Obrada informacija

Analiza multivarijatnih podataka - primjene u financijama

Zvonko Kostanjčar

Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Ak. god. 2020./2021.

 v1.0

Danas. . .

Indeks tržišta kapitala - težinski prosjek vrijednosti kompanija kojima se trguje na tom tržištu

Tržišta kapitala su iznimno važna! Unatoč tome što se na njih u većini tranzicijskih zemalja gleda kao na kockanje ona su ključni dio svakog gospodarskog sustava!

Investiranje

Investicija: Odricanje (danas) od sredstava (novac) na neko vrijeme kako bi se ostvarili budući povrati koji će kompenzirati investitora za

- **Vrijeme** na koji su sredstva uložena
- Očekivanu stopu **inflacije**
- **Rizik**: nesigurnost budućih isplata

Povrat na investiciju često se naziva i kamata te se može razložiti na tri komponente:

- Vremenska vrijednost novca r_{vrijeme} (godišnje oko 2%)
- Stopa inflacije $r_{\text{inflacija}}$
- Premija na rizik r_{rizik}

$$\text{Povrat} = (1 + r_{\text{vrijeme}})(1 + r_{\text{inflacija}})(1 + r_{\text{rizik}}) - 1$$

Povrat i rizik

- Početna vrijednost investicije: 100.000 EUR, horizont: 1 godina.
- Investicija A

Scenarij	povrat	vjerojatnost povrata	krajnja vrijednost
Optimistični	100%	0.2	200.000
Neutralni	0%	0.6	100.000
Pesimistični	-50%	0.2	50.000

- Investicija B

Scenarij	povrat	vjerojatnost povrata	krajnja vrijednost
Optimistični	20%	0.3	120.000
Neutralni	0%	0.6	100.000
Pesimistični	-10%	0.1	90.000

Povrat i rizik

- U ovim primjerima povrat je diskretna slučajna varijabla

$$R \sim \begin{pmatrix} r_1 & r_2 & \dots & r_N \\ p_1 & p_2 & \dots & p_N \end{pmatrix}$$

- Očekivani povrat investicije $E(R) = \sum_{i=1}^{N} r_i p_i$
- Rizik investicije $STD(R) = \sqrt{\sum_{i=1}^{N} p_i (r_i - E(r_i))^2}$
- Koliki su očekivani povrati i rizici investicije A i B?
- Koju investiciju odabrati?
- Općenito se povrat modelira s neprekidnom slučajnom varijablom s odgovarajućom gustoćom
- Očekivani povrat i rizik se onda modeliraju kao očekivanje i standardna devijacija te slučajne varijable

Funkcija distribucije

Neprekidna slučajna varijabla

Neka je $(\Omega, \mathfrak{F}, \mathrm{P})$ vjerojatnosni prostor. Preslikavanje $X : \Omega \to \mathbb{R}$ nazivamo slučajna varijabla ako je za svaki $x \in \mathbb{R}$ skup $\{\omega \in \Omega : X(\omega) < x\}$ događaj, dakle element od $\mathfrak{F}$.

Funkcija distribucije slučajne varijable

Neka je $(\Omega, \mathfrak{F}, \mathrm{P})$ vjerojatnosni prostor te neka je $X : \Omega \to \mathbb{R}$ slučajna varijabla. Funkcija $F = F_X : \mathbb{R} \in [0, 1]$ definirana formulom

$$F(x) = F_X(x) = \mathrm{P}(X < x), x \in \mathbb{R},$$

zove se funkcija distribucije slučajne varijable X.

Funkcija gustoće

Funkcija gustoće slučajne varijable

Slučajna varijabla $X : \Omega \to \mathbb{R}$ je neprekidna slučajna varijabla ako postoji funkcija $f = f_X : \mathbb{R} \to [0, \infty]$ takva da vrijedi

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \forall x \in \mathbb{R}$$

i f_X se zove vjerojatnosna funkcija gustoće.

Egzistencija i zadavanje neprekidne slučajne varijable

Svaka funkcija gustoće neprekidne slučajne varijable ima svojstva

$$f(x) \geq 0, \forall x \in \mathbb{R} \texttt{ te } \int_{-\infty}^{\infty} f(x)dx = 1.$$

Vrijedi i obrat.

