Theoretische Multivariate Statistik

January 21, 2016

Exercise 1 (13.10.2015)

Problem 1

For the matrix $\mathcal{A}$, such that

$$\mathcal{A} = \begin{pmatrix} 1 & -2 & 2 \\ -2 & -2 & 4 \\ 2 & 4 & -2 \end{pmatrix}$$

compute the eigenvalue and eigenvector and perform the spectral decomposition.

Problem 2

Rewrite the following quadratic form in matrix form, i.e. $f(x_1, x_2, \ldots, x_n) = x^\top \mathcal{A} x$, where $x = (x_1, x_2, \ldots, x_n)^\top$.

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\ &\quad + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ &\quad + \cdots \\ &\quad + a_{n1}x_nx_1 + a_{n2}x_nx_1 + \cdots + a_{nn}x_n^2. \end{aligned}$$

Problem 3

Rewrite the quadratic form $f(x_1, x_2, \ldots, x_n) = 2x_1^2 + 5x_2^2 + 5x_3^2 + 4x_1x_2 - 4x_1x_3 - 8x_2x_3$ to the quadratic form with a diagonal matrix by using spectral decomposition and Gram–Schmidt orthonormalization.

Problem 4

For the regression $Y = \mathcal{X}\beta + \varepsilon$, give the estimator of β based on the least square estimation.

Exercise 2 (20.10.2015)

Problem 5

For the linear regression $Y = \mathcal{X}\beta + \varepsilon$ the unconstrained ordinary least squares estimator is $\hat{\beta}_{OLS} = (\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top Y$. Now the linear constraint of coefficients $\mathcal{A}\beta = a$ is given, where $\mathcal{A}_{q\times p}$, $(q \le p)$ is the matrix with rank q and a is a $(q \times 1)$ vector. Show that

$$\hat{\beta} = \hat{\beta}_{OLS} - (\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{A}^\top \left\{\mathcal{A}(\mathcal{X}^\top\mathcal{X})\mathcal{A}^\top\right\}^{-1}(\mathcal{A}\hat{\beta}_{OLS} - a).$$

Problem 6

Show the variance decomposition

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

And show that the coefficient of determination,

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

is the square of the simple correlation between the random vector X and Y.

Problem 7

Show that $\operatorname{rank}(\mathcal{H}) = \operatorname{tr}(\mathcal{H}) = n - 1$, where $\mathcal{H} = \mathcal{I}_n - n^{-1}1_n1_n^\top$ is termed as the *centering matrix*.

Problem 8

Define $\mathcal{X}_* = \mathcal{H}\mathcal{X}\mathcal{D}^{-1/2}$, where $\mathcal{X}$ is a $(n \times p)$ matrix, $\mathcal{H}$ is the centering matrix, and $\mathcal{D}^{-1/2} = diag(s_{11}^{-1/2}, \dots, s_{pp}^{-1/2})$. Show that $\mathcal{X}_*$ is the standardized data matrix, i.e., $\bar{x}_* = 0_p$ and $\mathcal{S}_{\mathcal{X}_*} = \mathcal{R}_x$, the correlation matrix of $\mathcal{X}$.

Exercise 3 (27.10.2015)

Problem 9

Calculate the mean and the variance of the estimator $\hat{\beta} = (\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top Y$ in a linear model $Y = \mathcal{X}\beta + \epsilon,\ \mathrm{E}\epsilon = 0_n,\ \mathrm{Var}(\epsilon) = \sigma^2\mathcal{I}_n$.

Problem 10

Compute the conditional moments $\mathrm{E}(X_2|X_1)$ and $\mathrm{E}(X_1|X_2)$ for the two-dimensional pdf,

$$f(x_1, x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2 & 0 \leq x_1, x_2 \leq 1, \\ 0 & \textit{otherwise.} \end{cases}$$

Problem 11

Show that the following function is a pdf,

$$f(y) = \begin{cases} \frac{1}{2}y_1 - \frac{1}{4}y_2 & 0 \leq y_1 \leq 1,\ |y_2| \leq 1 - |1 - y_1|, \\ 0 & \textit{otherwise.} \end{cases}$$

Problem 12

Show that $\mathrm{Var}(X_2) = \mathrm{E}\{\mathrm{Var}(X_2|X_1)\} + \mathrm{Var}\{\mathrm{E}(X_2|X_1)\}$.

Exercise 4 (05.11.2015)

Problem 13

A company decides to compare the effect of three marketing strategies

1. advertisement in local newspaper,
2. presence of sales assistant,
3. special presentation in shop windows, on the sales of their portfolio in 30 shops.

