UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 1. stopnja

Jaka Munda

**Ocene parametrov večrazsežne normalne porazdelitve z metodo največjega verjetja**

Delo diplomskega seminarja

Mentor: doc. dr. Aljoša Peperko

Ljubljana, 2019

## Kazalo

**Ocene parametrov večrazsežne normalne porazdelitve z metodo največjega verjetja**

Povzetek

V diplomski nalogi obravnavamo primera podatkov z manjkajočimi podatki in primer brez manjkajočih podatkov, ki izhajajo iz zaporedja slučajnih vektorjev, ki so neodvisno enako porazdeljeni z večrazsežno normalno porazdelitvijo s parametroma vektorjem matematičnega upanja $\mu$ in kovariančno matriko $\Sigma$. Za vsako obliko podatkov lahko po metodi največjega verjetja izračunamo cenilki parametrov porazdelitve. Pristopov za izračun cenilke po metodi največjega verjetja je več, v delu obravnavamo pristopa z matričnim odvajanjem in matrično transformacijo. Obravnavamo še monoton vzorec, ki je poseben primer manjkajočih podatkov, za katerega prav tako izračunamo cenilki za parametra po metodi največjega verjetja.

**Estimation of multivariate normal distribution with maximum likelihood estimation**

Abstract

In this paper we consider sample with missing data and sample without missing data, that comes from multivariate normal distribution with parameters mean vector $\mu$ and covariance matrix $\Sigma$. No matter the shape of the data we can estimate parameters with maximum likelihood estimation. There are various techniques for estimating parameters with maximum likelihood estimation. We consider two techniques, namely, matrix differentiation and matrix transformation. With both techniques we must derivate likelihood function that we get from the sample. We also consider monotone sample, which is a special case of missing data for which we can also estimate parameters with method of maximum likelihood estimation.

**Math. Subj. Class. (2010):** 62H12, 62F10

**Ključne besede:** večrazsežna normalna porazdelitev, metoda največjega verjetja, matrično odvajanje, monoton vzorec

**Keywords:** multivariate normal distribution, maximum likelihood estimation, matrix differentiation, monotone sample

## 1. Uvod

Večrazsežna normalna porazdelitev s parametroma $\mu$ in $\Sigma$, ki predstavljata vektor matematičnih upanj in kovariančno matriko, v statistiki najpogosteje nastopa, kot vektor enorazsežnih normalnih porazdelitev. Za vzorec, ki pripada zaporedju $X_1, \ldots, X_N$ neodvisno enako porazdeljenih slučajnih vektorjev, ki so p-razsežno normalno porazdeljeni s parametroma $\mu$ in $\Sigma$, lahko z uporabo cenilk, ki ju dobimo po metodi največjega verjetja, ocenimo parametra porazdelitve. Cenilki po metodi največjega verjetja izračunamo na splošnem vzorcu tako, da zapišemo funkcijo verjetja $L(\mu, \Sigma; x) = f_{X_1}(x_1) \cdots f_{X_N}(x_N)$, iz katere izračunamo cenilki $\mu$ in $\Sigma$ tako, da je za njiju največja možna verjetnost oz. jo maksimiziramo glede na $\mu$ in $\Sigma$. Osnovni pristop je enačanje prvega odvoda z nič, ker maksimiziramo vektor in matriko moramo poznati nekaj formul za matrično odvajanje. Maksimizacijo si lahko olajšamo s transformacijo matrik in tako preuredimo matrično funkcijo v funkcijo realne spremenljivke. Razvitih je bilo še veliko drugih tehnik, vendar za noben pristop ne moremo reči, da je najučinkovitejši. Za praktično vsako obliko podatkov je bil razvit poseben pristop in nekateri pristopi delujejo samo na točno določenih oblikah podatkov.Vedno pa je potrebno na takšen ali drugačen način maksimizirati funkcijo verjetja. Če imamo poleg vzorca podane še kakšne dodatne pogoje, kot je omejitev ranga matrike, uporabimo Lagrangeve multiplikatorje.

V diplomski nalogi obravnavamo dva primera vzorcev in sicer vzorec z manjkajočimi podatki in vzorec brez manjkajočih podatkov.V prvem razdelku obravnavamo vzorec, kjer nimamo manjkajočih podatkov in izračunamo cenilki za $\mu$ in $\Sigma$ z odvajanjem funkcije verjetja. Cenilko za $\Sigma$ izračunamo tudi z transformacijo matrik. V drugem razdelku obravnavamo dva različna vzorca, v katerih nam del podatkov manjka, ker se nam matrična funkcija verjetja s transformacijo ne poenostavi v funkcijo realne spremenljivke, se transformacije ne poslužejemo in kar direktno odvajamo funkcijo verjetja. V zadnjem razdelku obravnavamo popoln monoton vzorec, ki je poseben primer manjkajočih podatkov in eden najpomembnejših, saj v takem vzorcu dobimo dovolj podatkov za oceno parametrov. Običajno namreč v vzorcih z manjkajočimi podatki ne dobimo dovolj podatkov za oceno parametrov porazdelitve. V tem razdelku je prav tako zapisan algoritem za uporabo v Mathematici.

## 2. Ocenjevanje večrazsežne normalne porazdelitve, ko nimamo manjkajočih podatkov

Za p-razsežno normalno porazdelitev $X \sim N_p(\mu, \Sigma)$, kjer je $\mu = (\mu_1, \ldots, \mu_p)^T$ vektor matematičnih upanj in $\Sigma = (\sigma_{ij})_{(p \times p)} > 0$ kovariančna matrika, zapišemo gostoto:

$$f_X(x) = (2\pi)^{-\frac{p}{2}} (det\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}. \tag{1}$$

Z $\Sigma > 0$ smo označili pozitivno definitno matriko. Za zaporedje slučajnih vektorjev $X = \{X_i = (X_{i,1}, \ldots, X_{i,p})^T; i = 1, \ldots, N\}$, ki so neodvisno enako porazdeljeni s p-razsežno normalno porazdelitvijo, lahko zapišemo funkcijo verjetja za nek poljuben vzorec:

$$
\begin{aligned}
L(\mu, \Sigma; x) &= \prod_{i=1}^{N} (2\pi)^{-\frac{p}{2}} (\det \Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)} \\
&= (2\pi)^{-\frac{pN}{2}} (\det \Sigma)^{-\frac{N}{2}} e^{-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)}.
\end{aligned}
$$

Iz funkcije verjetja želimo izračunati cenilki za $\mu$ in $\Sigma$ tako, da bo verjetnost za njiju največja možna. Ker je logaritemska funkcija naraščajoča, je ekvivalentno maksimizirati logaritem funkcije verjetja:

$$
\begin{aligned}
l(\mu, \Sigma; x) &= -\frac{pN}{2} \log 2\pi - \frac{N}{2} \log (\det \Sigma) + \log e^{-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)} \\
&= -\frac{pN}{2} \log 2\pi - \frac{N}{2} \log (\det \Sigma) - \frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu).
\end{aligned}
$$

Za izračun cenilke $\mu$ moramo maksimizirati samo vsoto. Preden maksimiziramo, se spomnimo dejstva, da, če je $A$ pozitivno definitna matrika, ima kvadratna forma oblike:

$$
\frac{1}{2} x^T A x - b^T x + c, \tag{2}
$$

enoličen minimum za $x$. Ker je $\Sigma$ pozitivno definitna, je potem tudi $\Sigma^{-1}$ pozitivno definitna. Vidimo, da sumandi iz vsote, ki nastopajo v logaritmu funkcije verjetja, ustrezajo kvadratni formi iz (2), kar pomeni, da ima vsak sumand enoličen minimum, iz česar sledi, da ga ima tudi vsota. Ker imamo vsoto pomnoženo z $-1$, bomo dobili z odvajanjem maksimum. Maksimum poiščemo z enačenjem prvega odvoda z 0, pri čemer moramo poznati spodnje formule za odvajanje vektorja. Spodnjo definicijo najdemo na [3, str. 4, Definicija 3].

**Definicija 2.1.** Naj bo $y = \Psi(x)$, kjer je $y$ $m \times 1$ in $x$ $n \times 1$, potem je odvod po $y$ glede na $x$ matrika dimenzije $m \times n$

$$
\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{dy_1}{dx_1} & \cdots & \frac{dy_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{dy_m}{dx_1} & \cdots & \frac{dy_m}{dx_n} \end{bmatrix}.
$$

Dokaz spodnje trditve za odvajanje vektorja najdemo na [3, str. 7, Trditev 9].

**Trditev 2.2.** *Naj bo $\alpha = x^T A x$, kjer je $x$ $n \times 1$ in $A$ je $n \times n$ simetrična matrika, ki ni odvisna od $x$, potem je odvod $\alpha$ po $x$ enak*

$$
\frac{\partial \alpha}{\partial x} = 2x^T A.
$$

Dokaz naslednje trditve najdemo na [3, str. 8, Trditev 14].

**Trditev 2.3.** *Naj bo $\alpha = x^T A x$, kjer je $x$ $n \times 1$ funkcija odvisna od vektorja $z$ in $A$ je $m \times n$ simetrična matrika ter neodvisna od $z$, potem je*

$$
\frac{\partial \alpha}{\partial z} = 2x^T A \frac{\partial x}{\partial z}.
$$

Z maksimizacijo logaritma funkcije verjetja dobimo:

$$\begin{aligned}\frac{\partial l(\mu,\Sigma;x)}{\partial \mu} &= \sum_{i=1}^{N} \Sigma^{-1}(x_i-\hat{\mu}) = 0\\ &\sum_{i=1}^{N}(x_i-\hat{\mu}) = 0\\ &\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i.\end{aligned}$$

Tako smo pokazali, da je cenilka za matematično upanje enaka $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i = \bar{X}$. Opazimo, da je cenilka enaka kot cenilka, ki jo dobimo po metodi največjega verjetja za enorazsežno normalno porazdelitev.

2.1. **Iskanje cenilke $\Sigma$ z odvajanjem.** Za izračun cenilke $\Sigma$ preuredimo logaritem funkcije verjetja tako, da zamenjamo matematično upanje z njegovo cenilko in upoštevamo, da je $(x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x})$ skalar za poljuben $i=1,\ldots,N$ ter, da je sled skalarja enaka skalarju. Tako dobimo sledeči logaritem funkcije verjetja:

$$\begin{aligned}l(\mu,\Sigma;x) &= -\frac{pN}{2}\log 2\pi - \frac{N}{2}\log(\det\Sigma) - \frac{1}{2}\sum_{i=1}^{N} tr\left((x_i-\bar{x})^T\Sigma^{-1}(x_i-\bar{x})\right)\\ &= -\frac{pN}{2}\log 2\pi - \frac{N}{2}\log(\det\Sigma) - \frac{N}{2}tr\left(\Sigma^{-1}\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})^T\right)\\ &= -\frac{pN}{2}\log 2\pi - \frac{N}{2}\log(\det\Sigma) - \frac{N}{2}tr(\Sigma^{-1}V)\\ &= -\frac{pN}{2}\log 2\pi + \frac{1}{2}Nf(\Sigma;V),\end{aligned}$$

kjer sta:

$$\begin{aligned}V &= \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})^T,\\ f(\Sigma;V) &= -\log(\det\Sigma) - tr(\Sigma^{-1}V).\end{aligned}$$

V izpeljavi smo v prehodu iz prve v drugo vrstico upoštevali cikličnost sledi in to, da je vsota sledi enaka sledi vsot. Matrika $V$ $p\times p$ predstavlja vzorčno kovariančna matriko in funkcija $f(\Sigma;V)$ je funkcija, katero bomo maksimizirali glede na $\Sigma$, saj samo v tej funkciji nastopa $\Sigma$. Če v funkcijo $f(\Sigma;V)$ vstavimo $\Psi=\Sigma^{-1}$ in se spomnimo, da velja $\det(A)^{-1} = (\det A)^{-1}$ dobimo:

(3) $$g(\Psi;V) = \log(\det\Psi) - tr(\Psi V).$$

Pripomnimo še, da je V vedno simetrična pozitivno semidefinitna matrika. Simetričnost sledi že iz same definicije matrike, semipozitivnosti pa ni težko pokazati:

$$
\begin{aligned}
y^T V y &= y^T (\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})^T)y \\
&= \frac{1}{N}\sum_{i=1}^{N} y^T(x_i-\bar{x})(x_i-\bar{x})^T y \\
&= \frac{1}{N}\sum_{i=1}^{N}((x_i-\bar{x})^T y)^2 \geq 0,
\end{aligned}
$$

kar velja za vsak neničeln vektor $y \in \mathbb{R}^p$, torej je $V$ pozitivno semidefinitna. Vidimo, da je zgornji izraz enak nič natanko tedaj, ko obstaja tak $y$, da za vsak $i = 1, \ldots, N$ velja $(x_i - \bar{x})^T y = 0$, kar je res, ko je $y$ pravokoten na vsak $(x_i - \bar{x})$. Če je $N > p$ potem je $V$ pozitivno definitna z verjetnostjo 1 (povzeto po [2, str. 149, Razdelek 2]).