Očekivanje

Neka je $X : \Omega \to \mathbb{R}$ neprekidna slučajna varijabla s funkcijom gustoće f_X. Ako integral $\int_{-\infty}^{\infty} |x| f_X(x) dx$ postoji, tada kažemo da slučajna varijabla ima matematičko očekivanje i definiramo

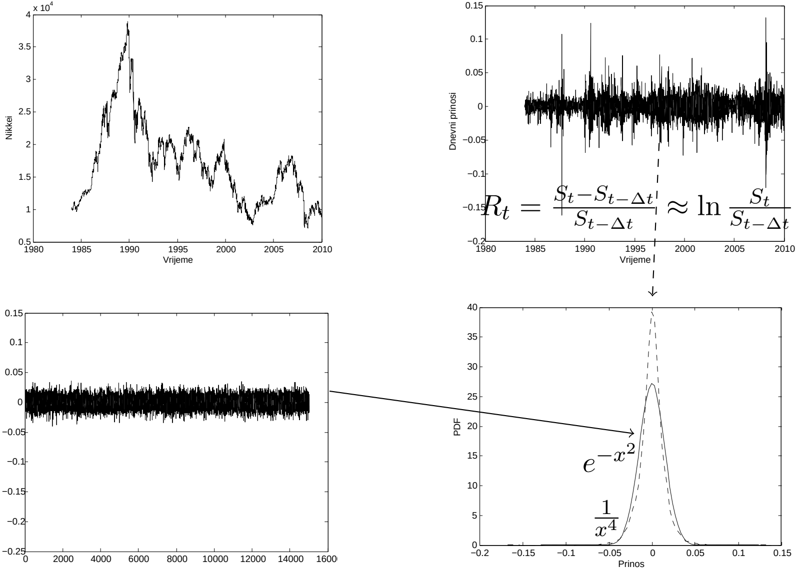
$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Varijanca

Neka je $X : \Omega \to \mathbb{R}$ neprekidna slučajna varijabla s funkcijom gustoće f_X. Ako integral $\int_{-\infty}^{\infty} x^2 f_X(x) dx$ postoji, tada kažemo da slučajna varijabla ima varijancu i definiramo

$$Var(X) = E\left((X - E(X))^2\right).$$

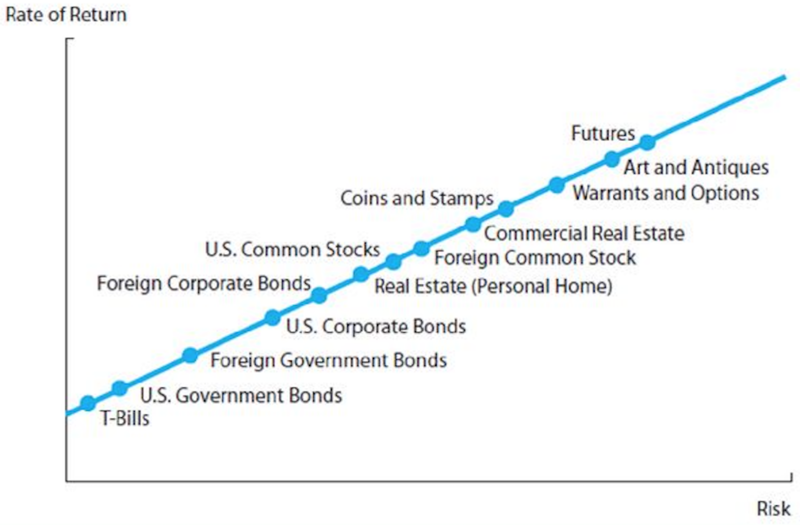
Funkcija gustoće povrata



Osnovne investicijske klase

- Instrumenti fiksnog prinosa (obveznice)
  - suštinski slično kreditu
  - ugovorena dinamika isplata investitoru
- Dionice
  - vlasnički udjel u trgovačkom društvu
  - prihod od dividende i/ili kapitalne dobiti
- Nekretnine
  - vlasnički udjel u stambenom/poslovnom objektu ili zemlji
  - prihod od najma i/ili kapitalne dobiti
- Izvedenice
  - vrijednost im ovisi o nekom drugom financijskom instrumentu
  - opcije, unaprijedni ugovori, swapovi, itd
- Alternativna ulaganja
  - venture investicije, umjetnine, collectibles
  - niska likvidnost

