

box-mueller preizkušanje domnev indikatorji neodvisnost  
slučajno polje monte carlo križni variogram gumbel  
semivariogram

fakulteta za gradbeništvo in geodezijo

navadno krigiranje

rezidual

univerza v ljubljani

weibull

2023

prostorska  
statistika

PDF  
poisson  
prag  
hipoteza  
moran geostatistika

gostota  
frechet

krigiranje s trendom  
odstopanje  
prejem zavrnitev  
robustna  
kovariogram  
krige  
varianca  
osamelec  
vzorec  
jura  
inverzna metoda  
porazdelitena funkcija  
buffonova igla  
vezani ekstrem  
geary  
preizkus  
krigiranje  
CDF

najmanjši kvadrati

goran turk

regresija kvadrati variogram cholesky





Univerza  
*v Ljubljani*

Fakulteta  
*za gradbeništvo  
in geodezijo*



GORAN TURK

# Prostorska statistika

UNIVERZITETNI UČBENIK

LJUBLJANA, OKTOBER 2023

## **PROSTORSKA STATISTIKA**

*Univerzitetni učbenik*

Avtor: *Goran Turk*

Recenzenta: *Gregor Dolinar, Samo Drobne*

Jezikovni pregled: *Nina Ambrožič*

Naslovnica: *Goran Turk*

Oblikovanje in prelom: *Goran Turk*

Založnik: *Založba Univerze v Ljubljani (University of Ljubljana Press)*

Za založbo: *Gregor Majdič, rektor Univerze v Ljubljani*

Izdajatelj: *Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo*

Za izdajatelja: *Violeta Bokan Bosiljkov, dekanja Fakultete za gradbeništvo in geodezijo*

Ljubljana, 2023

Prva e-izdaja.

Publikacija je v digitalni obliki prosto dostopna na: <https://ebooks.uni-lj.si>

DOI: 10.15292/9789612971991

Kataložni zapis o publikaicji (CIP) pripravili v  
Narodni in univerzitetni knjižnici v Ljubljani

COBISS.SI-ID 171155459

ISBN 978-961-297-199-1 (PDF)



To delo je ponujeno pod licenco Creative Commons Priznanje avtorstva-Nekomercialno-Brez predelav 4.0 Mednarodna licenca.

This work is licenced under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International licence.

# Predgovor

Prostorska statistika je razširitev običajne statistike, pri kateri nimamo podatka o legi ali času izvedbe poskusa ali pa nas to ne zanima. Pri prostorski statistiki obravnavamo slučajne spremenljivke z dodatno informacijo o legi ali času, za kateri smo dobili posamezen element vzorca. Takim slučajnim spremenljivkam pravimo *slučajna polja*, če imamo podatke o legi v prostoru, ali *slučajni procesi*, če poznamo podatke o času ali legi v prostoru in času.

Učbenik vključuje kratko uvodno poglavje, ki opredeli osnovne značilnosti prostorske statistike, prikaže razvoj oziroma zgodovino te relativno nove veje uporabne statistike in pojasni, s katerimi vrstami spremenljivk in podatkov se prostorska statistika ukvarja. V tem poglavju je tudi seznam različnih geostatističnih podatkov, ki jih študenti lahko uporabijo za vaje in preizkušanje različnih metod prostorske statistike. Večino teh podatkov sem pridobil na svetovnem spletu, primer podatkov iz Slovenije pa sta prispevala raziskovalca Kmetijskega inštituta Slovenije, Borut Vrščaj in Janez Bergant, za kar se jima iskreno zahvaljujem.

Za razumevanje prostorske statistike je nujno osnovno predznanje verjetnostnega računa in statistike. To predvsem vključuje poznavanje in razumevanje osnovnih pojmov, kot so slučajna spremenljivka, porazdelitev slučajne spremenljivke, porazdelitvena funkcija, slučajni vektor, momenti slučajnih spremenljivk in vektorjev (pričakovana vrednost, varianca, kovarianca), funkcije slučajnih spremenljivk in momenti funkcij slučajnih spremenljivk, osnovne statistike (povprečje, varianca vzorca, kovarianca vzorca), preizkušanje domnev ter regresija. Zato je v učbeniku kratko drugo poglavje, kjer so na kratko naštet in definirani ti osnovni pojmi. V zaključku tega poglavja je definirana tudi skalarna in vektorska slučajna funkcija.

V poglavju Geostatistika so opisane osnovne metode za analizo slučajnih polj in procesov. To vključuje definicijo razsevnih grafov, kovariančnih funkcij, variograma in dveh pogosto uporabljenih preizkusov domnev o lastnostih slučajnih polj (Moranov indeks in Gearyjevo razmerje).

Krigiranje je pomembno regresijsko orodje, zato ga tu opišemo v samostojnem poglavju. Opišemo tri osnovne metode krigiranja: preprosto, običajno in krigiranje s trendom.

Kot alternativa krigiranju lahko uporabimo tudi prostorsko regresijo, ki je pravzaprav običajna linearna regresija z utežmi, ki so določene tako, da bližnje točke močneje vplivajo na vrednost regresije. Postopki določitve uteži in parametrov prostorske linearne regresije so opisani v poglavju Prostorsko utežena re-

gresija.

V poglavju Prostorski vzorci analiziramo razpored točk v prostoru, same vrednosti polja pa nas tu ne zanimajo. Opišemo dve skupini metod: metoda kvadratov in metoda najbližjega soseda.

Zadnje poglavje na videz ne sodi v učbenik Prostorska statistika, saj je metoda Monte Carlo in generiranje vzorcev slučajnih spremenljivk bolj splošno področje. Kljub temu pa predstavlja osnovni gradnik pri generiranju slučajnih polj in procesov in zato kot zadnje poglavje ustrezno zaključuje učbenik.