Izkaže se, da je cenilka za $\Sigma$ po metodi največjega verjetja enaka $V = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(X_i - \bar{X})^T$. To bomo pokazali na dva načina. Prvi način je z matričnim odvajanjem funkcij $f(\Sigma; V)$ in $g(\Psi; V)$, drugi pa z transformacijo matrik $\Sigma$ in $V$ in odvajanjem poenostavljenih funkcij. Za odvajanje matričnih funkcij bomo potrebovali naslednjo definicijo in trditvi. Spodnjo definicijo najdemo na [3, str. 8, Definicija 5].

**Definicija 2.4.** Naj bo $A$ matrika dimenzije $m \times n$, katere elementi so funkcije parametra $\alpha$. Potem je odvod $A$ glede na $\alpha$ enak

$$
\frac{\partial A}{\partial \alpha} = \begin{bmatrix} \frac{da_{11}}{d\alpha} & \frac{da_{12}}{d\alpha} & \cdots & \frac{da_{1n}}{d\alpha} \\ \frac{da_{21}}{d\alpha} & \frac{da_{22}}{d\alpha} & \cdots & \frac{da_{2n}}{d\alpha} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{da_{m1}}{d\alpha} & \frac{da_{m2}}{d\alpha} & \cdots & \frac{da_{mn}}{d\alpha} \end{bmatrix} = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{da_{ij}}{d\alpha}E_{ij},
$$

kjer je $E_{ij}$ matrika z 1 na $(i, j)$-tem mestu in 0 drugje.

Dokaz trditve o odvodu inverza matrike, najdemo na [3, str. 8-9, Trditev 15].

**Trditev 2.5.** *Naj bo $A$ nesingularna matrika dimenzije $n \times n$ katere elementi so funkcije parametra $\alpha$. Potem velja*

$$
\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1}\frac{\partial A}{\partial \alpha}A^{-1}.
$$

Trditev o odvodu determinante, najdemo na [2, str. 150, Trditev 2].

**Trditev 2.6.** *Naj bo $A$ $n \times n$ nesingularna matrika, potem je odvod determinante matrike $A$ enak*

$$
\frac{\partial \det(A)}{\partial a_{ij}} = (2 - \delta_{ij}) A_{ij} (da_{ij}),
$$

*kjer je $\delta_{ij}$ Kroneckerjeva delta*

$$
\delta_{ij} = \begin{cases} 1; i = j \\ 0; sicer \end{cases}
$$

*Element na $(i, j) - tem$ mestu matrike $A$ označimo z $a_{ij}$, $A_{ij}$ označuje kofakor matrike glede na $i - to$ vrstico in $j - ti$ stolpec, torej je to predznačena poddeterminanta $A$ glede na $(i, j) - ti$ element matrike.*

2.1.1. *Odvajanje* $f(\Sigma; V)$. Preden se lotimo odvajanja se je potrebno prepričati o obstoju maksimuma. Ko se $\Sigma$ približuje meji pozitivno definitnih matrik oz., ko se vsaj ena lastna vrednost matrike $\Sigma$ približuje nič, gre funkcija $f(\Sigma; V)$ proti minus neskončno, kar pomeni, da mora biti maksimum na območju pozitivno definitnih matrik. Da gre $f(\Sigma; V)$ proti minus neskončno, ko se lastna vrednost $\Sigma$ približuje nič, vidimo iz dejstva, da je $\det \Sigma = \lambda_1 \cdots \lambda_p$, kjer je $\lambda_i$ lastna vrednost $\Sigma$ za $i = 1, \dots, p$, torej gre determinanta proti nič, ko gre lastna vrednost proti nič in $-\log(\det \Sigma)$ proti neskončno, $-tr(\Sigma^{-1}V)$ pa gre proti minus neskončno, ker je $\Sigma^{-1} = \frac{1}{\det \Sigma} adj(\Sigma)$ in ko gre determinanta proti nič, gre $\frac{1}{\det \Sigma}$ proti neskončno in zato gre nasprotna vrednost sledi proti minus neskončno. Funkcija $f(\Sigma; V)$ gre proti minus neskončno, ker sled hitreje pada proti minus neskončno kot logaritem narašča proti neskončno. Funkcija ni ne konveksna, ne konkavna, ker je $-\log(\det \Sigma)$ konveksna, $-tr(\Sigma^{-1}V)$ pa je konkavna. Vendar se kljub temu izkaže, da ima enačba, ki jo dobimo z odvajanjem samo eno rešitev, kar pomeni, da je maksimum enoličen. Za izračun maksimuma najprej odvajamo posebej logaritem determinante in sled produkta matrik, pri čemer uporabimo izrek za odvajanje determinante (glej 2.6), formulo za odvod matrike (glej 2.4) in formulo za odvod inverza matrike (glej 2.5):

$$
\begin{aligned}
\frac{\partial}{\partial \sigma_{ij}} \left(-\log\left(\det \Sigma\right)\right) &= (2 - \delta_{ij}) \left( -\frac{\Sigma_{ij}}{\det \Sigma} d\sigma_{ij} \right), \\
\frac{\partial}{\partial \Sigma} \left(-tr \Sigma^{-1} V\right) &= tr\left(\Sigma^{-1} \left(d\Sigma\right) \Sigma^{-1} V\right) \\
&= (2 - \delta_{ij}) tr\left(\Sigma^{-1} E_{ij} \Sigma^{-1} V\right) d\sigma_{ij}, \quad \forall i \leq j.
\end{aligned}
$$

Z $\Sigma_{ij}$ smo označili kofaktor matrike $\Sigma$ glede na i-to vrstico in j-ti stolpec, z $\delta_{ij}$ smo označili Kroneckerjevo delto, ki je 1, če je $i = j$ in 0 sicer in $E_{ij}$ je matrika, ki ima 1 na $(i, j)$-tem mestu in 0 povsod drugje. Sledi, da je odvod $f(\Sigma; V)$ enak:

$$
\begin{aligned}
d\left[f\left(\Sigma; V\right)\right]_{ij} = (2 - \delta_{ij}) &\left\{ -\frac{\Sigma_{ij}}{\det \Sigma} d\sigma_{ij} + tr\left(\Sigma^{-1} E_{ij} \Sigma^{-1} V\right) d\sigma_{ij} \right\} = 0 \\
(2 - \delta_{ij}) &\left\{ -\frac{\Sigma_{ij}}{\det \Sigma} + tr\left(E_{ij} \Sigma^{-1} V \Sigma^{-1}\right) \right\} = 0.
\end{aligned}
$$

Z upoštevanjem $\Sigma^{-1} = \frac{1}{\det \Sigma} adj\left(\Sigma\right)$ lahko zapišemo zgornji sistem enačb, kot matrično enačbo:

$$
\begin{aligned}
-\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1} V \hat{\Sigma}^{-1} &= 0 \\
\hat{\Sigma}^{-1} V \hat{\Sigma}^{-1} &= \hat{\Sigma}^{-1} \\
V \hat{\Sigma}^{-1} &= I \\
\hat{\Sigma} &= V
\end{aligned}
$$

2.1.2. *Odvajanje* $g(\Psi; V)$. Zopet se je potrebno prepričati o obstaju in enoličnosti maksimuma. Ko se $\Psi$ približuje meji pozitivno definitnih matrik, gre $g(\Psi; V)$ proti minus neskončno, kar vidimo na podoben način, kot smo za funkcijo $f(\Sigma; V)$ le, da gresta tu oba člena funkcije proti minus neskončno. Ker je $\log\left(\det \Psi\right)$ konkavna in

$tr\Psi V$ linearna, je celotna funkcija $g(\Psi;V)$ konkavna. Torej imamo enoličen maksimum na območju pozitivno definitnih matrik in bomo z odvajanjem dobili največjo vrednost funkcije. Če odvajamo $g(\Psi;V)$, dobimo:

$$d\left[g(\Psi;V)\right]_{ij} = (2-\delta_{ij})\left\{\tfrac{\Psi_{ij}}{\det\Psi}d\psi_{ij} - (trE_{ij}V)\,d\psi_{ij}\right\} = 0, \quad \forall\ i \le j.$$

Za ta sistem enačb lahko zapišemo matrično enačbo:

$$\hat{\Psi}^{-1} - V = 0.$$

Cenilka za $\Psi$ je enaka $\hat{\Psi} = V^{-1}$, kar pomeni, da je cenilka za $\Sigma$ enaka $V$. Tako smo pokazali, da dobimo isto cenilko za $\Sigma$ ne glede na to, če odvajamo $f(\Sigma;V)$ ali $g(\Psi;V)$, vendar imamo pri odvajanju slednje nekoliko manj dela.

2.2. **Iskanje cenilke za $\Sigma$ z transformacijo matrik.** Namesto, da odvajamo matrično funkcijo, jo lahko s transformacijo matrik preoblikujemo v funkcijo realne spremenljivke in se tako izognemo matričnemu odvajanju. Za poljubno transformacijo $V = CC^T$, kjer je $C$ kvadratna obrnljiva matrika, lahko zapišemo:

$$\tilde{\Sigma} = C^{-1}\Sigma(C^T)^{-1}, \quad \tilde{\Psi} = C^T\Psi C.$$

Z upoštevanjem tega se funkciji $f\left(\Sigma;V\right)$ in $g\left(\Psi;V\right)$ preoblikujeta v:

$$\begin{aligned}
f\left(\tilde{\Sigma};V\right) &= -\log\left(\det C\tilde{\Sigma}C^T\right) - tr\left(\left(C\tilde{\Sigma}C^T\right)^{-1}V\right)\\
&= -\log\left(\det C\det\tilde{\Sigma}\det C^T\right) - tr\left((C^T)^{-1}\tilde{\Sigma}^{-1}C^{-1}CC^T\right)\\
&= -\log\left(\det V\det\tilde{\Sigma}\right) - tr\left(\tilde{\Sigma}^{-1}C^T(C^T)^{-1}\right)\\
&= -\log\left(\det V\right) - \log\left(\det\tilde{\Sigma}\right) - tr\tilde{\Sigma}^{-1},\\
g\left(\tilde{\Psi};V\right) &= \log\left(\det(C^T)^{-1}\tilde{\Psi}C^{-1}\right) - tr\left((C^T)^{-1}\tilde{\Psi}C^{-1}CC^T\right)\\
&= \log\left(\det(\left(C^T\right)^{-1}C^{-1})\det\tilde{\Psi}\right) - tr(\tilde{\Psi}C^T(C^T)^{-1})\\
&= -\log\left(\det V\right) + \log\left(\det\tilde{\Psi}\right) - tr\tilde{\Psi}.
\end{aligned}$$

Ker že poznamo $V$, je $\log\left(\det V\right)$ konstanta in ne vpliva na maksimum. S transformacijo matrike $\tilde{\Sigma}$ oz. $\tilde{\Psi}$ lahko prevedemo zgornji matrični funkciji na funkcijo realne spremenljivke. V nadaljevanju bomo transformirali matriko $\tilde{\Psi}$ in tako prevedli matrično funkcijo $g(\tilde{\Psi};V)$ na funkcijo realne spremenljivke. Uporabili bomo naslednje tri transformacijske izreke. Dokaz razcepa Choleskega najdemo na [8, str. 84, Trditev 3.20].

**Trditev 2.7. Razcep Choleskega** *Če je $H > 0$, potem obstaja spodnje trikotna matrika $T = (t_{ij})$ tako, da velja $H = TT^T$, kjer so diagonalni elementi $t_{ii} > 0$, za $i = 1, \ldots, p$.*

Spodnji trditvi sta povzeti po [2, str. 153, Trditev 4 in Trditev 5].