Investicijske klase



Kovarijacija

Neka su $X, Y : \Omega \to \mathbb{R}$ slučajne varijable koje imaju varijancu, tada se njihova kovarijacija definirana s

$$Cov(X,Y) = E\left((X - EX)(Y - EY)\right).$$

Svojstva očekivanja i varijance

Neka su $X, Y : \Omega \to \mathbb{R}$ neprekidne slučajne varijable koje imaju očekivanje, odnosno varijancu. Tada vrijedi

$$E\left(aX + bY\right) = aE(X) + bE(Y).$$

$$Var\left(aX + bY\right) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y).$$

Portfelj

Portfelj je linearna kombinacija investicija, čiji povrat je dan s

$$R_p = \sum_{i=1}^{N} w_i R_i, \sum_{i=1}^{N} w_i = 1,$$

gdje je R_i povrat i-te investicije, a w_i udjel i-te investicije.

Očekivani povrat portfelja

$$E(R_p) = \sum_{i=1}^{N} w_i E(R_i)$$

Rizik portfelja (standardna devijacija)

$$STD(R_p) = \sqrt{\sum_{i=1}^{N} w_i^2 Var(R_i) + \sum_{i,j=1, i \neq j}^{N} w_i w_j Cov(R_i, R_j)}$$

Pearsonov koeficijent korelacije

Neka su $X, Y : \Omega \to \mathbb{R}$ slučajne varijable koje imaju varijance σ_X^2, σ_Y^2, tada je njihov Pearsonov koeficijent korelacije dan s

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Svojstva koeficijenta korelacije

Neka je ρ koeficijent korelacije slučajnih varijabli X i Y. Tada je $\rho \in [-1, 1]$ te vrijedi:

- $\rho = 0$ - slučajne varijable X i Y nisu linearno povezane (kažemo da su nekorelirane)
- $\rho = 1$ - slučajne varijable su pozitivno korelirane i postoji linearna veza između njih
- $\rho = -1$ - slučajne varijable su negativno korelirane i postoji linearna veza između njih

Svojstva varijance aritmetičke sredine slučajnih varijabli

Neka su X_i slučajne varijable koje imaju varijancu te neka je $Y = \frac{1}{n}\sum_{i=1}^{i=n} X_i$. Tada vrijedi

1. ako su varijable X_i nekorelirane

$$Var(Y) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} X_i = \frac{Var(X)}{n}$$

2. ako su varijable korelirane s prosječnom korelacijom ρ

$$Var(Y) = \frac{Var(X)}{n} + \frac{n-1}{n}\rho Var(X)$$

Markowitzev model, Nobelova nagrada 1990.

1. Harry Markowitz
   utjecajan ekonomist na Rady school of Managment i na University of California, San Diego.
2. Ključni doprinosi
   1. pokazao je da vrijedi diverzificirati
   2. pokazao je kako optimizirati portfelj

Moderna teorija portfelja

Optimizacijski problem

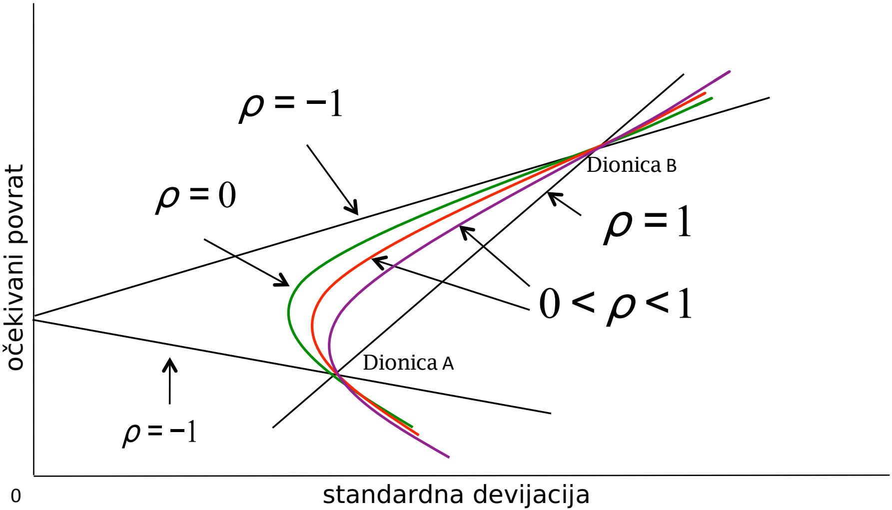
1. **minimiziraj** $\sum_{i,j=1}^{N} w_i w_j Cov(R_i, R_j)$, s obzirom na
   $\sum_{i=1}^{N} w_i E(R_i) = E(R_p)$,
   $\sum_{i=1}^{N} w_i = 1$

