Laskuharjoitustehtävät 5 syksy 2019

1.-2. (**e**) Vastaukset stackissa.

3. (**p**) Satunnaismuuttujat X ja Y ovat riippumattomia ja noudattavat eksponenttijakaumaa $\mathrm{Exp}\,(1)$.

   a) Mikä on yhteisjakauman tiheysfunktio $f_{X,Y}(x,y)$?

   b) Muodostetaan uudet satunnaismuuttujat Z ja W lineaarisella muunnoksella

   $$\begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} X+Y \\ Y \end{pmatrix}.$$

   Määrää muuttujien Z ja W yhteisjakauman tiheysfunktio $h_{Z,W}(z,w)$.

   c) Määrää integroimalla satunnaismuuttujan Z reunajakauman tiheysfunktio $h_Z(z)$.

   d) Mikä on summan $Z = X+Y$ tiheysfunktio yleisesti, kun X ja Y ovat riippumattomia ja niiden tiheysfunktiot ovat $f_X(x)$ ja $f_Y(y)$? Ohje: toista kohdat a)-c) yleisessä tapauksessa.

<u>Ratkaisu</u>

a) Tiheysfunktiot ovat $f_X(x) = \mathrm{e}^{-x}\mathrm{u}(x)$ ja $f_Y(y) = \mathrm{e}^{-y}\mathrm{u}(y)$. Koska satunnaismuuttujat ovat riippumattomia, niin $f_{X,Y}(x,y) = f_X(x)f_Y(y) = \mathrm{e}^{-x-y}\mathrm{u}(x)\mathrm{u}(y)$.

b) Lineaarinen muunnos

$$\begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} \Leftrightarrow \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} Z \\ W \end{pmatrix},$$

ja Jacobin determinantti on $J = \frac{\partial(z,w)}{\partial(x,y)} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$. Muuttujien Z ja W yhteisjakauman tiheysfunktio on siis

$$h_{Z,W}(z,w) = f_{X,Y}(z-w,w)\frac{1}{|J|} = \mathrm{e}^{-(z-w)}\,\mathrm{e}^{-w}\mathrm{u}(z-w)\mathrm{u}(w) = \mathrm{e}^{-z}\mathrm{u}(z-w)\mathrm{u}(w).$$

c) Muuttujan Z reunajakauma

$$h_Z(z) = \int_{-\infty}^{\infty} h_{Z,W}(z,w)\mathrm{d}w = \begin{cases} \displaystyle\int_0^z \mathrm{e}^{-z}\mathrm{d}w = z\,\mathrm{e}^{-z}, & z \geq 0 \\ 0, & z < 0 \end{cases} = z\,\mathrm{e}^{-z}\mathrm{u}(z).$$

(Gamma$(2,1)$-jakauma)

d) Edellisen perusteella yleisesti $h_Z(z) = \displaystyle\int_{-\infty}^{\infty} f_X(z-w)f_Y(w)\mathrm{d}w = (f_X * f_Y)(z)$, kun X ja Y ovat riippumattomat.

4. (**p**) Satunnaismuuttujien X ja Y kovarianssimatriisi on

$$\begin{pmatrix} 3.7 & 0.9 \\ 0.9 & 1.3 \end{pmatrix}.$$

Muodosta X:stä ja Y:stä lineaarisella muunnoksella uudet muuttujat U ja V, jotka ovat korreloimattomia. Säilyykö yhteenlaskettu varianssi muunnoksessa samana?

