TILTA1 Matemaattinen tilastotiede
6. harjoitukset, 42. viikko 2007
Vastauksia

Huom. merkintä (a nCr b) tarkoittaa ”a yli b:n” eli (a nCr b) = (a!)/(b!(a–b)!) = $a^{(b)}$/(b!)

6.1 (a) $Var(X+Y) = Var(X) + Var(Y) + 2\cdot Cov(X,Y) = 13$.
$Var(X–Y) = Var(X) + Var(Y) – 2\cdot Cov(X,Y) = 1$.
(b) $Var(X–X) = Var(X) + Var(X) – 2\cdot Cov(X,X) = 0$, koska $Cov(X,X) = Var(X)$.
$Cov(X,–X) = –Cov(X,X) = –Var(X) = –2$.
(c) $E(XY) = 4$, koska $Cov(X,Y) = E(XY) – E(X)E(Y) = 3$.
$Cov(2X–1, –3Y+9) = Cov(2X, –3Y) = 2\cdot(–3)\cdot Cov(X,Y) = –6\cdot Cov(X,Y) = –18$.

6.2 $E(X) = np = 6$ ja $Var(X) = np(1–p) = 2.4$. Saadaan ratkaistua $p = 0.6$ ja $n = 10$, joten $X \sim Bin(10,0.6)$.
Siis $P(X=5) = (10 \text{ nCr } 5)0.6^5 0.4^5 \approx 0.2007$.

6.3 $X_A \sim Bin(20,0.4)$, $X_B \sim Bin(20,0.3)$ ja $X_C \sim Bin(20,0.1)$. Nyt $X = X_A + X_B + X_C$.
$E(X) = E(X_A) + E(X_B) + E(X_C) = 20\cdot 0.4 + 20\cdot 0.3 + 20\cdot 0.1 = 16$.
$Var(X) =^* Var(X_A) + Var(X_B) + Var(X_C) = 20\cdot 0.4\cdot 0.6 + 20\cdot 0.3\cdot 0.7 + 20\cdot 0.1\cdot 0.9 = 10.8$ *) r:mattomuus!

6.4 Merk. X = toimivien komponenttien lkm 5:n systeemissä. Siis $X \sim Bin(5,p)$.
Merk. Y = toimivien komponenttien lkm 3:n systeemissä. Siis $Y \sim Bin(3,p)$.
5:n komponentin systeemi toimii, jos vähint. 3 komponenttia toimii; 3:n komponentin systeemi toimii, jos vähint. 2 komponenttia toimii.
5:n systeemi toimii todennäköisemmin, jos $P(X\geq 3) > P(Y\geq 2)$
$\Leftrightarrow (5 \text{ nCr } 3)p^3(1–p)^2 + (5 \text{ nCr } 4)p^4(1–p) + p^5 > (3 \text{ nCr } 2)p^2(1–p) + p^3$.
Epäyhtälö voidaan sieventää muotoon $3p^2(2p^3 – 5p^2 + 4p – 1) > 0$,
ja nollakohdat ratkaisemalla lauseke voidaan esittää: $3p^2(p–1)^2(2p–1) > 0$.
Oletetaan, että kelvolliset ratkaisut ovat välillä $0 < p < 1$.
Lauseketta tutkimalla havaitaan, että $1/2 < p < 1$ täyttää ehdot. Siis $1/2 < p < 1$.

6.5 (a) $X \sim HGeo(15,N,0.15)$
(b) (i) $P(X=0) = (150 \text{ nCr } 0)(850 \text{ nCr } 15)/(1000 \text{ nCr } 15) \approx 0.0857$.
(ii) $P(X\geq 2) = 1 – P(X\leq 1) = 1 – P(X=0) – P(X=1) \approx 0.6835$.
(c) (i) $P(X=0) = (750 \text{ nCr } 0)(4250 \text{ nCr } 15)/(5000 \text{ nCr } 15) \approx 0.0870$.
(ii) $P(X\geq 2) = 1 – P(X\leq 1) = 1 – P(X=0) – P(X=1) \approx 0.6818$.

R:llä:
```
# (b)
dhyper(0,0.15*1000,1000-0.15*1000,15) # kohta (i)
1-phyper(1,0.15*1000,1000-0.15*1000,15) # kohta (ii)
# (c)
dhyper(0,0.15*5000,5000-0.15*5000,15) # kohta (i)
1-phyper(1,0.15*5000,5000-0.15*5000,15) # kohta (ii)
```

6.6 ks. H6mtt07_6_R.txt (R-koodi). Kyseisessä tiedostossa on vertailtu todennäköisyysfunktion arvoja HGeo- ja Bin-jakaumilla, kun x = 0,1,2,…,15. Siinä ei varsinaisesti lasketa todennäköisyyksiä $P(X=0)$ ja $P(X\geq 2)$ kun N saa eri arvoja.