2. **maksimiziraj** $\sum_{i=1}^{N} w_i E(R_i)$, s obzirom na
   $\sum_{i,j=1}^{N} w_i w_j Cov(R_i, R_j) = {\sigma_p}^2$
   $\sum_{i=1}^{N} w_i = 1$

Primjer

	Dionica A		Dionica B		Kovarijanca
Vrijeme	R_1	$(R_{1i} - E(R_1))^2$	R_2	$(R_{2i}-E(R_2))^2$	$[R_{1i}-E(R_1)][R_{2i}-E(R_2)]$
sij. 2014.	-2,0%	0,1%	-3,0%	0,3%	0,16%
vlj. 2014.	6,4%	1,3%	8,5%	0,4%	0,34%
ožu. 2014.	13,8%	1,6%	5,8%	0,1%	0,47%
tra. 2014.	-1,9%	0,1%	8,1%	0,4%	-0,18%
svi. 2014.	3,4%	0,1%	-2,6%	0,2%	-0,11%
lip. 2014.	9,9%	0,8%	1,9%	0,0%	-0,01%
srp. 2014.	-5,2%	0,4%	-1,2%	0,1%	0,21%
kol. 2014.	-18,0%	3,6%	-5,0%	0,5%	1,37%
ruj. 2014.	-10,3%	1,3%	7,7%	0,3%	-0,64%
lis. 2014.	17,0%	2,5%	1,0%	0,0%	-0,17%
stu. 2014.	4,8%	0,1%	5,8%	0,1%	0,14%
pro. 2014.	-4,1%	0,3%	-2,1%	0,2%	0,22%
Očekivani prinos	1,16%		2,08%		
St. dev.		10,11%		4,91%	
Kovarijanca					0,16%
Koeficijent korelacije					32,96%