**Trditev 2.8.** *Če je $H > 0$, potem velja $H = DRD$, kjer je $D = Diag(\sqrt{h_{11}}, \ldots, \sqrt{h_{pp}})$, $h_{ii} > 0$ za $i = 1, \ldots, p$ in $R$ je korelacijska matrika, kar pomeni, da je pozitivno definitna matrika z enkami na diagonali.*

**Trditev 2.9. Schurov izrek** *Naj bo $H$ simetrična, potem obstaja ortogonalna matrika $U$ tako, da velja $H = UD_dU^T$, kjer je $D_d = Diag\,(d_1, \dots, d_p)$ in $d_1 \geq \cdots \geq d_p$ so urejene lastne vrednosti matrike $H$. Če je $H > 0$ potem je $d_p > 0$.*

2.2.1. *Transformacija z razcepom Choleskega.* Z uporabo (2.7) dobimo $\tilde{\Psi} = TT^T$, kjer je $T = (\tau_{ij})$ spodnje trikotna matrika s pozitivnimi diagonalnimi elementi. Če transformacijo upoštevamo v $g(\tilde{\Psi}; I)$ dobimo:

$$\begin{aligned} g\left(\tilde{\Psi}; I\right) &= \log\left(\det TT^T\right) - trTT^T \\ &= \log\left(\prod_{i=1}^{p} \tau_{ii} \prod_{i=1}^{p} \tau_{ii}\right) - \sum_{i=1}^{p}\left(\tau_{ii}^2 + \sum_{i>j}^{p} \tau_{ij}^2\right) \\ &= \sum_{i=1}^{p}\left(\log\left(\tau_{ii}^2\right) - \tau_{ii}^2 - \sum_{i>j}^{p} \tau_{ij}^2\right). \end{aligned}$$

Pri tem smo upoštevali, da je determinanta produkta matrik enaka produktu determinant matrik. Za sled pa upoštevamo, da je produkt matrik enak:

$$TT^T = \begin{bmatrix} \tau_{11} & 0 & 0 \\ \vdots & \ddots & 0 \\ \tau_{p1} & \cdots & \tau_{pp} \end{bmatrix} \begin{bmatrix} \tau_{11} & \cdots & \tau_{p1} \\ 0 & \ddots & \vdots \\ 0 & 0 & \tau_{pp} \end{bmatrix} = \begin{bmatrix} \tau_{11}^2 & * & * \\ * & \ddots & * \\ * & * & \tau_{p1}^2 + \cdots + \tau_{pp}^2 \end{bmatrix}.$$

Vidimo, da bo vrednost funkcije $g(\tilde{\Psi}; I)$ največja takrat, ko je $\sum_{i>j}^{p} \tau_{ij}^2 = 0$ oz. ko je vsak $\tau_{ij}$ enak nič za $i > j$. Preostali del enačbe moramo odvajati. Preden se lotimo dela, se je potrebno prepričati o obstoju maksimuma na danem območju. Lahko je preveriti, da je $\log\left(\tau_{ii}^2\right) - \tau_{ii}^2$ konkavna funkcija na $\tau_{ii} > 0$, torej maksimum na danem območju obstaja in je enoličen. Z odvajanjem dobimo spodnjo enačbo:

$$\begin{aligned} \frac{dg(\tilde{\Psi}; I)}{d\tau_{ii}} = \frac{2}{\hat{\tau}_{ii}} - 2\hat{\tau}_{ii} &= 0 \\ \hat{\tau}_{ii}^2 &= 1. \end{aligned}$$

Ker so diagonalni elementi matrike $T$ strogo večji od nič, mora biti maksimum enak $\tau_{ii} = 1$ za $i = 1 \dots, p$. Kar pomeni, da sta cenilki za matriki $T$ in $\tilde{\Psi}$ identiteti. Cenilka za $\Psi$ pa je enaka:

$$\hat{\Psi} = (C^T)^{-1}\hat{\tilde{\Psi}}C^{-1} = (C^T)^{-1}C^{-1} = V^{-1}.$$

Sledi, da je cenilka za $\Sigma$ enaka $V$.

2.2.2. *Transformacija z vektorsko matriko.* Z uporabo (2.8) imamo sledečo transformacijo:

$$\tilde{\psi} = DPD, \quad P = (\rho_{ij}), \quad D = Diag\left(\sqrt{\tilde{\psi}_{11}}, \dots, \sqrt{\tilde{\psi}_{pp}}\right).$$

Če to transformacijo upoštevamo v funkciji $g(\tilde{\Psi}; I)$, dobimo:

$$
\begin{aligned}
g\left(\tilde{\Psi}; I\right) &= \log\left(\det DPD\right) - tr(DPD) \\
&= \log\left(\det D^2\right) + \log\left(\det P\right) - tr(PD^2) \\
&= \log \prod_{i=1}^{p} \tilde{\psi}_{ii} + \log\left(\det P\right) - \sum_{i=1}^{p} \tilde{\psi}_{ii} \\
&= \sum_{i=1}^{p} \left(\log \tilde{\psi}_{ii} - \tilde{\psi}_{ii}\right) + \log\left(\det P\right).
\end{aligned}
$$

Upoštevali smo, da je produkt diagonalnih matrik enak $D^2 = Diag\left(\psi_{11}, \dots, \psi_{pp}\right)$ in da ima korelacijska matrika $P$ na diagonali enke in je zato sled enaka $tr\left(D^2 P\right) = \sum_{i=1}^{p} \psi_{ii}$. Maksimum $\log\left(\det P\right)$ je enak 0, kar pokažemo s Hadamardovo neenakostjo. Dokaz Hadamardove neenakosti najdemo v [10].

**Trditev 2.10. Hadamardova neenakost** *Naj bo $A$ $n \times n$ kvadratna nesingularna matrika in $a_i$ njeni stolpci. Potem velja neenakost*

$$
|\det A| \leq \prod_{i=1}^{n} \|a_i\|_2 .
$$

Ker je $P$ pozitivno definitna matrika, lahko zapišemo alternativno različico Hadamardove neenakosti. Za $P$ zapišemo razcep Choleskega $P = NN^T$, iz česar sledi neenakost $\det P = (\det N)^2 \leq \prod_{i=1}^{p} \|v_i\|_2^2 = \prod_{i=1}^{p} p_{ii}$. Ker je $P$ korelacijska matrika, so diagonalni elementi enaki 1. Opazimo, da enakost za Hadamardovo neenakost nastopi za $P = I$, iz česar sledi, da je maksimum $\log\left(\det P\right)$ enak 0. Še preden enačimo prvi odvod z nič preostalega dela funkcije, se moramo zopet prepričati o obstaju in enoličnosti maksimuma. Vidimo, da je funkcija $\log \tilde{\psi}_{ii} - \tilde{\psi}_{ii}$ konkavna, zato z odvajanjem res dobimo enoličen maksimum:

$$
\begin{aligned}
\frac{\partial g(\tilde{\Psi}; I)}{\partial \tilde{\psi}_{ii}} &= \frac{1}{\hat{\tilde{\psi}}_{ii}} - 1 = 0 \\
\hat{\tilde{\psi}}_{ii} &= 1.
\end{aligned}
$$

Sledi, da je $\hat{\tilde{\Psi}} = I$ in cenilka za $\Psi$ je enaka:

$$
\hat{\Psi} = C^{T^{-1}} \hat{\tilde{\Psi}} C^{-1} = C^{T^{-1}} C^{-1} = V^{-1}.
$$

2.2.3. *Transformacija z lastnimi vrednostmi.* Z uporabo transformacije (2.9) imamo sledečo transformacijo matrike $\tilde{\Psi}$:

$$
\tilde{\Psi} = G D_\lambda G^T, \quad D_d = Diag\left(\lambda_1, \dots, \lambda_p\right),
$$

kjer je $G$ ortogonalna matrika in $\lambda_1 \geq \cdots \geq \lambda_p$, so urejene lastne vrednosti matrike $\tilde{\Psi}$, ker je $\tilde{\Psi} > 0$, so lastne vrednosti pozitivne. Z upoštevanjem te transformacije dobimo:

$$
\begin{aligned}
g\left(\tilde{\Psi}; I\right) &= \log\left(\det GD_\lambda G^T\right) - tr\left(GD_\lambda G^T\right) \\
&= \log\left(\det\left(GG^T\right)\det D_\lambda\right) - tr\left(D_\lambda G^T G\right) \\
&= \log\left(\det D_\lambda\right) - trD_\lambda \\
&= \log\prod_{i=1}^{p}\lambda_i - \sum_{i=1}^{p}\lambda_i \\
&= \sum_{i=1}^{p}\left(\log\left(\lambda_i\right) - \lambda_i\right).
\end{aligned}
$$

Takšno funkcijo smo odvajali že v prejšnem razdelku in ima maksimum v $\lambda_i = 1$ za $i = 1, \dots, p$. Posledično je $\hat{D}_\lambda = I$, cenilka za $\Psi$ pa je enaka:

$$
\hat{\Psi} = C^{T^{-1}}\hat{\tilde{\Psi}}C^{-1} = C^{T^{-1}}C^{-1} = V^{-1}.
$$

2.2.4. *Redukcija dveh matrik hkrati.* Namesto da transformiramo matriki $V$ in $\Sigma$ posebej, lahko z transformacijo obeh matrik hkrati poenostavimo matrično funkcijo $f(\Sigma; V)$ oz. $g(\Psi; V)$ na funkcijo ene spremenljivke. Uporabili bomo spodnjo transformacijo, povzeto po [2, str. 154, Trditev 6].

**Trditev 2.11.** *Če sta $V$ in $\Sigma$ velikosti $p \times p$ pozitivno definitni matriki, potem obstaja nesingularna matrika $L$ tako, da velja $V = LL^T$ in $\Sigma = LD_\lambda L^T$, kjer je $D_\lambda = Diag\left(\lambda_1, \dots, \lambda_p\right)$ in $\lambda_1 \geq \dots \geq \lambda_p > 0$, ki so ničle karakterističnega polinoma* $\det\left(\Sigma - \lambda V\right) = 0$.

Z upoštevanjem transformacije v funkciji $f(\Sigma; V)$ dobimo

$$
\begin{aligned}
f\left(\Sigma; V\right) &= -\log\left(\det\Sigma\right) - tr\left(\Sigma^{-1}V\right) \\
&= -\log\left(\det LD_\lambda L^T\right) - tr\left(\left(L^T\right)^{-1}D_\lambda^{-1}L^{-1}LL^T\right) \\
&= -\log\left(\det LL^T\right) - \log\left(\det D_\lambda\right) - tr\left(D_\lambda^{-1}L^T(L^T)^{-1}\right) \\
&= -\log\left(\det V\right) - \log\left(\det D_\lambda\right) - trD_\lambda^{-1} \\
&= -\log\left(\det V\right) - \sum_{i=1}^{p}\left(\log\lambda_i + \frac{1}{\lambda_i}\right).
\end{aligned}
$$

Upoštevajmo, da $V$ že poznamo in je zato $\log\left(\det V\right)$ konstanta, torej ne vpliva na maksimum. Funkcija $-\log\lambda_i - \frac{1}{\lambda_i}$ sicer ni ne konveksna, ne konkavna, vendar, ker gre funkcija proti minus neskončno, ko $\lambda_i$ narašča preko vseh meja, se mora na danem območju nahajati maksimum in izkaže se, da je ta enoličen. Z odvajanjem dobimo maksimum v $\lambda_i = 1$ za $i = 1, \dots, p$, iz česar sledi, da je $\hat{D}_\lambda = I$, cenilka za $\Sigma$ pa je enaka:

$$
\hat{\Sigma} = L\hat{D}_\lambda L^T = LIL^T = LL^T = V.
$$

Tako smo na več načinov pokazali, da je cenilka za $\Sigma$ enaka $V$. Obstaja še nekaj metod s katerimi pokažemo $\hat{\Sigma} = V$ na vzorcu brez manjkajočih podatkov, vendar se bomo raje osredotočili na vzorce z manjkajočimi podatki.