The 30 shops were divided into 3 groups of 10 shops. The sales using the strategies1, 2, and 3 were $y_1 = (9, 11, 10, 12, 7, 11, 12, 10, 11, 13)^\top$, $y_2 = (10, 15, 11, 15, 15, 13, 7, 15, 13, 10)^\top$, and $y_3 = (18, 14, 17, 9, 14, 17, 16, 14, 17, 15)^\top$, respectively. Define x_i as the index of the shop, i.e., $x_i = i,\ i = 1, 2, ..., 30$. Using this notation, the null hypothesis corresponds to a constant regression line, $EY = \mu$. What does the alternative hypothesis involving a regression curve look like?

Problem 14

Let $X \sim N_p(\mu, \Sigma)$. Show the random variable $U = (X - \mu)^\top \Sigma^{-1}(X - \mu)$ is $\mathcal{X}^2_P$ distributed.

Problem 15

Suppose that$X$ has mean zero and covariance $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Let $Y = X_1 + X_2$. Write Y as a linear transformation, i.e., find the transformation matrix $\mathcal{A}$. Then compute $\text{Var}(Y)$.

Problem 16

For the following functions,

$$\begin{array}{lcll} f_1(x_1,x_2) & = & 4x_1x_2\exp(-x_1^2), & x_1,x_2 > 0, \\ f_2(x_1,x_2) & = & 1, & 0 < x_1,x_2 < 1 \text{ and } x_1 + x_2 < 1, \\ f_3(x_1,x_2) & = & \frac{1}{2}\exp(-x_1), & x_1 > |x_2|. \end{array}$$

check if they are pdfs and then compute the $E(X)$, $\text{Var}(X)$, $E(X_1|X_2)$, $E(X_2|X_1)$, $\text{Var}(X_1|X_2)$ and $\text{Var}(X_2|X_1)$.

Exercise 5 (12.11.2015)

Problem 17

Consider the pdf

$$f(x_1,x_2) = \frac{3}{4}x_1^{-\frac{1}{2}},\ 0 < x_1 < x_2 < 1.$$

Compute the following probability $P(X_1 < 0.25)$, $P(X_2 < 0.25)$ and $P(X_2 < 0.25|X_1 < 0.25)$.

Problem 18

Let X be exponentially distributed with following density function,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Show the characteristic function of the random variable X , i.e. $E(e^{itX})$.

Exercise 6 (19.11.2015)

Problem 19

Consider $X \sim N_2(\mu, \Sigma)$ with $\mu = (2, 2)^\top$ and $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and the matrices $\mathcal{A} = (1, 1)$ and $\mathcal{B} = (1, -1)$. Show that $\mathcal{A}X$ and $\mathcal{B}X$ are independent.

Problem 20

Given that $X_1 \sim N_r(\mu_1, \Sigma_{11})$ and $(X_2|X_1 = x_1) \sim N_{p-r}(\mathcal{A}x_1 + b, \Omega)$ where Ω does not depend on x_1, then $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_P(\mu, \Sigma)$, where $\mu = \begin{pmatrix} \mu_1 \\ \mathcal{A}\mu_1 + b \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}\mathcal{A}^\top \\ \mathcal{A}\Sigma_{11} & \Omega + A\Sigma_{11}\mathcal{A}^\top \end{pmatrix}$.

Problem 21

Compute the following complex problems i^{4n}, i^{4n+1},i^{4n+2},i^{4n+3}, $\sqrt{7 + 24i}$, $\frac{1+i}{7-i}$.

Exercise 7 (26.11.2015)

Problem 22

Let $X = (X_1, X_2)^\top \sim N_2(0_2, I_2)$. Determine the distribution of the random vector $Y = \mathcal{A}X$ with $\mathcal{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. And show that the transformed random variables Y_1 and Y_2 are independent.

Problem 23

Let $X \sim N_2\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & a \\ a & 2 \end{pmatrix}\right)$.

a) Represent the contour ellipses for $a = 0;\ -\frac{1}{2};\ \frac{1}{2};\ 1$.

b) For $a = \frac{1}{2}$ find the regions of X centered on μ which cover the area of the true parameter with probability 0.90 and 0.95.

Exercise 8 (03.12.2015)

Problem 24

Let $X \sim N_2\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2 & 1\\1 & 2\end{pmatrix}\right)$ and $Y|X \sim N_2\left(\begin{pmatrix}x_1\\x_1+x_2\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$.

a) Determine the distribution of $Y_2|Y_1$.

b) Determine the distribution of $W = X - Y$.

Problem 25

It is set that

$$\begin{aligned} Z &\sim N_1(0,1), \\ Y|Z &\sim N_1(1+z,1), \\ X|Y,Z &\sim N_1(1-y,1). \end{aligned}$$

a) Determine the distribution of $\begin{pmatrix}X\\Y\\Z\end{pmatrix}$ and of $Y|(X,Z)$.

b) Determine the distribution of $\begin{pmatrix}U\\V\end{pmatrix} = \begin{pmatrix}1+Z\\1-Y\end{pmatrix}$.

c) Determine the distribution of $Y|U=2$.