Primjer - mogući portfelji

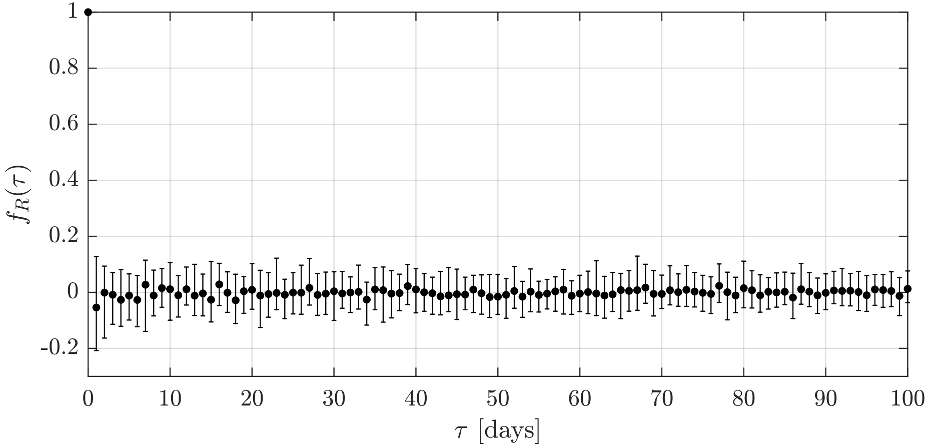


Normalizirana autokorelacijska funkcija

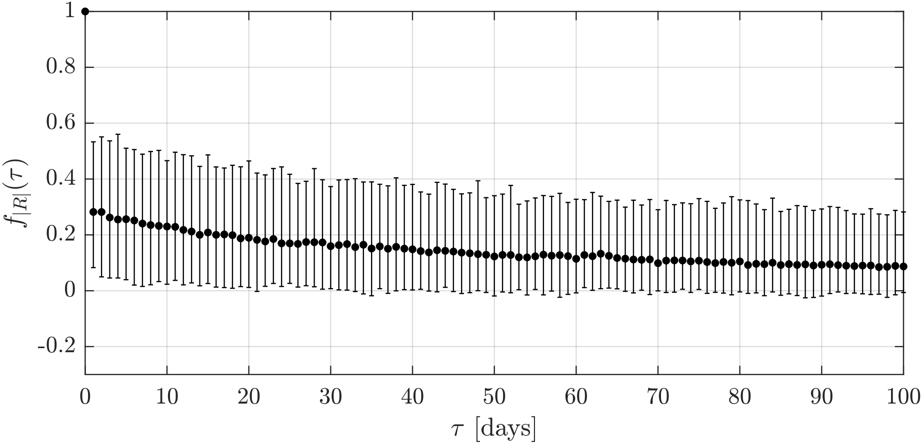
Neka je $\{X_i, i = 1, \ldots, T\}$ niz jednako distribuiranih slučajnih varijabli koje imaju očekivanje μ i varijancu σ^2. Normalizirana autokorelacijska funkcija definirana je izrazom

$$f_X(\tau) = \frac{Cov(X_t, X_{t+\tau})}{\sigma^2}$$

Normalizirana autokorelacijska funkcija povrata



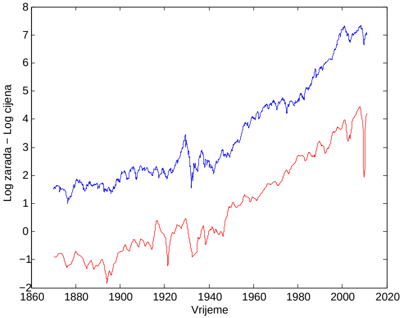
Normalizirana autokorelacijska funkcija apsolutnih povrata



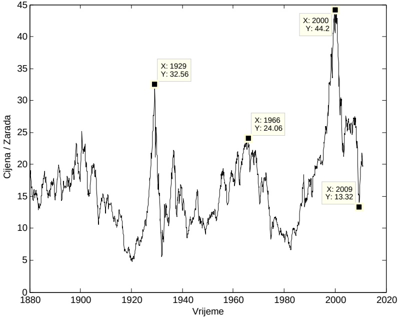
Financijski signali ili financijski vremenski nizovi

- Cijene financijskih instrumenata
  - cijene dionica
  - cijene obveznica
  - cijene opcija
  - itd
- Makroekonomske varijable
  - kamatne stope
  - tečajne liste (valute)
  - bruto domaći proizvod (BDP)
  - itd
- Podatci o kompanijama
  - zarada
  - promet
  - dug
  - itd

Odnos cijene i zarade

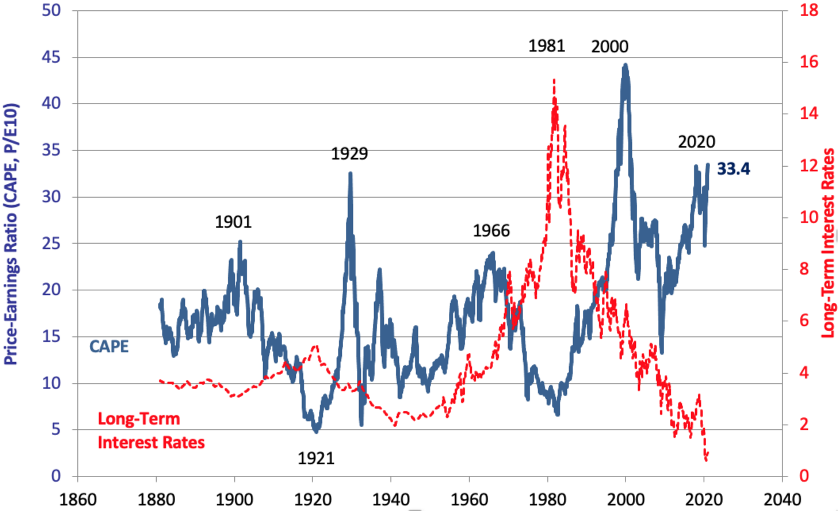


(a) Logaritam S&P 500 indeksa i logaritam odgovarajućih zarada



(b) P/E za S&P 500

Odnos cijene i zarade



Danas. . .