## 3. Ocenjevanje $\mu$ in $\Sigma$ na vzorcu, kjer deloma poznamo matematično upanje ali pa manjka del podatkov

Do sedaj smo obravnavali primer, kjer nimamo manjkajočih podatkov, vendar se v statistiki velikokrat srečamo s podatki, kjer del le teh manjka. Za številne vzorce z manjkajočimi podatki lahko zapišemo spodnjo funkcijo za ocenjevanje $\Sigma$:

$$f(\Sigma) \equiv -N \log(\det \Sigma) - Ntr\left(\Sigma^{-1} V\right) - M \log(\det \Sigma_{11}) - Mtr\left(\Sigma_{11}^{-1} W\right),$$

kjer je $\Sigma$ $p \times p$ bločna pozitivno definitna matrika oblike $\Sigma = (\Sigma_{ij})$ za $i, j = 1, 2$ z blokoma $\Sigma_{11}$ $k \times k$, $\Sigma_{22}$ $(p-k) \times (p-k)$, $V$ $p \times p$ pozitivno definitna matrika, $W$ $k \times k$ pozitivna semidefintna matrika in $M$ ter $N$ sta pozitivni konstanti oz. velikosti vzorca. Če označimo $\Psi = \Sigma^{-1}$ in Schurov komplement $\Psi_{11\cdot 2} = \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21}$ ter upoštevamo $\Sigma_{11}^{-1} = \Psi_{11\cdot 2}$, potem se zgornja funkcija preoblikuje v:

$$g(\Psi) = N \log(\det \Psi) - Ntr(\Psi V) + M \log(\det \Psi_{11\cdot 2}) - Mtr(\Psi_{11\cdot 2} W).$$

Da pokažemo enakost $\Sigma_{11}^{-1} = \Psi_{11\cdot 2}$, je potrebno poznati formulo za inverz bločne matrike:

$$\begin{aligned}
&\Sigma^{-1} \\
&= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} \\ -\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} & \left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1} \end{bmatrix},
\end{aligned}$$

ki jo upoštevamo v Schurovi formi:

$$\begin{aligned}
&\Psi_{11\cdot 2} \\
&= \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21} \\
&= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} \\
&- \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\left(\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\right)^{-1}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} \\
&= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\left(\Sigma_{22} - \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\Sigma_{12}^T\Sigma_{11}^{-1} \\
&= \Sigma_{11}^{-1}.
\end{aligned}$$

V naslednjih dveh razdelkih bomo predstavili dva primera vzorcev, iz katerih lahko zapišemo zgornji funkciji za izračun cenilke $\Sigma$.

### 3.1. Primer, ko delno poznamo matematično upanje.

Recimo, da imamo (k+l)-razsežno normalno porazdelitev oblike $Z = (X, Y) = (X_1, \ldots, X_k, Y_1, \ldots, Y_l)^T$, za katero matematično upanje $X$-a poznamo in brez škode za splošnost lahko predpostavimo, da je $E(X) = 0$. Matematičnega upanja $Y$-a pa ne poznamo in ga označimo z $E(Y) = \mu$. Kovariančna matrika za $(X, Y)$ je bločna matrika oblike $\Sigma = (\Sigma)$ za $i, j = 1, 2$ z $\Sigma_{11} : k \times k$ in $\Sigma_{22} : l \times l$. Za vzorec, ki izhaja iz zaporedja $Z_1, \ldots, Z_N$ neodvisno enako porazdeljenih slučajnih vektorjev, lahko zapišemo logaritem funkcije verjetja:

$$
\begin{aligned}
l(\mu, \Sigma; z) &= -\frac{(k+l)N}{2}\log 2\pi - \frac{N}{2}\log\left(\det \Sigma\right) - \frac{1}{2}\sum_{i=1}^{N}\left(x_i,\ y_i - \mu\right)^T \Sigma^{-1}\left(x_i,\ y_i - \mu\right) \\
&= -\frac{(k+l)N}{2}\log 2\pi + \frac{N}{2}\log\left(\det \Psi\right) - \frac{1}{2}\sum_{i=1}^{N}\left(x_i,\ y_i - \mu\right)^T \Psi\left(x_i,\ y_i - \mu\right).
\end{aligned}
$$

Za izračun cenilke $\mu$ moramo maksimizirati samo vsoto. Vidimo, da so sumandi vsote oblike (2) in da je $\Psi$ pozitivno definitna matrika, torej ima vsak sumand enoličen minimum, torej ga ima tudi vsota. Pred vsoto nastopa še minus, kar pomeni da imamo enoličen maksimum. Za izračun cenilke $\mu$ lahko izpustimo minus in iščemo minimum vsote. Vsoto lahko še nekoliko preoblikujemo:

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(x_i,\ y_i - \mu\right)^T \Psi\left(x_i,\ y_i - \mu\right) \\
&= \sum_{i=1}^{N}\left(x_i - \bar{x} + \bar{x},\ y_i - \bar{y} - (\mu - \bar{y})\right)^T \Psi\left(x_i - \bar{x} + \bar{x},\ y_i - \bar{y} - (\mu - \bar{y})\right) \\
&= \sum_{i=1}^{N}(\left(x_i - \bar{x}, y_i - \bar{y}\right)^T \Psi\left(x_i - \bar{x}, y_i - \bar{y}\right) - 2\left(x_i - \bar{x}, y_i - \bar{y}\right)^T \Psi\left(-\bar{x}, \mu - \bar{y}\right) \\
&+ \left(-\bar{x}, \mu - \bar{y}\right)^T \Psi\left(-\bar{x}, \mu - \bar{y}\right)) \\
&= \sum_{i=1}^{N}\Big(\left(x_i - \bar{x}, y_i - \bar{y}\right)^T \Psi\left(x_i - \bar{x}, y_i - \bar{y}\right)\Big) + N\left(\bar{x}, \bar{y} - \mu\right)^T \Psi\left(\bar{x}, \bar{y} - \mu\right).
\end{aligned}
$$

Ker v vsoti $\mu$ ne nastopa, jo zanemarimo, prav tako na minimum $N$ ne vpliva in ga zato zanemarimo. Če razpišemo izraz dobimo:

$$
\begin{aligned}
&\left(\bar{x}, \bar{y} - \mu\right)^T \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12}^T & \Psi_{22} \end{pmatrix} \left(\bar{x}, \bar{y} - \mu\right) \\
&= (\bar{x}^T\Psi_{11}\bar{x} + (\bar{y} - \mu)^T\Psi_{12}^T\bar{x} + \bar{x}^T\Psi_{12}(\bar{y} - \mu) + (\bar{y} - \mu)^T\Psi_{22}(\bar{y} - \mu)) \\
&= (\bar{x}^T\Psi_{11}\bar{x} + (\bar{y} - \mu)^T\Psi_{22}\Psi_{22}^{-1}\Psi_{12}^T\bar{x} + \bar{x}^T\Psi_{12}(\bar{y} - \mu) + (\bar{y} - \mu)^T\Psi_{22}(\bar{y} - \mu) \\
&+ \bar{x}^T\Psi_{12}\Psi_{22}^{-1}\Psi_{12}^T\bar{x} - \bar{x}^T\Psi_{12}\Psi_{22}^{-1}\Psi_{12}^T\bar{x}) \\
&= ((\bar{y} - \mu)^T + \bar{x}^T\Psi_{12}\Psi_{22}^{-1})\Psi_{22}((\bar{y} - \mu)^T + \bar{x}^T\Psi_{12}\Psi_{22}^{-1})^T + \bar{x}^T\Psi_{11}\bar{x} - \bar{x}^T\Psi_{12}\Psi_{22}^{-1}\Psi_{12}^T\bar{x} \\
&= ((\bar{y} - \mu)^T + \bar{x}^T\Psi_{12}\Psi_{22}^{-1})\Psi_{22}((\bar{y} - \mu)^T + \bar{x}^T\Psi_{12}\Psi_{22}^{-1})^T + \bar{x}^T\Psi_{11 \cdot 2}\bar{x}.
\end{aligned}
$$

Za fiksen $\Psi$ minimiziramo glede na $\mu$. Prvi člen je nenegativen za vsak $\mu$, kar pomeni, da bo najmanjši, kadar je enak nič. V drugem členu pa $\mu$ ne nastopa in ne vpliva na ekstrem. Očitno je cenilka enaka $\hat{\mu} = \bar{y}^T + \bar{x}^T\Psi_{12}\Psi_{22}^{-1}$, če cenilko upoštevamo v funkciji verjetja dobimo:

$$\begin{aligned}
l(\mu,\Sigma;z) &= -\frac{(k+l)N}{2}\log 2\pi - \frac{N}{2}\log\left(\det\Sigma\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{N}\Big(\left(x_i-\bar{x},y_i-\bar{y}\right)^T\Psi\left(x_i-\bar{x},y_i-\bar{y}\right)\Big) - \frac{N}{2}\bar{x}^T\Psi_{11\cdot 2}\bar{x} \\
&= -\frac{(k+l)N}{2}\log 2\pi - \frac{N}{2}g(\Psi),
\end{aligned}$$

kjer je:

$$\begin{aligned}
g\left(\Psi\right) &= \log\left(\det\Psi\right) - tr\left(\Psi\sum_{i=1}^{N}\frac{1}{N}\left(x_i,y_i-\bar{y}\right)\left(x_i,y_i-\bar{y}\right)^T\right) - \bar{x}^T\Psi_{11\cdot 2}\bar{x} \\
&= \log\left(\det\Psi\right) - tr\left(\Psi V\right) - tr\left(\Psi_{11\cdot 2}\bar{x}\bar{x}^T\right).
\end{aligned}$$

3.2. **Vzorec z dodatnimi podatki.** Drug primer, ki ga bomo obravnavali, je primer, ko imamo vzorec velikosti $N$ za $p$ razsežno normalno porazdelitev s kovariančno matriko $\Sigma$ in vektorjem matematičnega upanja $(\mu,\nu)$, kjer je $\mu$ k-razsežen in $\nu$ (p-k)-razsežen. Imamo še vzorec velikosti $M$, ki pa ima samo $k$ opažanj na isti $p$-razsežni porazdelitvi. Torej imamo vzorec velikosti $N+M$ na $p$-razsežni normalni porazdelitvi, kjer zadnjih $(p-k)$ opažanj manjka pri zadnjih $M$ podatkih iz vzorca. Naj bo $(\bar{x},\bar{y})$ vzorčno upanje na vzorcu $N$, kjer je $\bar{x}$ dimenzije $k$ in $\bar{y}$ dimenzije $(p-k)$. Vzorčno upanje vzorca $M$, kjer imamo samo prvih $k$ opažanj, je enako $\bar{z}$. Na takem vzorcu lahko zapišemo logaritem funkcije verjetja:

$$\begin{aligned}
&-\frac{pN}{2}\log\left(2\pi\right) - \frac{N}{2}\log\left(\det\Sigma\right) - \frac{1}{2}\sum_{i=1}^{N}\left(x_i-\mu,y_i-\nu\right)^T\Sigma^{-1}\left(x_i-\mu,y_i-\nu\right) \\
&\qquad -\frac{kM}{2}\log\left(2\pi\right) - \frac{M}{2}\log\left(\det\Sigma_{11}\right) - \frac{1}{2}\sum_{i=1}^{M}\left(z_i-\mu\right)^T\Sigma_{11}^{-1}\left(z_i-\mu\right).
\end{aligned}$$

Po istem premisleku kot v prejšnem primeru vidimo, da je za izračun cenilke matematičnega upanja dovolj, če minimiziramo sledeč izraz:

$$N(\bar{x}-\mu,\bar{y}-\nu)^T\Sigma^{-1}(\bar{x}-\mu,\bar{y}-\nu) + M(\bar{z}-\mu)^T\Sigma_{11}^{-1}(\bar{z}-\mu). \tag{4}$$

Za izračun cenilke $\mu$ imamo sledeč problem:

$$\min_{\mu}\left\{N(\bar{x}-\mu)^T\Sigma_{11}^{-1}(\bar{x}-\mu) + M(\bar{z}-\mu)^T\Sigma_{11}^{-1}(\bar{z}-\mu)\right\}. \tag{5}$$

Ker je $\Sigma$ pozitivno definitna, je potem tudi $\Sigma_{11}^{-1}$ in kvadratni formi iz (5) sta podobni (2), kar pomeni, da imamo enoličen minimum. Z odvajanjem, dobimo:

$$\begin{aligned}
&-2N\left(\bar{x}-\hat{\mu}\right)^T\Sigma_{11}^{-1} - 2M\left(\bar{x}-\hat{\mu}\right)^T\Sigma_{11}^{-1} = 0 \quad / \ \Sigma_{11} \\
&N\left(\bar{x}-\hat{\mu}\right)^T + M\left(\bar{x}-\hat{\mu}\right)^T = 0 \\
&\hat{\mu} = \frac{1}{N+M}\left(N\bar{x}+M\bar{z}\right).
\end{aligned}$$