Pisanje učbenika je dolgotrajen postopek, ki se je pri meni vedno začel s pisanjem zapiskov pred in po predavanjih. Po nekaj letih je nastal osnutek učbenika, ki sem ga nato uredil in pripravil v obliki, ki je bila že precej podobna zdajšnji verziji. V nekaj naslednjih letih smo jaz in moji študenti učbenik uporabljali in ob tem odkrivali različne napake, pomanjkljivosti in možnosti za izboljšanje. Študentom, ki so s pripombami aktivno sodelovali pri tem, se na tem mestu lepo zahvaljujem. V zaključni fazi priprave učbenika sta učbenik skrbno prebrala recenzenta Gregor Dolinar in Samo Drobne. Za njune pripombe, ki nedvomno izboljšujejo učbenik, se jima iskreno zahvaljujem. Posebej pa se moram zahvaliti še Nini Ambrožič, ki je učbenik zelo natančno lektorirala.

Že na začetku pisanja sem se odločil, da učbenik ne bo šel v klasični tisk, temveč bo na voljo prvenstveno v digitalni obliki. Zato učbenik vključuje povezave na različne spletne strani z dodatnimi informacijami, vključene pa so tudi kratke animacije, ki še dodatno pojasnijo vplive različnih parametrov.

Učbenik je v prvi vrsti namenjen študentom druge stopnje pri pouku Prostorske statistike, lahko pa tudi kot osnovno gradivo študentom tretje stopnje, ki jih zanima oziroma želijo uporabiti metode te veje statistike. Upam, da vam bo učbenik pomagal do znanja, ki si ga želite pridobiti.

Goran Turk

Ljubljana, jesen 2023

# Vsebina

<b>Predgovor</b>	<b>i</b>
<b>Vsebina</b>	<b>iii</b>
<b>1 Uvod</b>	<b>1</b>
1.1 Zgodovina prostorske statistike . . . . .	1
1.2 Podatki za prostorsko statistiko . . . . .	2
1.3 Vrste obravnavnih slučajnih spremenljivk . . . . .	3
1.3.1 Geostatistični podatki . . . . .	4
1.3.2 Podatki v mreži . . . . .	6
1.3.3 Prostorski točkovni vzorci . . . . .	8
<b>2 Slučajne funkcije, polja in procesi</b>	<b>9</b>
2.1 Slučajne spremenljivke . . . . .	9
2.2 Slučajni vektorji . . . . .	10
2.3 Momenti slučajnih spremenljivk in vektorjev . . . . .	12
2.4 Slučajne funkcije . . . . .	13
2.5 Vektorske slučajne funkcije . . . . .	18
<b>3 Geostatistika</b>	<b>23</b>
3.1 Geostatistični podatki . . . . .	23
3.2 Priprava podatkov . . . . .	25
3.3 Razsevni graf . . . . .	28
3.4 Kovariančna funkcija . . . . .	29
3.5 Variogram . . . . .	30
3.6 Določitev empirične funkcije variograma . . . . .	31

3.7	Geostatistična analiza z indikatorji . . . . .	33
3.8	Preizkušanje domnev o prostorski povezanosti . . . . .	35
3.8.1	Moranov indeks . . . . .	37
3.8.2	Gearyjevo razmerje . . . . .	40
<b>4</b>	<b>Krigiranje</b>	<b>45</b>
4.1	Preprosto krigiranje . . . . .	46
4.2	Običajno krigiranje . . . . .	50
4.3	Krigiranje s trendom . . . . .	51
4.4	Primeri krigiranja . . . . .	54
4.4.1	Preprosto krigiranje . . . . .	55
4.4.2	Običajno krigiranje . . . . .	56
4.4.3	Krigiranje s trendom . . . . .	57
<b>5</b>	<b>Prostorsko utežena regresija</b>	<b>61</b>
5.1	Linearna regresija . . . . .	61
5.2	Utežena linearna regresija . . . . .	64
5.2.1	Določitev uteži . . . . .	64
5.2.2	Izbira funkcije za določitev uteži in kritične razdalje . . . . .	65
5.2.3	Primeri . . . . .	66
<b>6</b>	<b>Prostorski vzorci</b>	<b>69</b>
6.1	Metoda kvadratov . . . . .	70
6.2	Primeri – metoda kvadratov . . . . .	73
6.3	Pomanjkljivosti metode kvadratov . . . . .	77
6.4	Metoda najbližjega soseda . . . . .	77
6.5	Primeri – metoda najbližjega soseda . . . . .	81
<b>7</b>	<b>Metoda Monte Carlo</b>	<b>87</b>
7.1	Generiranje vzorca slučajne spremenljivke . . . . .	89
7.1.1	Inverzna metoda generiranja vzorca . . . . .	89
7.1.2	Generiranje vzorca normalno porazdeljenih slučajnih spremenljivk . . . . .	91
7.1.3	Metoda sprejema/zavrnitve . . . . .	92
7.1.4	Primeri . . . . .	93
7.2	Generiranje vzorca slučajnega vektorja . . . . .	98
7.2.1	Neodvisne slučajne spremenljivke – inverzna metoda . . . . .	98

7.2.2	Odvisne slučajne spremenljivke – inverzna metoda . . . . .	100
7.2.3	Korelirane normalno porazdeljene spremenljivke . . . . .	101
7.2.4	Primeri . . . . .	102
<b>8</b>	<b>Literatura</b>	<b>113</b>
	<b>Stvarno kazalo</b>	<b>115</b>





# 1 Uvod

## 1.1 Zgodovina prostorske statistike

Prostorska statistika se je podobno kot druge statistične metode razvijala vzporedno s potrebami pri različnih aplikativnih problemih. Elemente prostorske statistike najdemo v 17. stoletju, ko so naredili prvo karto vetrov na Zemlji (Halley<sup>1</sup>, 1686). Znan je problem Buffonove igle, ki ga prav tako lahko smatramo za enega od začetkov prostorske statistike (Comte de Buffon<sup>2</sup>, 1780). Slučajni procesi (ali polja), kot osnova prostorske statistike, so se prvič uporabili v najpreprostejši obliki Poissonovega procesa pri izračunu srednje dolžine poti molekule v plinu (Clausius<sup>3</sup>, 1858). Poissonov proces še danes predstavlja osnovno merilo o enakomerni in neodvisni, a slučajni razporeditvi elementov v prostoru ali času. Že v začetku 20. stoletja so določili porazdelitev razdalje med poljubno točko v Poissonovem procesu in njeno najbližjo sosedo (Hertz<sup>4</sup>, 1909), kar je še danes osnova za nekatere metode prepoznavanja prostorskih vzorcev.