Ratkaisu

Muodostetaan lineaarisella muunnoksella $\begin{pmatrix} U \\ V \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}$ korreloimattomat satunnaismuuttujat U ja V (eli $\mathrm{Cov}(U, V) = 0$)

$$C_{(U,V)} = AC_{(X,Y)}A^T = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}.$$

Matriisi A muodostetaan riveittäin $C_{(X,Y)}$:n ominaisvektoreista. Ominaisarvot λ saadaan yhtälöstä $\det(C_{(X,Y)} - \lambda I) = 0$

$$\begin{vmatrix} 3.7-\lambda & 0.9 \\ 0.9 & 1.3-\lambda \end{vmatrix} = (3.7-\lambda)(1.3-\lambda) - 0.81 = \lambda^2 - 5\lambda + 4 = 0 \;\Rightarrow\; \begin{cases} \lambda_1 = 1 \\ \lambda_2 = 4 \end{cases}.$$

Ominaisvektorit saadaan yhtälöstä $\left(C_{(X,Y)} - \lambda I\right)\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\underline{\lambda = 1:} \qquad \begin{pmatrix} 2.7 & 0.9 \\ 0.9 & 0.3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \;\Rightarrow\; b = -3a, \quad \text{valitaan } \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\underline{\lambda = 4:} \qquad \begin{pmatrix} -0.3 & 0.9 \\ 0.9 & -2.7 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \;\Rightarrow\; a = 3b, \quad \text{valitaan } \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Lineaarisen muunnoksen matriisi

$$A = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}.$$

Uudet korreloimattomat muuttujat

$$\begin{pmatrix} U \\ V \end{pmatrix} = A\begin{pmatrix} X \\ Y \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} X - 3Y \\ 3X + Y \end{pmatrix}.$$

Nyt: $C_{(U,V)} = AC_{(X,Y)}A^T = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ eli $\mathrm{Cov}(U, V) = 0$. Lisäksi $\sigma_X^2 + \sigma_Y^2 = 5 = \sigma_U^2 + \sigma_V^2$.

Huomio 1. Normitetuilla ominaisvektoreilla muunnoksesta seuraa $\sigma_U^2 = \lambda_1$, $\sigma_V^2 = \lambda_2$.

Huomio 2. Jos valitsit muunnosmatriisin riveiksi normittamattomat ominaisvektorit, kokonaisvarianssi ei pysy samana. Esim. muunnosmatriisilla $A = \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}$ saadaan $\sigma_U^2 + \sigma_V^2 = 50$.

5. (**n**) Satunnaismuuttujien X ja Y odotusarvot ovat $\mu_X = 2$ ja $\mu_Y = -1$ sekä varianssit $\sigma_X^2 = 1$ ja $\sigma_Y^2 = 16$. Korrelaatiokerroin on $\rho_{X,Y} = \frac{1}{4}$. Muodostetaan uudet muuttujat $U = X + 2Y$ ja $V = 2X - Y$. Määrää

   a) $E(U)$ ja $E(V)$
   b) Kovarianssimatriisit $C_{X,Y}$ ja $C_{U,V}$ sekä korrelaatiokerroin $\rho_{U,V}$
   c) $E(U^2)$, $E(V^2)$ ja $E(UV)$.

   Ratkaisu

   a)
$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$
$$\begin{pmatrix} E(U) \\ E(V) \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}$$

   b) Kovarianssi $\mathrm{Cov}(X,Y) = \sigma_X \sigma_Y \rho(X,Y) = 1 \cdot 4 \cdot \frac{1}{4} = 1$, joten
$$C_{(X,Y)} = \begin{pmatrix} 1 & 1 \\ 1 & 16 \end{pmatrix}.$$
   Joten
$$C_{(U,V)} = AC_{(X,Y)}A^\top = \begin{pmatrix} 69 & -27 \\ -27 & 16 \end{pmatrix},$$
   eli $D^2(U) = 69,\ D^2(V) = 16, \mathrm{Cov}_{U,V} = -27$ sekä $\rho_{U,V} = \frac{-27}{4\sqrt{69}} \approx -0.8126$.

   c) Koska $\mathrm{Var}(U) = E(U^2) - (E(U))^2$, niin $E(U^2) = \mathrm{Var}(U) + (E(U))^2 = 69 + 0 = 69$. Vastaavasti $E(V^2) = 16 + 5^2 = 41$. $E(UV) = \mathrm{Cov}(U,V) + E(U) \cdot E(V) = -27$.