Za izračun cenilke $\nu$ moramo minimizirati samo levi člen izraza (4). Ta je nenegativen in bo najmanjši, ko bo enak 0. Če razpišemo ta člen, kot smo to naredili v prejšnem razdelku, dobimo spodnjo enačbo:

$$N((\bar{y}-\nu)^T + (\bar{x}-\hat{\mu})^T\Psi_{12}\Psi_{22}^{-1})\Psi_{22}((\bar{y}-\nu)^T + (\bar{x}-\hat{\mu})^T\Psi_{12}\Psi_{22}^{-1})^T = 0.$$

Vidimo, da je cenilka za $\nu$ enaka $\hat{\nu} = \bar{y}^T + (\bar{x}-\hat{\mu})^T\Psi_{12}\Psi_{22}^{-1} = \bar{y}^T + \frac{M}{N+M}(\bar{x}-\bar{z})^T\Psi_{12}\Psi_{22}^{-1}$. Če izračunani cenilki vključimo v funkcijo verjetja, dobimo:

$$N\log(\det\Psi) - Ntr\Psi V + M\log(\det\Psi_{11\cdot 2}) - Mtr\Psi_{11\cdot 2}W,$$

kjer je:

$$W = \frac{1}{M}\sum_{i=1}^{M}\left((z_i-\bar{z})(z_i-\bar{z})^T\right) + \frac{N}{N+M}(\bar{x}-\bar{z})(\bar{x}-\bar{z})^T.$$

3.3. **Izračun cenilke kovariančne matrike.** Za izračun cenilke za kovariančno matriko bi lahko zopet uporabili transformacijski izrek, vendar se nam v tem primeru ohrani matrična funkcija in moramo še vedno matrično odvajati funkcijo. Zato je enostavneje, če direktno odvajamo funkcijo $f(\Sigma)$ oz. $g(\Psi)$. Premislek o obstoju in enoličnosti maksimuma je enak kot za funkcijo, ki jo dobimo na vzorcu brez manjkajočih podatkov. Z uporabo izrekov za odvajanje matrik dobimo:

$$\begin{aligned}\frac{\partial}{d\sigma_{ij}}f(\Sigma) = (2-\delta_{ij})\{&-N\frac{\Sigma_{ij}}{\det\Sigma}d\sigma_{ij} - M\frac{A_{ij}}{\det A}da_{ij}\epsilon_{ij}\\ &+ Ntr\left(\Sigma^{-1}E_{ij}\Sigma^{-1}V\right)d\sigma_{ij} + Mtr\left(\Sigma_{11}^{-1}E_{ij}\Sigma_{11}^{-1}W\right)da_{ij}\epsilon_{ij}\} = 0\end{aligned}$$

Zaradi poenostavitve smo zapisali $A = (a_{ij})_{k\times k} \equiv \Sigma_{11}$, $\Sigma_{ij}$ in $A_{ij}$ sta kofaktorja matrik $\Sigma$ in $A$ za $i,j = 1,\ldots,p$ in $\epsilon_{ij}$ pa je 1, če velja $i,j \leq k$ in 0 sicer. Enačbo lahko zapišemo v matrični obliki:

$$-N\Sigma^{-1} + N\Sigma^{-1}V\Sigma^{-1} - M\begin{pmatrix}\Sigma_{11}^{-1} & 0\\ 0 & 0\end{pmatrix} + M\begin{pmatrix}\Sigma_{11}^{-1}W\Sigma_{11}^{-1} & 0\\ 0 & 0\end{pmatrix} = 0,$$

z množenjem z leve in desne z $\Sigma$ dobimo:

$$\begin{aligned}-N\Sigma + NV - M\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}\begin{pmatrix}\Sigma_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix} +\\ + M\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}\begin{pmatrix}\Sigma_{11}^{-1}W\Sigma_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix} = 0,\end{aligned}$$

nato še zmnožimo bločne matrike med sabo:

$$-N\Sigma + NV - M\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix} + M\begin{pmatrix}W & W\Sigma_{11}^{-1}\Sigma_{12}\\ \Sigma_{12}^T\Sigma_{11}^{-1}W & \Sigma_{12}^T\Sigma_{11}^{-1}W\Sigma_{11}^{-1}\Sigma_{12}\end{pmatrix} = 0.$$

Matrično enačbo lahko zapišemo kot sistem enačb:

$$
\begin{aligned}
-N\Sigma_{11} + NV_{11} - M\Sigma_{11} + MW &= 0, \\
-N\Sigma_{12} + NV_{12} - M\Sigma_{12} + MW\Sigma_{11}^{-1}\Sigma_{12} &= 0, \\
-N\Sigma_{12}^T + NV_{12}^T - M\Sigma_{12}^T + M\Sigma_{12}^T\Sigma_{11}^{-1}W &= 0, \\
-N\Sigma_{22} + NV_{22} - M\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} + M\Sigma_{12}^T\Sigma_{11}^{-1}W\Sigma_{11}^{-1}\Sigma_{12} &= 0.
\end{aligned}
$$

Iz prve enačbe lahko izračunamo cenilko za $\Sigma_{11}$:

$$
\begin{aligned}
\hat{\Sigma}_{11}(N+M) &= NV_{11} + MW \\
\hat{\Sigma}_{11} &= \frac{1}{N+M}(NV_{11} + MW).
\end{aligned}
$$

Cenilko za $\Sigma_{12}$ izrazimo iz druge enačbe tako, da zamenjamo $\Sigma_{11}^{-1}$ z $V_{11}^{-1}$ in v zadnjem členu enačbe zamenjamo $\Sigma_{12}$ z $V_{12}$ in tako dobimo:

$$
\begin{aligned}
&-N\hat{\Sigma}_{12} + NV_{12} - M\hat{\Sigma}_{12} + MWV_{11}^{-1}V_{12} = 0 \\
&(N+M)\hat{\Sigma}_{12} = NV_{12}^T + MWV_{11}^{-1}V_{12} \\
&\hat{\Sigma}_{12} = \frac{1}{N+M}(NV_{11} + MW)V_{11}^{-1}V_{12}.
\end{aligned}
$$

Ker, je $\Sigma_{21} = \Sigma_{12}^T$ je tudi cenilka za $\Sigma_{21}$ enaka transponirani cenilki za $\Sigma_{12}$. Cenilko za $\Sigma_{22}$ dobimo iz zadnje enačbe tako, da zamenjamo $\Sigma_{12}$ z $V_{12}$ in $\Sigma_{11}^{-1}$ z $V_{11}^{-1}$:

$$
\begin{aligned}
&-N\hat{\Sigma}_{22} + NV_{22} - MV_{12}^TV_{11}^{-1}V_{12} + MV_{12}^TV_{11}^{-1}WV_{11}^{-1}V_{12} = 0 \\
&-(N+M)\hat{\Sigma}_{22} + M\Sigma_{22} + NV_{22} - M\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12} + M\Sigma_{12}^T\Sigma_{11}^{-1}W\Sigma_{11}^{-1}\Sigma_{12} = 0 \\
&(N+M)\hat{\Sigma}_{22} = M(V_{22} - V_{12}^TV_{11}^{-1}V_{12}) + NV_{22} + MV_{12}^TV_{11}^{-1}WV_{11}^{-1}V_{12} \\
&\hat{\Sigma}_{22} = \frac{1}{N+M}\left(MV_{22\cdot 1} + NV_{22} + MV_{12}^TV_{11}^{-1}WV_{11}^{-1}V_{12}\right).
\end{aligned}
$$

## 4. Popoln monoton vzorec podatkov

Poseben primer vzorca z manjkajočimi podatki je popoln monoton vzorec. Če imamo zaporedje slučajnih vektorjev $X = \{X_i = (X_{i1}, \ldots, X_{ip})^T; i = 1, \ldots, n\}$, iz katerega izhaja vzorec, ki ga lahko uredimo v $p$ podvzorcev tako, da v prvem podvzorcu ni manjkajočih podatkov, v drugem podvzorcu imamo v opažanjih en manjkajoč podatek, v tretjem podvzorcu dva manjkajoča podatka in $k$-tem podvzorcu $k-1$ manjkajočih prvih podatkov. V tem primeru rečemo, da imamo popoln monoton vzorec:

$$
\begin{array}{rcl}
x_{i1}^{(1)}, \ldots, x_{ik}^{(1)}, \ldots, x_{ip}^{(1)} & za & i = 1, \ldots, n_1; \\
\ddots & \vdots & \vdots \\
x_{ik}^{(k)}, \ldots, x_{ip}^{(k)} & za & i = 1, \ldots, n_k; \\
\ddots & \vdots & \vdots \\
x_{ip}^{(p)} & za & i = 1, \ldots, n_p
\end{array} \tag{6}
$$

Za $p$-razsežno normalno porazdelitev $X \sim N_p(\mu, \Sigma)$, kjer sta $\mu = (\mu_1, \ldots, \mu_p)^T$ vektor matematičnega upanja in $\Sigma = (\sigma_{ij})_{(p\times p)} > 0$ kovariančna matrika, lahko za $k = 1, \ldots, p$ zapišemo $(p-k+1)$-razsežno normalno porazdelitev $N_{p-k+1}(\mu^{(k)}, \Sigma^{(k)})$, kjer sta $\mu^{(k)}$ podvektor vektorja matematičnega upanja $\mu$ in $\Sigma^{(k)}$ desna spodnja podmatrika matrike $\Sigma$ oblike

$$\mu^{(k)} = \begin{bmatrix} \mu_k \\ \vdots \\ \mu_p \end{bmatrix}, \quad \Sigma^{(k)} = \begin{bmatrix} \sigma_{kk} & \ldots & \sigma_{kp} \\ \vdots & \ddots & \vdots \\ \sigma_{pk} & \ldots & \sigma_{pp} \end{bmatrix}. \tag{7}$$

Zaporedje slučajnih vektorjev, ki je oblike popolnega monotonega vzorca in ustreza $p$-razsežni normalni porazdelitvi, je za $k = 1, \ldots, p$ in $i = 1, \ldots, n_k$ enak $X_i^{(k)} \sim N_{p-k+1}(\mu^{(k)}, \Sigma^{(k)})$. Za vzorec oblike $\{x_i^{(j)} = (x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T : i = 1, \ldots, n_j;$ $j = 1, \ldots, k\}$, ki izhaja iz opisanega zaporedja slučajnih vektorjev, lahko definiramo za vsak $k = 1, \ldots, p$ vzorčno upanje:

$$N_k = \sum_{j=1}^{k} n_j,$$

$$\bar{x}_k = \frac{1}{N_k} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T.$$

Za vsak $k = 1, \ldots, p$ definiramo matriki $S_k$ in $R_k$

$$S_k = \sum_{j=1}^{k} \sum_{i=1}^{n_j} ((x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T - \bar{x}_k)((x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T - \bar{x}_k)^T,$$

$$R_k = \sum_{j=1}^{k} \sum_{i=1}^{n_j} ((x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T - \mu^{(k)})((x_{ik}^{(j)}, \ldots, x_{ip}^{(j)})^T - \mu^{(k)})^T.$$

$R_k$ lahko zapišemo tudi v drugačni obliki. Če v vsakem vektorju, ki tvorita skalarni produkt prištejemo in odštejemo $\bar{x}_k$ dobimo

$$\begin{aligned} R_k &= \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_i^{(k)} - \bar{x}_k - (\mu^{(k)} - \bar{x}_k))((x_i^{(k)} - \bar{x}_k - (\mu^{(k)} - \bar{x}_k))^T \\ &= \sum_{j=1}^{k} \sum_{i=1}^{n_j} ((x_i^{(k)} - \bar{x}_k)(x_i^{(k)} - \bar{x}_k)^T - 2(x_i^{(k)} - \bar{x}_k)(\mu^{(k)} - \bar{x}_k)^T \\ &+ (\mu^{(k)} - \bar{x}_k)(\mu^{(k)} - \bar{x}_k)^T) \\ &= S_k + N_k(\mu^{(k)} - \bar{x}_k)(\mu^{(k)} - \bar{x}_k)^T \end{aligned}$$

Za inverz kovariančne matrike $\Sigma^{-1}$ obstaja razcep Choleskega $\Sigma^{-1} = HH^T$, kjer je $H$ spodnje trikotna matrika in $h_k$ k-ti stolpec matrike $H$, v katerem prvih $(k-1)$ elementov manjka. Za popoln monoton vzorec na $p$-razsežni normalni porazdelitvi, je logaritem funkcije verjetja enak:

$$
\begin{aligned}
l(\mu,\Sigma;x) &= \sum_{k=1}^{p}(-\frac{n_k}{2}\log(\det\Sigma^{(k)}) - \frac{1}{2}\sum_{i=1}^{n_k}(x_i^{(k)}-\mu^{(k)})^T(\Sigma^{(k)})^{-1}(x_i^{(k)}-\mu^{(k)}))\\
&= \sum_{k=1}^{p}(-\frac{n_k}{2}\log(\det\Sigma^{(k)}) - \frac{1}{2}\sum_{i=1}^{n_k}tr((x_i^{(k)}-\mu^{(k)})^T(\Sigma^{(k)})^{-1}(x_i^{(k)}-\mu^{(k)})))\\
&= \sum_{k=1}^{p}(-\frac{n_k}{2}\log(\det\Sigma^{(k)}) - \frac{1}{2}tr(\sum_{i=1}^{n_k}(x_i^{(k)}-\mu^{(k)})(x_i^{(k)}-\mu^{(k)})^T(\Sigma^{(k)})^{-1}))\\
&= \sum_{k=1}^{p}(-\frac{n_k}{2}\log(\det\Sigma^{(k)}) - \frac{1}{2}tr((\Sigma^{(k)})^{-1}R^{(k)}))
\end{aligned}
$$

Zanemarili smo konstanto, ki ne vpliva na ekstrem in upoštevali, da je sled matrike ciklična. Za nadaljno izpeljavo bomo potrebovali Liu-jevo lemo, ki je povzeta po [5, str.201, Lema 1].