Analiza glavnih komponenti - uvod

Analiza glavnih komponenti bavi se objašnjavanjem kovarijancijske strukture skupa varijabli s nekoliko linearnih kombinacija tih varijabli

Glavni ciljevi analize glavnih komponenti su:

- redukcija dimenzionalnosti
- interpretacija

Algebarski, glavne komponente su određene linearne kombinacije slučajnih varijabli $X_1, \ldots, X_p$
Geometrijski, te linearne kombinacije predstavljaju nove koordinatne osi dobivene rotacijom originalnog sustava s $X_1, \ldots, X_p$ kao originalnim osima.

Analiza glavnih komponenti - uvod

Neka je $X' = [X_1, \dots, X_p]$ slučajni vektor s kovarijacijskom matricom Σ sa svojstvenim vrijednostima $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$.
Razmatramo linearne kombinacije:

$$Y_1 = a_1'X = a_{11}X_1 + a_{12}X_2 + \dots a_{1p}X_p$$

$$Y_2 = a_2'X = a_{21}X_1 + a_{22}X_2 + \dots a_{2p}X_p$$

$$\dots$$

$$Y_p = a_p'X = a_{p1}X_1 + a_{p2}X_2 + \dots a_{pp}X_p$$

Očigledno vrijedi:

$$Var(Y_i) = a_i'\Sigma a_i, i = 1, 2, \dots, p$$

$$Cov(Y_i, Y_k) = a_i'\Sigma a_k, i, k = 1, 2, \dots, p$$

Glavne komponente su one nekorelirane linearne kombinacije $Y_1, \dots, Y_p$ čije varijance su najveće moguće.

Analiza glavnih komponenti - zadatak

Prva glavna komponenta = linearna kombinacija $a_1'X$ koja

$$\text{maksimizira } Var(a_1'X), \text{uz uvjet } a_1'a_1 = 1$$

Druga glavna komponenta = linearna kombinacija $a_2'X$ koja

$$\text{maksimizira } Var(a_2'X), \text{uz uvjete } a_2'a_2 = 1, Cov(a_1'X, a_2'X) = 0$$

Analiza glavnih komponenti - uvod

Glavne komponente

Neka je Σ kovarijacijska matrica pridružena slučajnom vektoru $X' = [X_1, \ldots X_p]$. Neka Σ ima parove vlastitih vrijednosti i vlastitih vektora $(\lambda_1, e_1), \ldots, (\lambda_p, e_p)$, pri čemu vrijedi $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$. Tada je i-ta glavna komponenta dana s

$$Y_i = e_i' X = e_{i1} X_1 + \ldots e_{ip} X_p, i = 1, \ldots, p,$$

te vrijedi

$$Var(Y_i) = e_i' \Sigma e_i = \lambda_i$$

$$Cov(Y_i, Y_k) = e_i' \Sigma e_k = 0, i \neq k$$

Ako su neki λ_i jednaki, rastav nije jedinstven.

Analiza glavnih komponenti - uvod

Glavne komponente - dekompozicija varijance

Neka je Σ kovarijacijska matrica pridružena slučajnom vektoru $X' = [X_1, \dots X_p]$. Neka Σ ima parove vlastitih vrijednosti i vlastitih vektora $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$, pri čemu vrijedi $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$. Neka su $Y_1 = e_1' X, Y_2 = e_2' X, \dots, Y_p = e_p' X$ glavne komponente. Tada za ukupnu varijancu vrijedi

$$\sigma_{11}^2 + \sigma_{22}^2 + \dots + \sigma_{pp}^2 = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \dots + \lambda_p = \sum_{i=1}^{p} Var(Y_i)$$