6. (**n**) Kirjoita Matlabin komentorivillä `load fisheriris`. Tämä lataa työmuistin muuttujaan `meas` kolmen eri iirislajin, Iris setosan, Iris virginican ja Iris versicolorin verho- ja terälehtien leveyden ja pituuden mittausdataa neljässä sarakkeessa. Kunkin rivin lajitieto löytyy muuttujasta `species`. Määrää (otos)kovarianssimatriisi ja tutki, onko tulos "laillinen" kovarianssimatriisi, ts. onko matriisi symmetrinen ja positiivisesti semidefiniitti. Määrää lineaarisella muunnoksella uudet muuttujat, jotka ovat korreloimattomat. Laske alkuperäisten ja uusien muuttujien yhteenlasketut varianssit. Mitkä kaksi uusista muuttujista kuvaavat parhaiten mittaustulosten vaihtelun? Piirrä näiden muuttujien hajontakuvio kaksiulotteiseen koordinaatistoon, eri lajit eri värein. Voitko erottaa eri lajit toisistaan suoralla viivalla?

   Komentoja: `cov`, `mean`, `eig`, `trace`, `gscatter`. Lisätietoa komennolla `help cov` jne.

   Ratkaisu

```
>> load fisheriris
>> X=meas;
>> Cx=cov(X)
>> diag(Cx)
```

```
ans =
0.6857
0.1900
3.1163
0.5810

>> [Q,D]=eig(Cx);
>> A=Q';
>> Y=X*A';
>> Cy=A*Cx*A';
>> diag(Cy)

ans =
0.0238
0.0782
0.2427
4.2282

>> trace(Cx)

ans =
4.5730

>> trace(Cy)

ans =
4.5730

>> gscatter(Y(:,4),Y(:,3),species)
```

Olkoon mittaushavainnot satunnaismuuttujat $[X_1,\ X_2,\ X_3,\ X_4]$. Yhden kasvin mittaukset muodostavat havaintomatriisin X rivin.

Otoskovarianssimatriisi C_X saadaan Matlabin komennolla `cov`. Sen diagonaalilla on mittaustulosten varianssit, yhteensä 4.5730 (saadaan komennolla `trace`). Selvästi kyseessä on symmetrinen matriisi, ja koska ominaisarvot ovat positiiviset, se on positiivisesti definiitti.

Kysytty lineaarinen muunnos saadaan diagonalisoimalla kovarianssimatriisi, eli muodostamalla hajotelma $C_X = QDQ^T$, missä Q on ortogonaalinen (yksikköominaisvektorit sarakkeina) ja D diagonaalinen (ominaisarvot diagonaalilla). Ominaisvektorit $Q = [q_1\ q_2\ q_3\ q_4]$ ja ominaisarvot D lasketaan komennolla `eig`, ja muunnoksen matriisi A on Q:n transpoosi. Jokainen X:n rivi saadaan nyt muunnettua kaavalla $Y = XA^T = XQ$. Uusien muuttujien $Y = [Y_1\ Y_2\ Y_3\ Y_4]$ kovarianssimatriisi on tällöin $C_Y = AC_XA^T = Q^TC_XQ = D$ (tai komennolla `cov`).

Kun Q on ortogonaalinen, muuttujien varianssien summa pysyy samana. Suurin varianssi on muuttujilla Y_3 ja Y_4: $\sigma_{Y_3}^2 + \sigma_{Y_4}^2 = 4.4709$; ne siis selittävät vaihtelusta suurimman osan. Hajontakuvaa tutkimalla Iris setosan havainnot erottuuvatkin omaksi ryhmäkseen, sen sijaan Iris versicolorin ja Iris virginican ei saada selvää eroa.