Pomemben napredek v razvoju vseh statističnih metod, tudi prostorske statistike, se je odvijal v prvi polovici 20. stoletja v Angliji. Gosset<sup>5</sup> in Fisher<sup>6</sup> sta uporabljala metode prostorske statistike in o njih pisala, prvi v pivovarni Guinness (pod psevdenimom *Student*), drugi na posestvu Rothamsted Experimental Station v Hertfordshiru. Fisher je precejšnji del svojega časa namenil statistični analizi podatkov s kmetijskih polj.

Naslednji pomemben napredek se je zgodil v povezavi z geološkimi raziskavami za potrebe rudarjenja v Južni Afriki (Krige<sup>7</sup>, 1951) in v Franciji (Matheron<sup>8</sup>, 1954–64) v sredini 20. stoletja. V tem obdobju je bil pomemben tudi prispevek švedskega avtorja Matérna<sup>9</sup> (1960), ki je deloval na Gozdarski fakulteti Švedske univerze za kmetijske znanosti. Z uporabo modela zveznega prostorskega spreminjanja so postavili osnove za Gaussov<sup>10</sup> prostorski slučajni proces, na osnovi katerega temelji pomembna veja

---

<sup>1</sup> Edmond Halley, angleški astronom in matematik, znan po izračunu orbite kometa (Halleyev komet) (1656–1742).

<sup>2</sup> Georges Louis Leclerc Comte de Buffon, francoski naravoslovec in matematik (1707–1788).

<sup>3</sup> Rudolf Clausius, nemški matematik in fizik (1822–1888).

<sup>4</sup> Heinrich Hertz, nemški fizik (1857–1894).

<sup>5</sup> William Sealy Gosset, angleški statistik (1876–1937).

<sup>6</sup> Sir Ronald Aylmer Fisher, angleški statistik in genetik (1890–1962).

<sup>7</sup> Danie Gerhardus Krige, južnoafriški rudarski inženir, avtor postopka aproksimacije kriging (1919–2013).

<sup>8</sup> Georges Francois Paul Marie Matheron, francoski statistik in geolog (1930–2000).

<sup>9</sup> Bertil Matérn, švedski statistik (1907–2007).

<sup>10</sup> Johann Friedrich Carl Gauss, nemški matematik (1777–1855).

prostorske statistike: geostatistika.

Danes se prostorska statistika uporablja v različnih aplikacijah:

- epidemiološke študije,
- analize razvoja bolezni,
- prostorsko načrtovanje,
- določanje kakovosti rudnine v nahajališčih,
- določanje globine in debeline geoloških plasti,
- določanje prepustnosti in poroznosti v nehomogenih poroznih materialih,
- določanje gostote dreves določene vrste v gozdu,
- določanje lastnosti tal na izbranem področju,
- analiza padavin na izbranem področju,
- določanje meteoroloških parametrov v atmosferi,
- določanje koncentracije onesnaževanja na onesnaženem področju,
- arheološka izkopavanja,
- seizmološke študije in
- mnoge druge.

## 1.2 Podatki za prostorsko statistiko

Osnovna razlika med podatki za običajno statistiko in prostorsko statistiko je v tem, da pri prostorski statistiki vsakemu elementu vzorca pripada informacija o legi v prostoru in/ali času, medtem ko pri običajni statistiki tega podatka ne upoštevamo. Na preprost način lahko to prikažemo v dveh vzporednih preglednicah:

PREGLEDNICA 1.1: Primera podatkov pri prostorski in običajni statistiki

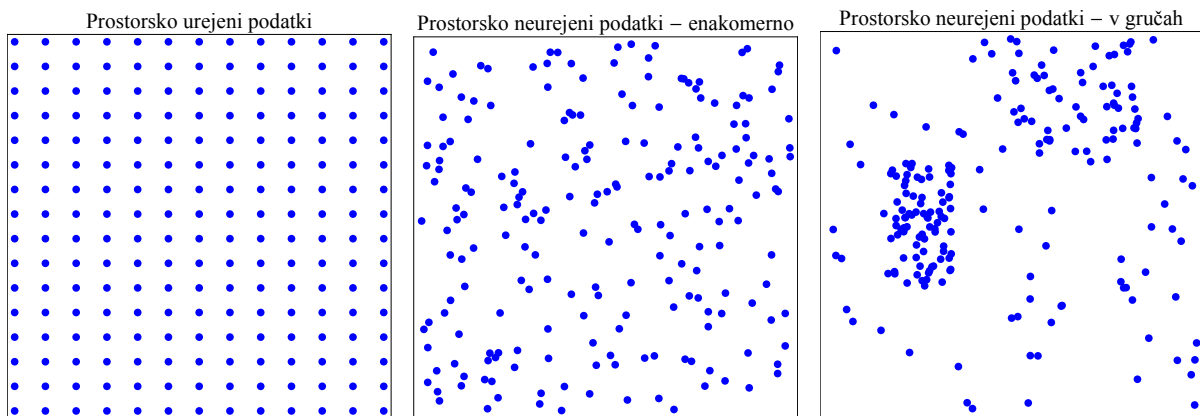
Prostorska statistika			Običajna statistika
$T[^\circ\text{C}]$	$x[\text{m}]$	$y[\text{m}]$	$T[^\circ\text{C}]$
22	112	70	22
15	142	107	15
27	47	78	27
18	30	145	18
...	...		...

V preglednicah 1.1 prikazujemo temperaturo v določenem času na obravnavanem območju. V obeh primerih obravnavamo isto slučajno spremenljivko, torej temperaturo. Razlika je le v tem, da so pri prostorski statistiki na voljo tudi informacije o tem, kje je bila temperatura merjena.

### 1.3 Vrste obravnavnih slučajnih spremenljivk

Slučajne spremenljivke, ki jih obravnavamo v okviru prostorske statistike, so lahko diskretne ali zvezne, podobno kot smo to obravnavali v okviru osnov verjetnostnega računa in statistike. Informacija o legi v prostoru in času ima lahko različne značilnosti:

- prostorska informacija je urejena ali neurejena (glej sliko 1.1),
- prostorska informacija je lahko realizacija diskretne ali zvezne slučajne spremenljivke,
- prostorska informacija je lahko podana v točkah (na primer vodnjak ali vrtina), na krivuljah (na primer cesta ali vodotok), ravninskih območjih (na primer parcela, občina ali država) ali po prostorskih območjih (na primer nahajališče rude).