**Lema 4.1. Liu-jeva lema** *Naj bo $\Sigma^{-1} = HH^T$ razcep Choleskega, kjer je $H$ spodnje trikotna matrika, potem je $(\Sigma^{(k)})^{-1} = H^{(k)}(H^{(k)})^T$, kjer je $H^{(k)}$ $(p-k+1)\times(p-k+1)$ spodnja desna podmatrika $H$.*

Če upoštevamo razcep Choleskega za matriko $\Sigma^{-1}$ in uporabimo Liu-jevo lemo dobimo:

$$
\begin{aligned}
l(\mu,H;x) &= \frac{1}{2}\sum_{k=1}^{p} n_k \log\det(H^{(k)}H^{(k)^T}) - \frac{1}{2}\sum_{k=1}^{p} tr(H^{(k)}H^{(k)^T}R^{(k)})\\
&= \frac{1}{2}\sum_{k=1}^{p} n_k \log(\prod_{i=1}^{p-k+1} h_{ii}^2) - \frac{1}{2}\sum_{k=1}^{p} tr(H^{(k)}R^{(k)}H^{(k)^T})\\
&= \sum_{k=1}^{p}\sum_{i=1}^{p-k+1} n_k\log h_{ii} - \frac{1}{2}\sum_{k=1}^{p} {h_k}^T R_k h_k\\
&= \sum_{k=1}^{p} N_k \log h_{kk} - \frac{1}{2}\sum_{k=1}^{p} h_k^T S_k h_k - \frac{1}{2}\sum_{k=1}^{p} h_k^T N_k(\mu_k-\bar{x}_k)(\mu_k-\bar{x}_k)^T h_k.
\end{aligned}
$$

Zapišimo matriko $D = \big(HDiag(N_1,\ldots,N_p)H^T\big)^{-1}$ in vektor $\theta = (H^T)^{-1}(h_1^T\bar{x}_1,\ldots,h_p^T\bar{x}_p)^T$. Razpišimo zadnji člen zgornje funkcije in upoštevajmo ravnokar zapisano matriko in vektor:

$$
\begin{aligned}
l(\mu,H;x) =& -\frac{1}{2}\sum_{k=1}^{p} N_k(\mu_k h_k^T h_k\mu_k^T - \mu_k h_k^T h_k\bar{x}_k^T - h_k^T\bar{x}_k h_k\mu_k^T + h_k^T\bar{x}_k h_k\bar{x}_k^T)\\
=& -\frac{1}{2}(\mu HDiag(N_1,\ldots,N_p)H^T\mu\\
&-\mu HDiag(N_1,\ldots,N_p)(h_1^T\bar{x}_1,\ldots,h_p^T\bar{x}_p)^T\\
&-(h_1^T\bar{x}_1,\ldots,h_p^T\bar{x}_p)Diag(N_1,\ldots,N_p)H^T\mu\\
&+(h_1^T\bar{x}_1,\ldots,h_p^T\bar{x}_p)Diag(N_1,\ldots,N_p))(h_1^T\bar{x}_1,\ldots,h_p^T\bar{x}_p)^T)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}(\mu - (H^T)^{-1}(h_1^T\bar{x}_1, \ldots, h_p^T\bar{x}_p)^T)^T(HDiag(N_1, \ldots, N_p)H^T) \\
&\cdot(\mu - (H^T)^{-1}(h_1^T\bar{x}_1, \ldots, h_p^T\bar{x}_p)^T) \\
&= -\frac{1}{2}(\mu - \theta)^T D^{-1}(\mu - \theta).
\end{aligned}
$$

Za nadaljno izpeljavo potrebujemo razcep Choleskega $S_k^{-1} = C_k C_k^T$. Definirajmo še vektor $t_k = (t_{kk}, \ldots, t_{kp})^T = C_k^{-1}h_k$ za $k = 1, \ldots, p$ in matriko $T = (t_1, \ldots, t_p)$, če ju upoštevamo v funkciji verjetja skupaj z zgornjim izračunom dobimo:

$$
\begin{aligned}
l(\mu, H; x) &= \sum_{k=1}^{p} N_k \log h_{kk} - \frac{1}{2}\sum_{k=1}^{p} h_k^T \left(C_k^T\right)^{-1} C_k^{-1} h_k - \frac{1}{2}(\mu - \theta)^T D^{-1}(\mu - \theta) \\
&= \sum_{k=1}^{p} N_k \log t_{kk} - \frac{1}{2}\sum_{k=1}^{p} t_k^T t_k - \frac{1}{2}(\mu - \theta)^T D^{-1}(\mu - \theta).
\end{aligned}
$$

Iz te funkcije želimo izračunati cenilko za $\mu$ in $\Sigma$. Za izračun cenilke $\Sigma$ moramo izračunati cenilko za matriko $H$ iz razcepa Choleskega. Cenilko za $H$ dobimo iz stolpcev matrike $T$, saj v njih nastopajo stolpci matrike $H$. Če odvajamo zgornjo funkcijo po matriki $T$, bomo dobili maksimum funkcije in ta je enoličen. Premislek je enak, kot v prejšnih primerih. Z odvajanjem logaritma funkcije verjetja glede na matriko $T$ dobimo:

$$
\begin{aligned}
\frac{\partial l(\mu, H; x)}{\partial T} &= \sum_{k=1}^{p} \frac{N_k}{\hat{t}_{k,k}}(1, 0, \ldots, 0) - \frac{1}{2}\sum_{k=1}^{p} 2\hat{t}_k = 0 \\
\sum_{k=1}^{p} \begin{bmatrix} \hat{t}_{k,k}^2 \\ \hat{t}_{k,k} \cdot \hat{t}_{k,k+1} \\ \vdots \\ \hat{t}_{k,k} \cdot \hat{t}_{k,p} \end{bmatrix} &= \sum_{k=1}^{p} \begin{bmatrix} N_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}.
\end{aligned}
$$

Iz česar vidimo:

$$
\begin{aligned}
&\hat{t}_k = (\sqrt{N_k}, 0, \ldots, 0)^T \quad za \quad k = 1, \ldots, p, \\
&\hat{h}_k = C_k(\sqrt{N_k}, 0 \ldots, 0)^T \quad za \quad k = 1, \ldots, p, \\
&\hat{\Sigma} = (\hat{H}\hat{H}^T)^{-1}.
\end{aligned} \tag{8}
$$

Za izračun cenilke $\mu$ odvajamo samo zadnji člen funkcije, ker samo tam nastopa $\mu$. Z odvajanjem bomo dobili maksimum, saj je $D$ produkt treh pozitivno definitnih matrik in je zato tudi D pozitivno definitna matrika. Vemo, da imajo funkcije oblike (2) enoličen minimum, zaradi minusa pa bomo dobili maksimum. Z odvajanjem funkcije glede na $\mu$ dobimo:

$$
\begin{aligned}
\frac{\partial l(\mu, H; Y)}{\partial \mu} &= (\hat{\mu} - \theta)D^{-1} = 0 \\
&\hat{\mu}D^{-1} = \theta D^{-1} \\
&\hat{\mu} = (\hat{H}^T)^{-1}(\hat{h}_1\bar{x}_1, \ldots, \hat{h}_p\bar{x}_p)^T.
\end{aligned}
$$

Tako smo izpeljali cenilki za $\mu$ in $\Sigma$ na popolnem monotonem vzorcu. Če nimamo popolnega monotonega vzorca, ga lahko dopolnimo s pomočjo dela vzorca, kjer nimamo manjkajočih podatkov, do popolnega monotonega vzorca. Vzorec dopolnimo z EM algoritmom, ki pa ga ne obravnavamo.

4.1. **Implementacija.** Primer ocenjevanja parametrov večrazsežne normalne porazdelitve na popolnem monotonem vzorcu naredimo v Mathematici. Ker nimamo dejanskih podatkov, jih zgeneriramo sami. S spremenljivko $p$ označimo dimenzijo normalne porazdelitve. Pozitivno definitno kovariančno matriko zgeneriramo z uporabo transformacije (2.9), ki nam pove, da se da pozitivno definitno matriko razcepiti kot $\Sigma = Q^T DQ$, kjer je $Q$ neka ortogonalna matrika, $D$ pa diagonalna matrika z lastnimi vrednostmi $\Sigma$ na diagonali. Torej, če nastavimo $D = I$ in zmnožimo naključno matriko z njeno transponirano matriko, bomo dobili neko pozitivno definitno matriko. Ustvarimo še naključen vektor matematičnih upanj, velikost vzorca in velikosti podvzorcev. Za ustvarjen vektor matematičnih upanj in kovariančno matriko ustvarimo vzorec določene velikosti.

```
p = 2;
q = Table[RandomReal[{-2, 2}], {p}, {p}];
kovarianca = Transpose[q].q;
upanje = Table[RandomReal[{2, 10}], p];
velikosti = RandomInteger[100, p];
n = Total[velikosti];
vzorec = RandomVariate[
   MultinormalDistribution[upanje, kovarianca], n];
```

Dobljeni vzorec je sicer vzorec brez manjkajočih podatkov, vendar ga obravnavamo kot da je popoln monoton vzorec. To naredimo tako, da pri računanju na vsakem podvzorcu zanemarimo en podatek več. Na tak način izračunamo vzorčna upanja $\bar{x}_k$ in matrike $S_k$ za vsak podvzorec.

```
precne = {};
For[k = 1, k <= p, k++,
 delno = Total[velikosti[[1 ;; k]]];
 precne = Append[precne, Total[vzorec[[1 ;;
  Total[velikosti[[1 ;; k]]], k ;;]]]/delno]]
S = {};
For[k = 1, k <= p, k++,
 t = ConstantArray[0, {p - k + 1, p - k + 1}];
 For[l = 1, l <= p - k + 1, l++,
  For[v = l, v <= p - k + 1, v++,
   t[[l, v]] = Sum[(vzorec[[i, k + l - 1]] - precne[[k, l]])
      *(vzorec[[i, k + v - 1]] - precne[[k, v]]),
     {i, 1, Total[velikosti[[1 ;; k]]]}];
   t[[v, l]] = t[[l, v]]]];
 S = Append[S, t]]
```

