = \frac{\dfrac{(10\ \ nCr\ 2)(20\ \ nCr\ y)(10\ \ nCr\ (6-y))}{(40\ \ nCr\ 8)}}{\dfrac{(10\ \ nCr\ 2)(30\ \ nCr\ 6)}{(40\ \ nCr\ 8)}} = \frac{(20\ \ nCr\ y)(10\ \ nCr\ (6-y))}{(30\ \ nCr\ 6)},$$

kun y = 0,1, …,6. Siis kyseessä hypergeometrinen jakauma, Y| X=2 ~ HGeo(6,30, $^{20}\!/_{30}$).

9.7 $f_X(x) = (1-\pi_1)\pi_1^{x-1}$, kun x = 1, 2, 3,… ja $f_Y(y) = (1-\pi_2)\pi_2^{y-1}$, kun y = 1, 2, 3,…

(a) $f(x,y) \overset{\Delta}{=} f_X(x) f_Y(y) = (1-\pi_1)\pi_1^{x-1}(1-\pi_2)\pi_2^{y-1}$, kun x = 1, 2, 3,… ja y = 1, 2, 3,…

(b) $P(Z > n) = P(X > n, Y > n) \overset{\Delta}{=} P(X > n)P(Y > n) \overset{\Delta\Delta}{=} \pi_1^n \pi_2^n = (\pi_1\pi_2)^n$.

Edellistä hyödyntäen saadaan

$$P(Z = n) = P(Z \le n) - P(Z \le n-1) = \ldots = P(Z > n-1) - P(Z > n) = (\pi_1\pi_2)^{n-1} - (\pi_1\pi_2)^n$$

$= (1-\pi_1\pi_2)(\pi_1\pi_2)^{n-1}$. Siis Z ~ Geo($1-\pi_1\pi_2$).

$^{\Delta}$) X ja Y riippumattomat

$^{\Delta\Delta}$) Vertaa edelliskerran harjoitus 8.1. ja luentomateriaalin esimerkki 4.6 (s. 111).

9.8 Merkitään X_1 = 'bakteerien lkm 1 litran näytteessä', X_1 ~ Poi(λ),
X_2 = 'bakteerien lkm 2 litran näytteessä', X_2 ~ Poi(2λ),
X_3 = 'bakteerien lkm 4 litran näytteessä', X_3 ~ Poi(4λ).

Tiedetään, että näytteet ovat riippumattomia.

(a) P('analysoitava kaikki kolme näytettä') = P(X_1 = 0, X_2 = 0) = P(X_1=0)P(X_2=0)

$$= \frac{e^{-\lambda}\lambda^0}{0!} \cdot \frac{e^{-2\lambda}(2\lambda)^0}{0!} = e^{-\lambda}e^{-2\lambda} = e^{-3\lambda}.$$

(b) Etsitään sellainen λ, että epäyhtälö $e^{-3\lambda} \le 0.01$ toteutuu.
Kun otetaan luonnollinen logaritmi puolittain saadaan $-3\lambda \le \ln 0.01$, jolloin λ:n suhteen saadaan ratkaistua $\lambda \ge (\ln 0.01)/(-3) \approx 1.535$.