SLIKA 1.1: Različne razporeditve točk, v katerih so merjeni podatki

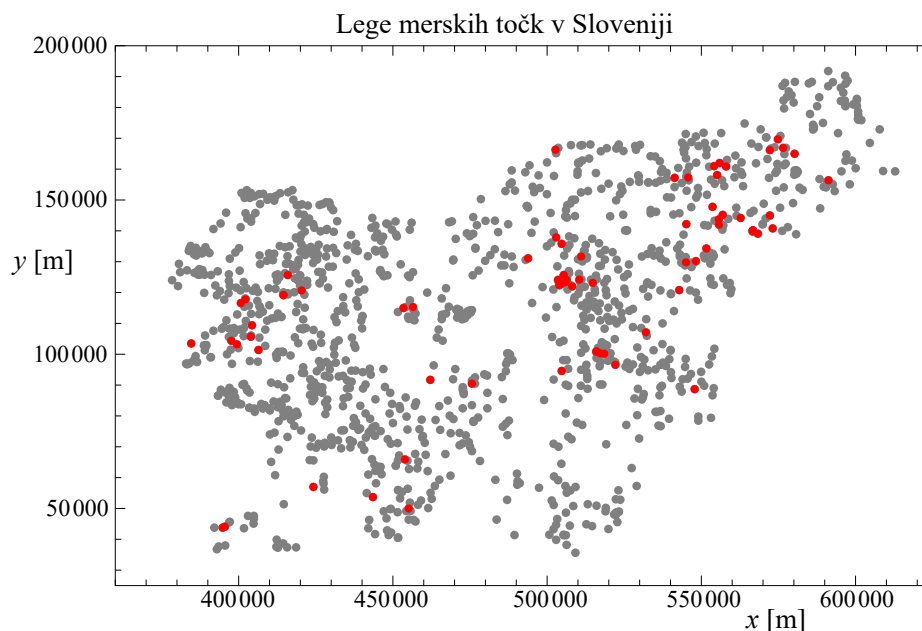
Vrste prostorskih podatkov so:

- geostatistični podatki (ang. geostatistical data),
- mrežna struktura podatkov (ang. lattice data),
- prostorski vzorci točk (ang. point patterns).

### 1.3.1 Geostatistični podatki

Podatki so pripravljene tako, da so vrednosti slučajne spremenljivke, ki jo obravnavamo, merjene v točkah v prostoru, ki so lahko razporejene urejeno ali neurejeno. Tu opisujemo nekaj primerov geostatističnih podatkov, ki jih bralec lahko uporabi pri svojih analizah. Podatke lahko dobite na [spletnih straneh UL FGG](#).

- Podatki **KIS - podatki o pedoloških profilih** predstavljajo vrednosti treh količin na območju Slovenije na globini 0 - 20 cm: vrednost pH (v KCl), količina izmenljivega fosforja P (mg/100 g) in količina izmenljivega kalija K (mg/100 g) v tleh. Podatke so pripravili raziskovalci Kmetijskega inštituta Slovenije (B. Vrščaj in J. Bergant). Na sliki 1.2 so z rdečo označene točke, kjer količina prostega fosforja presega 25 mg/100 g.

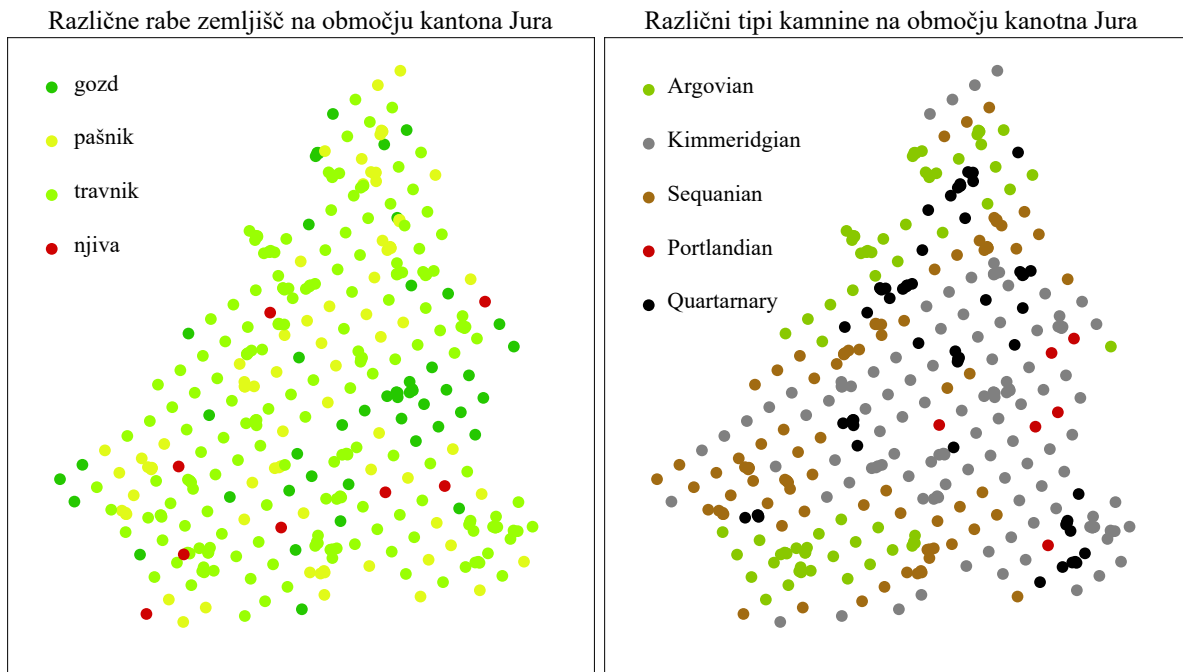


SLIKA 1.2: Razpored točk na območju Slovenije z označenimi točkami z višjo koncentracijo fosforja