Za izračun razcepa Choleskega matrike $S_k^{-1}$ za $k = 1, \dots, p$ bi lahko najprej izračunali inverz matrike $S_k$ in nato izračunali razcep Choleskega $S_k^{-1} = C_k C_k^T$. Vendar je numerično stabilneje, če najprej izračunamo razcep Choleskega $S_k = B_k B_k^T$

in nato izračunamo inverz razcepa $(B_k^T)^{-1} = C_k$. Ta postopek deluje, ker velja $S_k^{-1} = (B_k^T)^{-1}B_k^{-1} = C_kC_k^T$. Za izračun razcepa Choleskega bi lahko uporabili že vgrajeno funkcijo, vendar smo natančnejši, če uporabimo spodaj zapisani algoritem, ki deluje tako, da izračunamo vrstice matrike od spodaj navzgor in vsako vrstico tako, da najprej izračunamo diagonalni element, nato pa se pomikamo od prvega elementa vrstice proti diagonalnemu:

$$b_{ii} = (s_{ii} - \textstyle\sum_{j=i+1}^{p} b_{ji}^2)^{1/2},$$
$$b_{ij} = (s_{ij} - \textstyle\sum_{l=i+1}^{p} b_{li}b_{lj})/b_{ii}, \quad i = p-k+1,\dots,1, \quad j = 1,\dots,i-1.$$

Za izračun inverza uporabimo že vgrajeno funkcijo.

```
B = {};
For[k = 1, k <= p, k++,
 b = ConstantArray[0, {p - k + 1, p - k + 1}];
 For[i = p - k + 1, i >= 1, i--,
  b[[i, i]] = Sqrt[S[[k]][[i, i]] -
      Sum[(b[[j, i]])^2, {j, i + 1, p - k + 1}]];
  For[v = 1, v <= i - 1, v++,
   b[[i, v]] = (S[[k]][[i, v]] - Sum[b[[l, i]]*b[[l, v]],
    {l, i + 1, p - k + 1}])/b[[i, i]]]];
B = Append[B, b]]
inverz = {};
For[k = 1, k <= p, k++, inv = Inverse[B[[k]]];
inverz = Append[inverz, inv]]
```

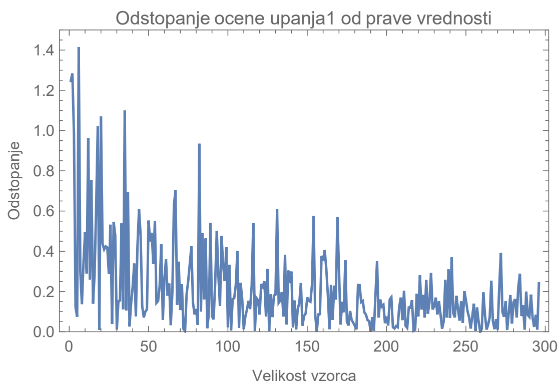
Sedaj lahko ocenimo $\Sigma^{-1}$ tako, da izračunamo oceno za matriko $H$ po postopku opisanem v (8), ki izhaja iz razcepa Choleskega $\Sigma^{-1} = HH^T$. Oceno $\Sigma$ bi lahko direktno izračunali iz njenega inverza, vendar je, kot že vemo, numerično stabilneje, če izračunamo inverz matrike H in njene transponirane matrike in ju nato zmnožimo med sabo. Torej oceno dobimo po enačbi $\Sigma = (H^T)^{-1}H^{-1}$.

```
H = ConstantArray[0, {p, p}];
For[k = 1, k <= p, k++,
 vektor = ConstantArray[0, p - k + 1];
 vektor[[1]] = Sqrt[Total[velikosti[[1 ;; k]]]];
 stolpec = inverz[[k]].vektor;
 H[[k ;; p, k]] = stolpec]
cenilka = Inverse[Transpose[H]].Inverse[H];
```

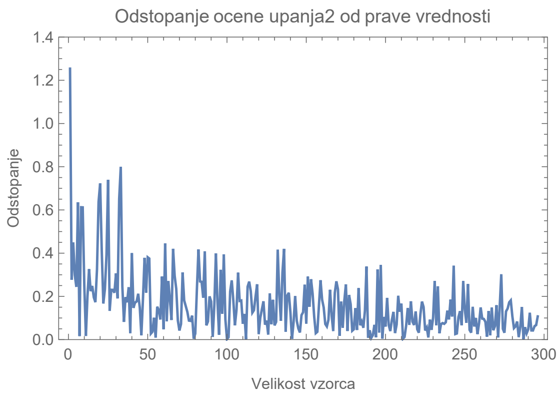
Za konec še ocenimo matematično upanje z uporabo cenilke za matematično upanje. Za razliko od ocene kovariančne matrike tu nimamo opravka z numeričnimi napakami in dobimo s tem postopkom zelo natančno aproksimacijo matematičnega upanja. Če imamo opravka z celimi števili, napak ni, če pa ocenjujemo matematično upanje z realnimi števili, pa so razlike med dejanskimi in ocenjenimi vrednostmi skrivajo na drugem decimalnem mestu.

```
vektor = ConstantArray[0, p];
For[k = 1, k <= p, k++,
 vektor[[k]] = H[[k ;; p, k]].precne[[k]]]
mu = Inverse[Transpose[H]].vektor
```

4.2. **Rezultati.** Če obravnavamo 2-razsežno normalno porazdelitev, vidimo, da se z večanjem vzorca ocena matematičnega upanja vedno bolj približuje pravi vrednosti, kar prikazujeta spodnja grafa. Za majhne vzorce se ocene matematičnega upanja od prave vrednosti razlikujejo za približno 1, z večanjem vzorca pa ta razlika hitro pada. Za vzorec velikosti 300 se ocena od prave vrednosti razlikuje le še za 0.2.

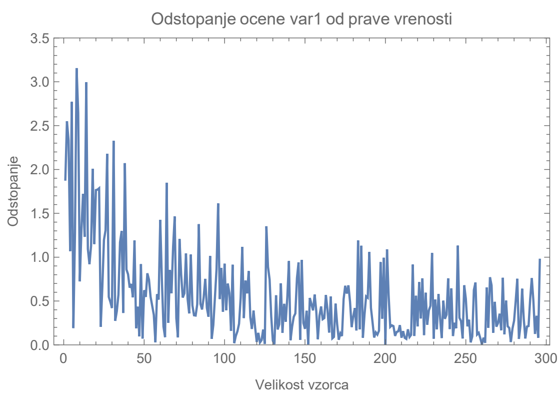



Slika 1. Odstopanje ocene $\mu_1$ od prave vrednosti z večanjem vzorca.

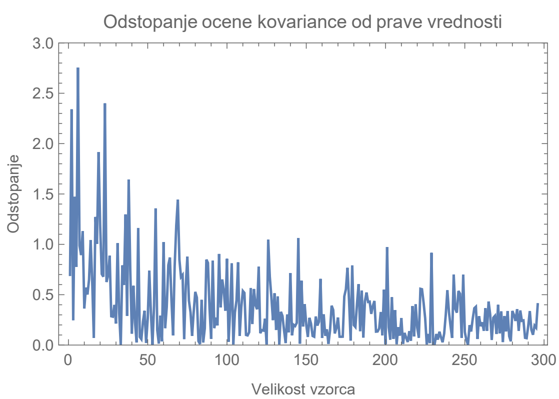



Slika 2. Odstopanje ocene $\mu_2$ od prave vrednosti z večanjem vzorca.

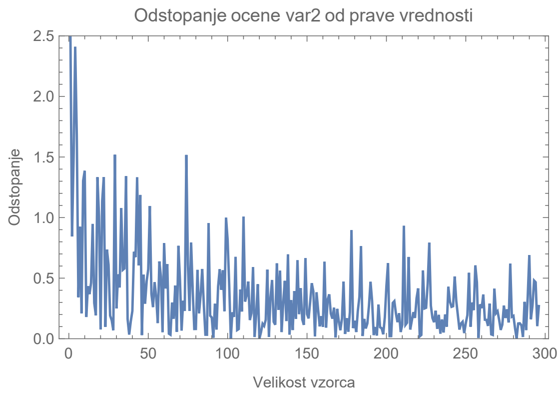
Ocena kovariančne matrike prav tako z večanjem vzorca pada. Sicer so za majhne vzorce odstopanja od prave vrednosti večja, kot so odstopanja ocen matematičnih upanj, vendar se z večanjem vzorca ocena kovariančne matrike prav tako približuje pravi vrednosti. Sicer moramo za zelo natančne ocene obravnavati večje vzorce kot za oceno matematičnega upanja, saj je za vzorec velikosti 100 napaka še vedno približno 1, medtem ko je napaka ocene matematičnega upanja približno 0,3. Opisana opažanja predstavljajo spodnji trije grafi.




Slika 3. Odstopanje ocene $\sigma_1$ od prave vrednosti z večanjem vzorca.

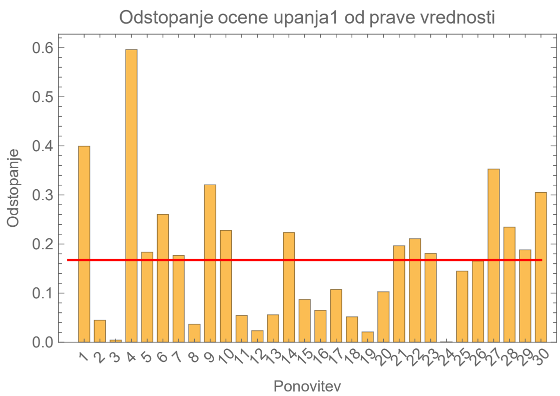



Slika 4. Odstopanje ocene $\sigma_{12}$ od prave vrednost z večanjem vzorca.

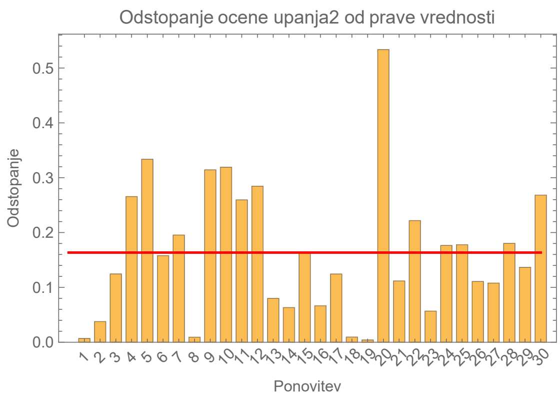



SLIKA 5. Odstopanje ocene $\sigma_2$ od prave vrednosti z večanjem vzorca.

Natančnost ocen matematičnega upanja opazujemo še v povprečju. Na vzorcu velikosti 100 za 30 ponovitev opazimo, da ocena v povprečju odstopa za 0,2 od prave vrednosti. Sicer se pojavljajo tudi nekoliko večja odstopanja, vendar je v povprečju odstopanje $0,2$.

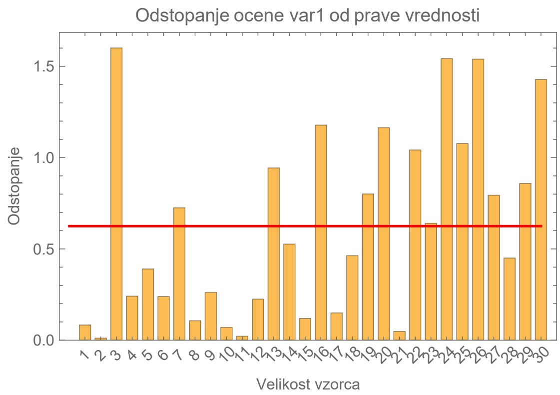



SLIKA 6. Odstopanje ocene $\mu_1$ od prave vrednosti v povprečju.

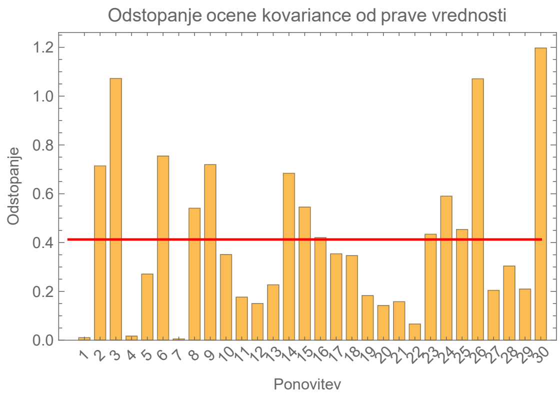



SLIKA 7. Odstopanje ocene $\mu_2$ od prave vrednosti v povprečju.