Analiza glavnih komponenti - uvod

Glavne komponente - korelacije

Neka je Σ kovarijacijska matrica pridružena slučajnom vektoru $X' = [X_1, \ldots X_p]$. Neka Σ ima parove vlastitih vrijednosti i vlastitih vektora $(\lambda_1, e_1), \ldots, (\lambda_p, e_p)$, pri čemu vrijedi $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$. Neka su $Y_1 = e_1' X, Y_2 = e_2' X, \ldots, Y_p = e_p' X$ glavne komponente. Tada je koeficijent korelacije između glavne komponente Y_i i slučajne varijable X_k dan s

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sigma_{kk}}, i, k = 1, 2, \ldots, p$$

Analiza glavnih komponenti - primjer

Glavne komponente - primjer 1

Pretpostavimo da slučajne varijable X_1, X_2 i X_3 imaju kovarijacijsku matricu

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- Odredite svojstvene vrijednosti i svojstvene vektore matrice
- Odredite glavne komponente
- Provjerite svojstvo dekompozicije varijance

Analiza glavnih komponenti - standardizirane varijable

Glavne komponente možemo dobiti iz standardiziranih varijabli

$$Z_1 = \frac{X_1 - \mu_1}{\sigma_{11}}$$

$$Z_2 = \frac{X_2 - \mu_2}{\sigma_{22}}$$

$$\dots$$

$$Z_p = \frac{X_p - \mu_p}{\sigma_{pp}}$$

U matričnoj formi to se može zapisati

$$Z = (V^{\frac{1}{2}})^{-1}(X - \mu),$$

gdje je $V^{\frac{1}{2}}$ dijagonalna matrica standardnih devijacija.
Očigledno vrijedi $E(Z) = 0$ i

$$Cov(Z) = (V^{\frac{1}{2}})^{-1}\Sigma(V^{\frac{1}{2}})^{-1} = C,$$

gdje je C korelacijska matrica slučajnog vektora X.

Analiza glavnih komponenti - uvod

Glavne komponente - korelacije

Neka je i-ta glavna komponenta standardiziranih varijabli $Z' = [Z_1, \ldots, Z_p]$ s kovarijacijskom matricom $Cov(Z) = C$, dana s

$$Y_i = e_i' Z = e_i' (V^{\frac{1}{2}})^{-1} (X - \mu), i = 1, 2, \ldots, p$$

Tada vrijedi,

$$\sum_{i=1}^{p} = Var(Y_i) = \sum_{i=1}^{p} = Var(Z_i) = p,$$

i

$$\rho_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i}, i, k = 1, \ldots, p.$$

U tom slučaju $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_p, e_p)$ su parovi svojstvenih vrijednosti i svojstvenih vektora matrice ρ, uz $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$

Analiza glavnih komponenti - posebne strukture

Primjer 1 - dijagonalna struktura

Neka je kovarijacijska matrica dijagonalna

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{bmatrix}$$

- i-ta svojstvena vrijednost i svojstveni vektor dani su s (σ_{ii}, e_i), gdje je $e_i' = [0, \dots, 0, 1, 0 \dots, 0]$
- Skup glavnih komponenti je originalni (polazni) skup nekoreliranih slučajnih varijabli
- Standardizacija vodi do suštinski istog rezultata

Analiza glavnih komponenti - posebne strukture

Primjer 2 - jednake korelacije

Neka je kovarijacijska matrica ($\rho > 0$)

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

Tada je korelacijska matrica

$$C = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

Analiza glavnih komponenti - posebne strukture

- Tada je najveća svojstvena vrijednost $\lambda_1 = 1 + (p-1)\rho$, s pridruženim svojstvenim vektorom $e_1' = [\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}}]$
- Ostale svojstvene vrijednosti su $\lambda_2 = \lambda_3 = \ldots = \lambda_p = 1 - \rho$ s pripadnim svojstvenim vektorima

$$e_i' = \left[\frac{1}{\sqrt{(i-1)i}}, \dots, \frac{1}{\sqrt{(i-1)i}}, \frac{-(i-1)}{\sqrt{(i-1)i}}, 0, \dots, 0\right]$$

- Prva glavna komponenta je $Y_1 = e_1' Z_1 = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} Z_i$ proporcionalna zbroju p standardiziranih varijabli (često se naziva indeks jednakih težina)
- Prva glavna komponenta objašnjava $\frac{\lambda_1}{p} = \rho + \frac{1-\rho}{p}$ proporciju ukupne varijance.