- Podatki **GASA - Geostatistical Association of Southern Africa** so pridobljeni z meritvami vsebnosti zlata v rudnini (g/t) in debeline plasti rudnine (cm) v globokih vrtnah v Južnoafriški republiki. Meritve so predstavljale osnovo za načrtovanje rudnika zlata Wits.
- Podatki **Wolfcamp** predstavljajo meritve vodnega tlaka v 85 vodnjakih v New Mexicu in Texasu. Meritve je naročil *Office for Nuclear Waste Isolation* ob načrtovanju odlagališča radioaktivnih odpadkov. Meritve so uporabili za oceno, kako hitro bi morebitno puščanje radioaktivnih odpadkov onesnažilo vodo na tem območju.
- Podatke **Dioxin** so izmerili na območju okoli ceste v Missouriju, kjer je iz tovarnjaka iztekla

neznana količina strupene tekočine. V želji po racionalizaciji meritev so pripravili načrt meritev v urejenem vzorcu okoli ceste. Merili so koncentracijo strupene snovi TCDD (tetraklordibenzo-p-dioksin) v  $\mu\text{g}/\text{kg}$ .

- Podatki **Scallops** predstavljajo podatke o številu ulovljenih školjk na priobalnem območju vzhodne obale ZDA. Poleg koordinat, kjer je bil vzet vzorec, so prešteli število školjk, manjših in večjih od 7 cm, ter skupno število školjk.
- Podatke **Jura** so zbrali na *Swiss Federal Institute of Technology* v Lozani (fr. Lousanne) in predstavljajo značilne geostatistične podatke. Na znanih koordinatah 359 točk v švicarskem kantonu Jura so v tleh merili koncentracije kadmija, kobalta, kroma, bakra, niklja, svinca in cinka. V teh točkah so določili tudi tip kamnine (pet različnih kamnin) in trenutno rabo tal (gozd, pašnik, travnik ali njiva), glej sliko 1.3. Podatke o pokrajini Jura bomo večkrat uporabili pri obravnavanju različnih geostatističnih metod.

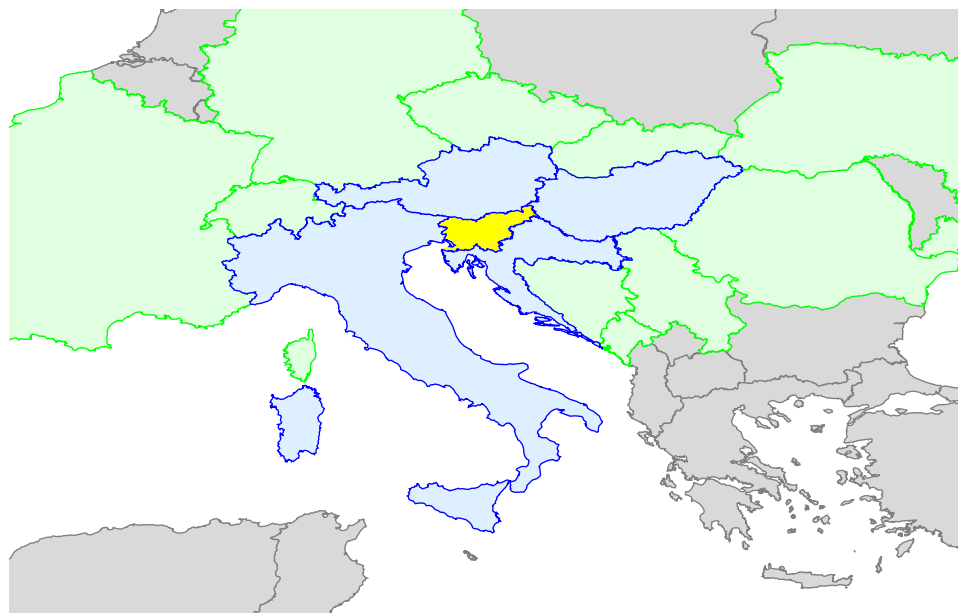


SLIKA 1.3: Razpored točk na območju kantona Jura z oznakami rabe zemljišča in tipa kamnine

### 1.3.2 Podatki v mreži

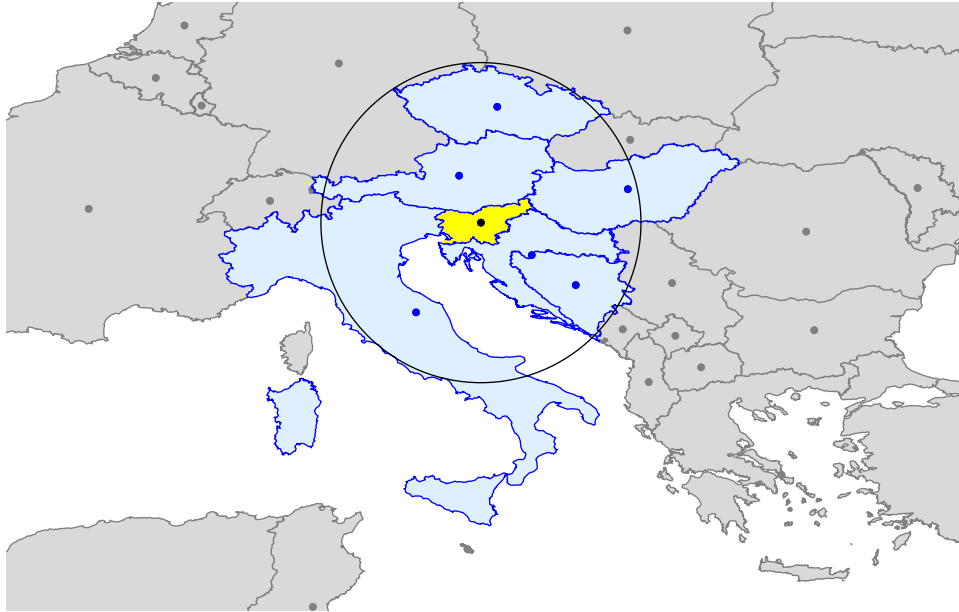
Mreže so lahko geometrijsko oziroma prostorsko pravih oblik, na primer kvadratne ali trikotne, ali pa nepravilnih oblik, kot so na primer geografska območja, občine, države in podobno.