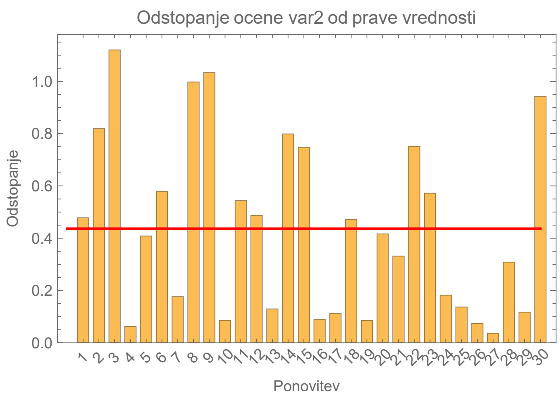
Z obravnavanjem ocene kovariančne matrike na vzorcu velikosti 100 v 30-tih ponovitvah opazimo, da ocenjene vrednosti kovariančne matrike od dejanskih vrednosti odstopajo okrog $0,4$, le ocena variance $\sigma_1$ odstopa nekoliko več, kar je posledica, da imamo za oceno tega dela variančne matrike na razpolago zgolj del vzorca, kjer nimamo manjkajočih podatkov.




SLIKA 8. Odstopanje ocene $\sigma_1$ od prave vrednosti v povprečju.




SLIKA 9. Odstopanje ocene $\sigma_{12}$ od prave vrednosti v povprečju.

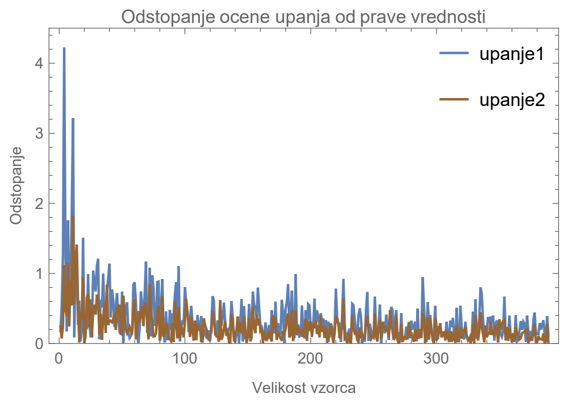



SLIKA 10. Odstopanje ocene $\sigma_2$ od prave vrednosti v povprečju.

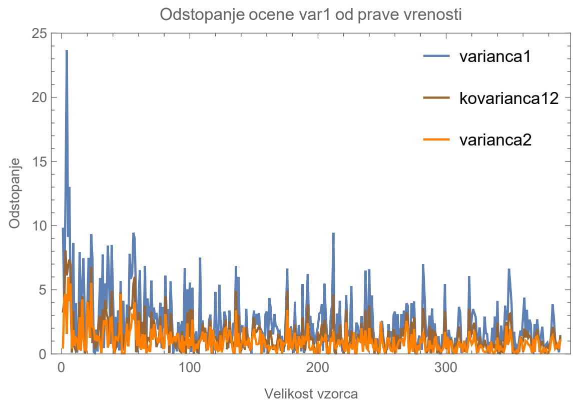
Z večanjem absolutnih vrednosti v kovariančni matriki prihaja do večjih napak pri oceni le te. Predvsem za majhne vzorce so napake velike, z večanjem vzorca dobimo vedno bolj natančne ocene, vendar moramo za njih obravnavati veliko večje vzorce kot za ocene kovariančne matrike, ki nima zelo velikih vrednosti. Prav tako je ocena matematičnega upanja nekoliko slabša, kot pa za oceno, kjer je kovariančna matrika z nizkimi vrednostmi. Obravnavamo 2-razsežno normalno porazdelitev, ki ima za kovariančno matriko spodnjo matriko:

$$\Sigma = \begin{bmatrix} 23,8826 & -14,712 \\ -14,712 & 10,951 \end{bmatrix}.$$

Za majhne vzorce je napaka ocene matemtičnega upanja kar velika, vendar ta napaka z večanjem vzorca hitro pada. Pri ocenjevanju kovariančne matrike so na manjših vzorcih ocene zelo slabe, z večanjem vzorca pa padajo. Za zelo natančne ocene potrebujemo zelo velik vzorec. Za vzorec velikosti 300 se napake gibljejo med 2 in 3. Spodnja grafa prikazujeta opaženo.

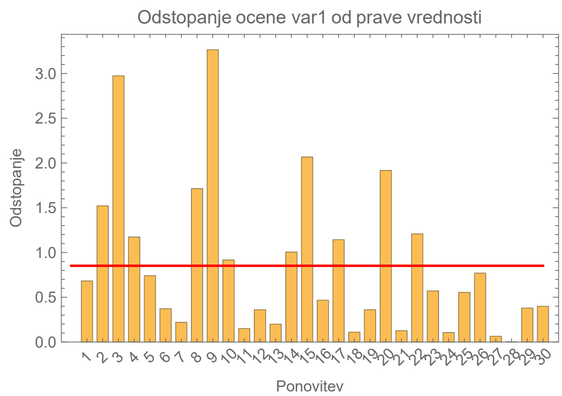



SLIKA 11. Napake ocene upanja od prave vrednosti z večanjem vzorca.

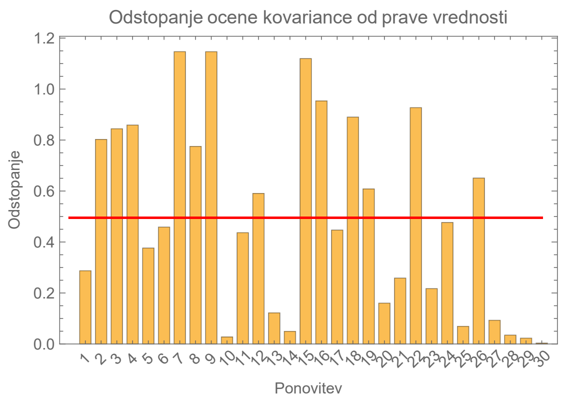



SLIKA 12. Napake ocene kovariančne matrike z večanjem vzorca.

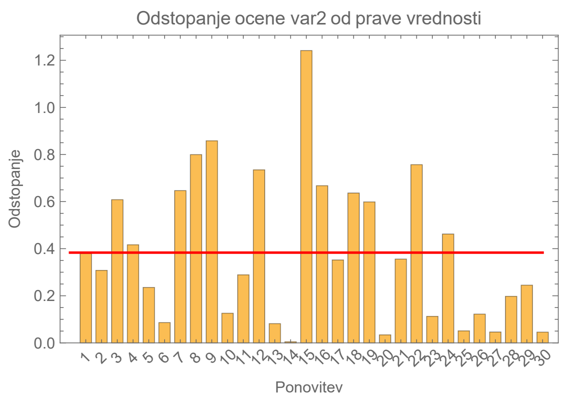
Da so ocene matematičnega upanja v povprečju natančne, moramo obravnavati zelo velik vzorec. Za velikost vzorca 1000 oz. za dva podvzorca velikosti 500 dobimo kar natančne ocene kovariančne matrike. Zopet so največja odstopanja pri oceni $\sigma_1$. Spodnji grafi prikazujejo napake ocene kovariančne matrike za 30 ponovitev.




SLIKA 13. Napaka ocene $\sigma_1$ v povprečju.




SLIKA 14. Napaka ocene $\sigma_{12}$ v povprečju.




SLIKA 15. Napaka ocene $\sigma_2$ v povprečju.

Za zaključek pripomnimo, da se z večanjem dimenzije večrazsežne normalne porazdelitve ocene matematičnega upanja in kovariančne matrike ne slabšajo. Tako lahko zaključimo, da je natančnost cenilke za matematično upanje zelo natančna ne glede na podatke medtem ko je cenilka kovariančne matrike za nizke vrednosti, natančna, za velike vrednosti pa natančnost močno pade oz. moramo obravnavati dosti večji vzorec.

## Slovar strokovnih izrazov

**Hadamard's inequality** Hadamardova neenakost – za determinanto matrike $A$, z stolpci $a_i$, velja neenakost $\det A \leq \prod_{i=1}^{n} \|a_i\|_2$
**matrix differentation** matrično odvajanje
**monotone sample** monoton vzorec
**multivariate normal distribution** večrazsežna normalna porazdelitev
**sample covariance matrix** vzorčno kovariančna matrika
**sample mean vector** vektor vzorčnih upanj

## Literatura


[1] T. W. Anderson, *Maximum Likelihood Estimates for a Multivariate Normal Distribution when some Observations are Missing*, Journal of the American Statistical Association **52** (1957), 200–203.
[2] T. W. Anderson in I. Olkin, *Maximum-Likelihood Estimation of the Parameters of a Multivariate Normal Distribution*, Linear algebra and it's applications **70** (1985) 147–171.
[3] R. J. Barnes, *Matrix Differentation*, [ogled 20. 11. 2018], dostopno na https://atmos.washington.edu/~dennis/MatrixCalculus.pdf.
[4] M. I. Jordan, *The Multivariate Gaussian*, [ogled 1. 12. 2018], dostopno na https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/other-readings/chapter13.pdf.
[5] C. Liu, *Bartlett's Decomposition of the Posterior Distribution of the Covariance for Normal Monotone Ignorable Missing Data*, Journal of Multivariate Analysis **46** (1993), 198–206.
[6] C. Liu, *Efficient ML estimation of the multivariate normal distribution from incomplete data*, Journal of Multivariate Analysis **69** (1999), 206–217.
[7] Y. Liu, *Statistical Estimation in Multivariate Normal Distribution With a Block of Missing Observations*, [ogled 10. 2. 2019], dostopno na https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/27145/LIU-DISSERTATION-2017.pdf?sequence=1&isAllowed=y.
[8] B. Plestenjak, *Numerične metode*, verzija 4. 3. 2010, [ogled 10. 8. 2019], dostopno na https://www.fmf.uni-lj.si/~plestenjak/Vaje/NaFgg/Predavanja/OldKnjiga_NM.pdf.
[9] *Determinant*, v: Wikipedia:Te Free Encyclopedia, [ogled 23. 1. 2019], dostopno na https://en.wikipedia.org/wiki/Determinant.
[10] *Hadamard's inequality*, v: Wikipedia: The Free Encyclopedia, [ogled 12. 8. 2019], dostopno na https://en.wikipedia.org/wiki/Hadamard%27s_inequality.
[11] *Is a sample covariance matrix always symmetric and positive definite?*, v: Stats.Stackexchange, [ogled 20. 8. 2019], dostopno na https://stats.stackexchange.com/questions/52976/is-a-sample-covariance-matrix-always-symmetric-and-positive-definite.
[12] *Is the trace of inverse matrix convex?*, v: Stats.Stackexchange, [ogled 20. 8. 2019], dostopno na https://math.stackexchange.com/questions/297635/is-the-trace-of-inverse-matrix-convex.
[13] *Matrix calculus*, v: Wikipedia:The Free Encyclopedia, [ogled 27. 10. 2018], dostopno na https://en.wikipedia.org/wiki/Matrix_calculus.
[14] *Maximum likelihood estimate*, v: Wikipedia: The Free Encyclopedia, [ogled 8. 11. 2018] dostopno na https://en.wikipedia.org/wiki/Maximum_likelihood_estimation.
[15] *Maximum Likelihood Estimation* v: PennState: Eberly College of Science, [ogled 8. 11. 2018] dostopno na https://newonlinecourses.science.psu.edu/stat414/node/191/.

[16] *Maximum Likelihood Estimators - Multivariate Gaussian*, v: Wikipedia: The Free Encyclopedia, [ogled 15. 11. 2018] dostopno na https://stats.stackexchange.com/questions/351549/maximum-likelihood-estimators-multivariate-gaussian?noredirect=1&lq=1.

[17] *Multivariate normal distribution*, v: Wikipedia: The Free Encyclopedia, [ogled 19 .12. 2018] dostopno na https://en.wikipedia.org/wiki/Multivariate_normal_distribution.

[18] *Positive Definite Matrix*, v: Wikipedia: The Free Encyclopedia, [ogled 20. 1. 2019], dostopno na https://en.wikipedia.org/wiki/Definiteness_of_a_matrix.

[19] *Schur complement*, v: Wikipedia: The Free Encyclopedia, [ogled 10. 11. 2018], dostopno na https://en.wikipedia.org/wiki/Schur_complement.

[20] *Trace (linear algebra)*, v: Wikipedia: The Free Encyclopedia, [ogled 2. 3. 2019], dostopno na https://en.wikipedia.org/wiki/Trace_(linear_algebra).

[21] *Why are symetric positive definite(SPD) matrices so important?*, v: Stats.Stackexchange, [ogled 27. 4. 2019], dostopno na https://stats.stackexchange.com/questions/224005/why-are-symmetric-positive-definite-spd-matrices-so-important.