Exercise 9 (10.12.2015)

Problem 26

Show that $E(X_2) = E\{E(X_2|X_1)\}$, where $E(X_2|X_1)$ is the conditional expectation of X_2 given X_1.

Problem 27

Compute the pdf of the random vector $Y = \mathcal{A}X$ with $\mathcal{A} = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}$ for the random vector X with the pdf:

$$f_X(x) = f_X(x_1,x_2) = \begin{cases} \frac{1}{2}x_1 + \frac{3}{2}x_2, & 0 \le x_1, x_2 \le 1, \\ 0, & else. \end{cases}$$

Problem 28

Set $f(x_1,x_2,x_3) = k(x_1 + x_2x_3);\ 0 < x_1,x_2,x_3 < 1$.

a) Determine k so that f is a valid pdf of $X := (X_1, X_2, X_3)$.
b) Compute the (3×3) variance-covariance matrix Σ_X.
c) Compute the (2×2) conditional variance-covariance matrix of (X_2, X_3) given $X_1 = x_1$.

Exercise 10 (17.12.2015)

Problem 29

Consider the pdf

$$f(x_1, x_2) = \frac{1}{8x_2} \exp\left\{-\left(\frac{x_1}{2x_2} + \frac{x_2}{4}\right)\right\}, \; x_1, x_2 > 0.$$

Compute $f(x_2)$ and $f(x_1|x_2)$. Also give the best (MSE) approximation of X_1 by a function of X_2. Compute the variance of the error of the approximation.

Problem 30

Consider $X \sim N_3 \left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix} \right)$.

a) Find the best linear approximation of X_3 by a linear function of X_1 and X_2 and compute the multiple correlation coefficient between X_3 and (X_1, X_2).
b) Let $Z_1 = X_2 - X_3, Z_2 = X_2 + X_3$ and $(Z_3|Z_1, Z_2) \sim N(Z_1 + Z_2, 10)$. Compute the distribution of $\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$.

Problem 31

Consider $X \sim N \left(\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 & 1 & 2 & 4 \\ 1 & 4 & 2 & 1 \\ 2 & 2 & 16 & 1 \\ 4 & 1 & 1 & 9 \end{pmatrix} \right)$.

a) Give the best linear approximation of X_2 as a function of (X_1, X_4).
b) Give the best linear approximation of X_2 as a function of (X_1, X_3, X_4) .

Problem 32

Let $(X, Y, Z)^\top$ be a tri-variate normal r.v. with $Y|Z \sim N(2Z, 24)$, $Z|X \sim N(2X+3, 14)$, $X \sim N(1, 4)$ and $\rho_{XY} = 0.5$. Find the distribution of $(X, Y, Z)^\top$

and compute the partial correlation between X and Y for fixed Z. Do you think it is reasonable to approximate X by a linear function of Y and Z?

Exercise 11 (14.01.2016)

Problem 33

Consider an iid sample of size n from a bivariate population with pdf $f(x_1, x_2) = \frac{1}{\theta_1\theta_2} \exp\left\{-\left(\frac{x_1}{\theta_1} + \frac{x_2}{\theta_2}\right)\right\}$, $x_1, x_2 > 0$. Compute the MLE of $\theta = (\theta_1, \theta_2)^\top$. Find the Cramer-Rao lower bound. Is it possible to derive a minimum variance unbiased estimator of θ?

Problem 34

Consider an iid sample of size n from the bivariate population with pdf

$$f(x_1, x_2) = \frac{1}{\theta_1^2\theta_2 x_2} \exp\left\{-\left(\frac{x_1}{\theta_1 x_2} + \frac{x_2}{\theta_1\theta_2}\right)\right\}, \; x_1, x_2 > 0.$$

Compute the MLE, $\hat{\theta}$, of the unknown parameter $\theta = (\theta_1, \theta_2)^\top$. Find the Cramer-Rao lower bound and the asymptotic variance of $\hat{\theta}$.

Exercise 12 (21.01.2016)

Problem 35

Consider a sample $\{x_i\}_{i=1}^n$ from $N_p(\theta, I_p)$, where $\theta \in \mathbb{R}^p$ is the mean vector parameter. Show that the MLE of θ is the minimum variance estimator.

Problem 36

Consider an iid sample $\{x_i\}_{i=1}^n$ from $N_p(\theta, \mathit{\Sigma}_0)$ where $\mathit{\Sigma}_0$ is known. Compute the Cramer-Rao lower bound for μ. Can you derive a minimum variance unbiased estimator for μ?

Problem 37

Let $X \sim N_p(\mu, \Sigma)$ where Σ is unknown but we know that $\Sigma = diag(\sigma_{11}, \sigma_{22}, ..., \sigma_{pp})$. From an iid sample of size n, find the MLE of μ and of Σ.