Prostorska informacija je podana topološko. To pomeni, da za posamezne elemente vemo, kateri drugi elementi so obravnavanemu elementu sosed. Na ta način lahko ugotavljamo, v kolikšni meri elementi vplivajo na svoje sosede. Podatki so pripravljene tako, da imamo za vsak element seznam sosedov. Soseda opredelimo kot element, ki ima z obravnavanim skupno mejo. Med sosede lahko vključimo tudi tiste elemente, ki imajo skupno mejo z neposrednim sosedom. Na primer: Avstrija ima s Slovenijo skupno mejo in je po prvem kriteriju Sloveniji sosed. Če uporabimo še drugi kriterij, lahko med sosede štejemo tudi države, ki mejijo na Italijo, Avstrijo, Madžarsko ali Hrvaško, to so na primer Francija, Slovaška, Češka, Nemčija in Srbija (glej sliko 1.4).



SLIKA 1.4: Slovenija s svojimi sosedomi v prvem (modre) in drugem kolenu (zelene)

Sosede lahko definiramo tudi kot elemente, katerih koordinate, na primer geometrijsko središče elementa, so dovolj blizu obravnavanega elementa oziroma njegovega geometrijskega središča. Tak primer je prikazan na sliki 1.5, kjer smo med sosede Slovenije šteli Italijo, San Marino, Avstrijo, Češko, Madžarsko, Hrvaško ter Bosno in Hercegovino.



SLIKA 1.5: Slovenija s svojimi sosedami glede na oddaljenost geometrijskih središč

Tak način podajanja in obravnavanja prostorskih podatkov lahko uporabimo na primer pri obravnavanju pojavov, kot sta širjenje nalezljivih bolezni ali razvoj kriminalnih aktivnosti.



### 1.3.3 Prostorski točkovni vzorci

O prostorskih vzorcih govorimo, ko ima lega oziroma razporeditev točk v prostoru pomembno vlogo in jo želimo obravnavati. V tem primeru je eno prvih vprašanj, kako so točke razporejene, slučajno, v gručah ali urejeno, kar je grafično prikazano na slikah 1.1 in 1.6. Primere take obravnave najdemo v gozdarskih raziskavah, kjer na primer ugotavljajo razporejenost in lastnosti dreves smreke v mešanem gozdu, v bioloških raziskavah rasti celic v pogojih in-vitro in v mnogih drugih.



SLIKA 1.6: Različni vzorci geometrijske porazdelitve dreves: zgoraj slučajno enakomerno, spodaj levo v gručah, spodaj desno urejen vzorec (foto: G. Turk)



## 2 Slučajne funkcije, polja in procesi

V tem poglavju bomo obravnavali osnovne definicije slučajne funkcije, slučajnega polja oziroma procesa. Najprej pa bomo na kratko ponovili definicije, povezane s slučajno spremenljivko, tako da bodo posebnosti, ki so značilne za slučajno funkcijo, ki je osnova prostorske statistike, bolj razumljive.

### 2.1 Slučajne spremenljivke

Slučajno spremenljivko označimo z veliko tiskano črko (na primer  $X, Y, Z$ ), vrednosti, ki jih slučajna spremenljivka lahko zavzame, pa z malimi črkami (na primer  $x, y, z$ ). Dogodke, povezane s slučajno spremenljivko, lahko zapišemo na primer takole:  $Z \leq z$ , kar bi prebrali kot „slučajna spremenljivka  $Z$  zavzame vrednosti manjše ali enake od vrednosti  $z$ “, ali pa  $Z = 4$ , kar bi prebrali kot „slučajna spremenljivka zavzame vrednost 4“.

Pomemben pojem pri obravnavanju slučajnih spremenljivk je **zaloga vrednosti**, ki predstavlja tisto množico vrednosti, ki jih slučajna spremenljivka lahko zavzame. Zaloga vrednosti je lahko diskretna množica vrednosti, v tem primeru govorimo o **diskretni slučajni spremenljivki**. Zaloga vrednosti je lahko tudi množica vrednosti na intervalu ali uniji intervalov, v tem primeru govorimo o **zvezni slučajni spremenljivki**. Primer zaloge vrednosti diskretne slučajne spremenljivke so naravna števila od 1 do 6 ( $z \in \mathbb{N}, 1 \leq z \leq 6$ ), kar lahko predstavlja zalogo vrednosti rezultata meta kocke. Primer zaloge vrednosti zvezne slučajne spremenljivke so pozitivna realna števila ( $z \in \mathbb{R}, z > 0$ ), kar lahko predstavlja čas do naslednjega potresa v Ljubljani.

Za popolno predstavitev slučajne spremenljivke moramo poznati njeno **porazdelitev**. Le-ta je lahko podana na različne načine. S **porazdelitveno funkcijo**, ki je definirana z enačbo:

$$F_Z(z) = P[Z \leq z], \quad (2.1)$$

lahko opišemo porazdelitev poljubne diskretne in zvezne slučajne spremenljivke. Če preberemo enačbo (2.1), je porazdelitvena funkcija  $F_Z(z)$  slučajne spremenljivke  $Z$  definirana kot verjetnost, da slučajna spremenljivka  $Z$  zavzame vrednosti, ki so manjše ali enake vrednosti  $z$ . Pomembni lastnosti porazdelitvene funkcije, ki izvirata iz aksiomov verjetnostnega računa, sta, da je omejena med 0 in 1 ter da je nepadajoča:

$$\begin{aligned}
0 &\leq F_Z(z) \leq 1, \\
\lim_{z \rightarrow -\infty} F_Z(z) &= 0, \\
\lim_{z \rightarrow \infty} F_Z(z) &= 1.
\end{aligned}$$

Porazdelitev diskretne slučajne spremenljivke lahko opišemo z **verjetnostno funkcijo**, ki je definirana z enačbo:

$$p_Z(z_i) = P[Z = z_i]. \quad (2.2)$$

Verjetnostna funkcija  $p_Z(z_i)$  je torej definirana kot verjetnost, da slučajna spremenljivka  $Z$  zavzame vrednost  $z_i$ . Lastnosti verjetnostne funkcije povzamemo z naslednjima izrazoma:

$$\begin{aligned}
0 &\leq p_Z(z_i) \leq 1, \\
\sum_{i=1}^n p_Z(z_i) &= 1,
\end{aligned}$$

kjer je  $n$  število vrednosti v zalogi vrednosti slučajne spremenljivke  $Z$ .