Jos lasketaan tehtävässä 6.5 kysytyt todennäköisyydet $P(X=0)$ ja $P(X\geq 2)$ saadaan:
Binomijakauma: $X_{Bin} \sim Bin(15,0.15)$, jolloin $P(X_{Bin} = 0) \approx 0.0874$ ja $P(X_{Bin} \geq 2) \approx 0.6814$.
Verrataan tilanteeseen, jossa $X_{HGeo} \sim HGeo(15,N,0.15)$:
(a) N=100. $P(X_{HGeo} = 0) \approx 0.0710$ ja $P(X_{HGeo} \geq 2) \approx 0.7040$.
(b) N=500. $P(X_{HGeo} = 0) \approx 0.0841$ ja $P(X_{HGeo} \geq 2) \approx 0.6857$.
(c) N=1000. Kuten 6.5 (b):ssä $P(X_{HGeo} = 0) \approx 0.0857$ ja $P(X_{HGeo} \geq 2) \approx 0.6835$. Binomijakauma on jo hyvin lähellä.

6.7 Huom. Koska S = $\{a_1, a_2, a_3\}$, niin $P(X = a_3) = 1 - P(X = a_1) - P(X = a_2)$ ja vastaavasti $P(Y = a_3) = 1 - P(Y = a_1) - P(Y = a_2)$. Käytetään näitä tietoja apuna odotusarvojen muodostuksessa.

$$E(X) = \sum_{i=1}^{3} a_i P(X = a_i) = \ldots = (a_1 - a_3)P(X = a_1) + (a_2 - a_3)P(X = a_2) + a_3 \text{ ja}$$

$$E(Y) = \sum_{i=1}^{3} a_i P(Y = a_i) = \ldots = (a_1 - a_3)P(Y = a_1) + (a_2 - a_3)P(Y = a_2) + a_3.$$

Oletuksen mukaan $E(X) = E(Y)$

$$\Leftrightarrow (a_1 - a_3)[P(X = a_1) - P(Y = a_1)] + (a_2 - a_3)[P(X = a_2) - P(Y = a_2)] = 0 \qquad (1)$$

Vastaavalla tavalla voidaan muodostaa $E(X^2)$ ja $E(Y^2)$, ja oletuksen mukaan

$$E(X^2) = E(Y^2)$$

$$\Leftrightarrow (a_1^2 - a_3^2)[P(X = a_1) - P(Y = a_1)] + (a_2^2 - a_3^2)[P(X = a_2) - P(Y = a_2)] = 0 \qquad (2)$$

Merk. $z_1 = P(X = a_1) - P(Y = a_1)$ ja $z_2 = P(X = a_2) - P(Y = a_2)$.

Tällöin yhtälöiden (1) ja (2) muodostama yhtälöpari voidaan esittää:

$$\begin{cases} (a_1 - a_3)z_1 + (a_2 - a_3)z_2 = 0 \\ (a_1^2 - a_3^2)z_1 + (a_2^2 - a_3^2)z_2 = 0 \end{cases}$$

Oletusten perusteella kaikki z_i:iden kertoimet eivät ole nollia ja yhtälöparin ratkaisuksi saadaan $z_1 = z_2 = 0$. Tällöin $P(X = a_1) = P(Y = a_1)$ ja $P(X = a_2) = P(Y = a_2)$.

Lisäksi tästä seuraa, että $P(X = a_3) = P(Y = a_3)$. Siis X ~ Y.

6.8 $f(2;30,p) = (30 \text{ nCr } 2)p^2(1-p)^{28} = 435p^2(1-p)^{28}$.

Siis f(2;30,p) maksimoituu, kun lauseke $p^2(1-p)^{28}$ maksimoituu.

Merk. $g(p) = p^2(1-p)^{28}$. Etsitään derivaatan nollakohdat ja tutkitaan, missä on maksimi.

$g'(p) = 2p(1-p)^{28} - 28p^2(1-p)^{27} = 0 \Leftrightarrow 2p(1-p)^{27}[(1-p) - 14p] = 0$.

Nyt $g'(p) = 0$, kun $p = 0$ tai $p = 1$ tai $p = 1/15$.

$g''(p) = \ldots = 2(1-p)^{28} - 112p(1-p)^{27} + 756p^2(1-p)^{26}$, ja $g''(1/15) < 0$, joten se on maksimi.

Lisäksi voidaan todeta, että $g(0) = 0$, $g(1) = 0$ ja $g(1/15) > 0$.

Siis P('otoksessa 2 pilaantunutta') on suurin silloin, kun p = 1/15.

Arvo 1/15 (=2/30) on p:n suurimman uskottavuuden estimaatti.