Porazdelitev zvezne slučajne spremenljivke lahko opišemo tudi z **gostoto verjetnosti**  $f_Z(z)$ , s katero lahko izračunamo verjetnost, da slučajna spremenljivka zavzame vrednost na določenem intervalu:

$$P[a < Z \leq b] = \int_a^b f_Z(z) dz. \quad (2.3)$$

Lastnosti gostote verjetnosti lahko povzamemo z naslednjima izrazoma:

$$\begin{aligned}
0 &\leq f_Z(z), \\
\int_{-\infty}^{\infty} f_Z(z) dz &= 1.
\end{aligned}$$

Pomembni sta tudi enačbi, ki povezujeta gostoto verjetnosti in porazdelitveno funkcijo

$$F_Z(z) = \int_{-\infty}^z f_Z(\tilde{z}) d\tilde{z} \quad \longleftrightarrow \quad f_Z(z) = \frac{dF_Z(z)}{dz}.$$

## 2.2 Slučajni vektorji

Če obravnavamo pojav, ki ga moramo opisati z več kot eno spremenljivko, le-te sestavimo v **slučajni vektor**. Slučajni vektor več slučajnih spremenljivk zapišemo takole:

$$\mathbf{Z} = \{Z_1, Z_2, \dots, Z_n\},$$

kjer je  $n$  število slučajnih spremenljivk v slučajnem vektorju oziroma število dimenzij, ki jih slučajni vektor opisuje.

Podobno kot pri slučajni spremenljivki zaloga vrednosti slučajnega vektorja predstavlja vse vrednosti, ki jih slučajna spremenljivka lahko zavzame. Slučajne spremenljivke, ki sestavljajo slučajni vektor, so lahko tako zvezne kot diskretne.

Tudi slučajni vektor v celoti predstavimo z njegovo porazdelitvijo. Porazdelitvena funkcija slučajnega vektorja je definirana z enačbo:

$$F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) = P[Z_1 \leq z_1 \cap Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n]. \quad (2.4)$$

Nekaj pomembnih lastnosti porazdelitvene funkcije slučajnega vektorja povzamemo z izrazi:

$$\begin{aligned} 0 &\leq F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) \leq 1, \\ \lim_{z_i \rightarrow -\infty} F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= 0 \text{ za vsak } i = 1, \dots, n, \\ \lim_{z_1 \rightarrow \infty \cap z_2 \rightarrow \infty \cap \dots \cap z_n \rightarrow \infty} F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= 1, \\ \lim_{z_1 \rightarrow \infty} F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= F_{Z_2, Z_3, \dots, Z_n}(z_2, z_3, \dots, z_n). \end{aligned}$$

Z zadnjim izrazom definiramo porazdelitev dela slučajnega vektorja, taki porazdelitveni funkciji pravimo **robna porazdelitvena funkcija**. Podobno kot pri slučajni spremenljivki definiramo tudi verjetnostno funkcijo in gostoto verjetnosti slučajnega vektorja. Verjetnostna funkcija je:

$$p_{Z_1, Z_2, \dots, Z_n}(z_{1i}, z_{2j}, \dots, z_{nk}) = P[Z_1 = z_{1i} \cap Z_2 = z_{2j} \cap \dots \cap Z_n = z_{nk}], \quad (2.5)$$

z gostoto verjetnosti pa lahko izračunamo verjetnost, da slučajni vektor leži na določenem intervalu:

$$\begin{aligned} P[a_1 < Z_1 \leq b_1 \cap a_2 < Z_2 \leq b_2 \cap \dots \cap a_n < Z_n \leq b_n] &= \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} f_{Z_1, Z_2, \dots, Z_n} dz_1 dz_2 \dots dz_n. \end{aligned} \quad (2.6)$$

Pri slučajnem vektorju lahko definiramo tudi pogojno porazdelitev, ki predstavlja porazdelitev slučajnega vektorja ob pogoju, da so za nekatere slučajne spremenljivke v slučajnem vektorju izpolnjeni določeni pogoji. Na primer, **pogojno porazdelitveno funkcijo** definiramo kot:

$$\begin{aligned} F_{Z_1 | (Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n)}(z_1) &= P[(Z_1 \leq z_1) | (Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n)] = \\ &= \frac{P[Z_1 \leq z_1 \cap Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n]}{P[Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n]} = \frac{F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n)}{F_{Z_2, \dots, Z_n}(z_2, \dots, z_n)}. \end{aligned} \quad (2.7)$$

V primeru, da je slučajna spremenljivka  $Z_1$  **neodvisna** od drugih slučajnih spremenljivk v slučajnem vektorju, lahko zapišemo:

$$\begin{aligned} F_{Z_1 | (Z_2 \leq z_2 \cap \dots \cap Z_n \leq z_n)}(z_1) &= F_{Z_1}(z_1) \longrightarrow \\ \longrightarrow F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) &= F_{Z_1}(z_1) F_{Z_2, \dots, Z_n}(z_2, \dots, z_n), \end{aligned} \quad (2.8)$$

kar pomeni, da lahko porazdelitveno funkcijo slučajnega vektorja, sestavljenega iz neodvisnih slučajnih spremenljivk, zapišemo kot produkt robnih porazdelitvenih funkcij. Torej, če so vse slučajne spremenljivke medsebojno neodvisne, lahko porazdelitveno funkcijo slučajnega vektorja zapišemo kot produkt vseh porazdelitvenih funkcij slučajnih spremenljivk, ki sestavljajo vektor:

$$F_{Z_1, Z_2, \dots, Z_n}(z_1, z_2, \dots, z_n) = F_{Z_1}(z_1)F_{Z_2}(z_2) \dots F_{Z_n}(z_n). \quad (2.9)$$

## 2.3 Momenti slučajnih spremenljivk in vektorjev

**Momenti** in **centralni momenti** slučajnih spremenljivk opisujejo lastnosti slučajnih spremenljivk in vektorjev. Najpomembnejši momenti slučajnih spremenljivk in vektorjev so: pričakovana vrednost, varianca in kovarianca ter simetričnost in sploščenost. V nadaljevanju bomo obravnavali le momente zveznih slučajnih spremenljivk in vektorjev.

**Pričakovana vrednost** slučajne spremenljivke  $Z$  je moment prvega reda, definiran z izrazom:

$$m_Z = E[Z] = \int_{-\infty}^{\infty} z f_Z(z) dz, \quad (2.10)$$

in predstavlja osnovno informacijo o vrednosti slučajne spremenljivke. S črko  $E$  smo označili pričakovano vrednost, povzeto po angleških virih (angl. expected value).

**Varianca** slučajne spremenljivke  $Z$  je centralni moment drugega reda, definiran z izrazom:

$$\sigma_Z^2 = \text{var}[Z] = \int_{-\infty}^{\infty} (z - m_Z)^2 f_Z(z) dz = E[Z^2] - E[Z]^2 \quad (2.11)$$

in predstavlja osnovno informacijo o razpršenosti slučajne spremenljivke. Centralna momenta tretje in četrte stopnje predstavljata meri za simetričnost in sploščenost:

$$\begin{aligned} \mu_Z^{(3)} &= \int_{-\infty}^{\infty} (z - m_Z)^3 f_Z(z) dz, & \text{mera simetričnosti,} \\ \mu_Z^{(4)} &= \int_{-\infty}^{\infty} (z - m_Z)^4 f_Z(z) dz, & \text{mera sploščenosti.} \end{aligned} \quad (2.12)$$

Brezdimenzijska koeficienta simetričnosti in sploščenosti definiramo z enačbama:

$$\begin{aligned} \gamma_{1Z} &= \frac{\mu_Z^{(3)}}{\sigma_Z^3}, & \text{koeficient simetričnosti,} \\ \gamma_{2Z} &= \frac{\mu_Z^{(4)}}{\sigma_Z^4}, & \text{kurtosis.} \end{aligned} \quad (2.13)$$

**Kovarianca** med slučajnima spremenljivkama  $Z_1$  in  $Z_2$  je centralni moment drugega reda, definiran z izrazom:

$$\begin{aligned}\sigma_{Z_1, Z_2} = \text{cov}[Z_1, Z_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z_1})(z_2 - m_{Z_2}) f_{Z_1, Z_2}(z_1, z_2) dz_1 dz_2 \\ &= E[Z_1 Z_2] - E[Z_1]E[Z_2],\end{aligned}\quad (2.14)$$

in predstavlja mero za linearno povezanost med dvema slučajnima spremenljivkama. Brezdimenzijska mera linearne povezanosti je **korelacijski koeficient**  $\rho_{Z_1, Z_2}$ , ki je definiran z enačbo:

$$\rho_{Z_1, Z_2} = \frac{\sigma_{Z_1, Z_2}}{\sigma_{Z_1} \sigma_{Z_2}} \quad (2.15)$$

in lahko zavzame vrednosti med  $-1$  in  $1$ :

$$-1 \leq \rho_{Z_1, Z_2} \leq 1.$$

## 2.4 Slučajne funkcije

Slučajna funkcija je slučajna spremenljivka ali slučajni vektor, ki je odvisna tudi od lege v prostoru in časa. Če je slučajna funkcija odvisna le od lege v prostoru, ji pravimo **slučajno polje**. Če je slučajna funkcija odvisna tudi od časa, ji rečemo **slučajni proces**. Omejili se bomo le na zvezna slučajna polja in procese, saj le-ti predstavljajo pomembnejši del prostorske statistike. Skalarno slučajno polje zapišemo kot:

$$Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^n, \quad (2.16)$$

skalarni proces pa kot:

$$Z(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^n, t \in \mathbb{R}, t > 0, \quad \text{oziroma} \quad Z(t), t \in \mathbb{R}, t > 0, \quad (2.17)$$

kjer  $\mathbf{s}$  predstavlja krajevni vektor do točke, kjer zapišemo obravnavano slučajno polje,  $t$  pa čas, pri katerem obravnavamo slučajni proces. Definirali smo torej slučajno spremenljivko  $Z(\mathbf{s})$  oziroma  $Z(\mathbf{s}, t)$  ali  $Z(t)$ , ki jo lahko obravnavamo podobno kot običajno slučajno spremenljivko. Pomen zaloge vrednosti je povsem enak kot pri slučajni spremenljivki, to so vrednosti, ki jih lahko ima slučajno polje. Primer slučajnega polja so trenutne temperature v Sloveniji, kjer lahko ugotovimo, da so tako pričakovane vrednosti kot tudi varianca in porazdelitev različni v različnih točkah.

Najprej bomo obravnavali le skalarna slučajna polja, torej bomo obravnavali le slučajno spremenljivko, odvisno od lege. **Porazdelitvena funkcija slučajnega polja**  $F_{Z(\mathbf{s})}(z)$  je:

$$F_{Z(\mathbf{s})}(z) = P[Z(\mathbf{s}) \leq z] \quad (2.18)$$

in pomeni verjetnost, da slučajno polje v točki, ki je določena s krajevnim vektorjem  $\mathbf{s}$ , zavzame vrednost, manjšo ali enako  $z$ . Podobno kot moramo pri slučajnem vektorju  $\{Z_1, Z_2\}$  poznati porazdelitveno

funkcijo  $F_{Z_1, Z_2}(z_1, z_2)$  in ne le robnih porazdelitvenih funkcij  $F_{Z_1}(z_1)$  in  $F_{Z_2}(z_2)$ , moramo za celovit opis skalarnega slučajnega polja definirati porazdelitveno funkcijo slučajnega polja  $Z(\mathbf{s})$  za dve različni točki  $\mathbf{s}_1$  in  $\mathbf{s}_2$ :

$$F_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)}(z_1, z_2) = P[Z(\mathbf{s}_1) \leq z_1 \cap Z(\mathbf{s}_2) \leq z_2], \quad (2.19)$$

kar je povsem analogno definiciji porazdelitvene funkcije slučajnega vektorja po enačbi (2.4). V primeru slučajnega vektorja smo porazdelitveno funkcijo definirali za dve ali več slučajnih spremenljivk, pri slučajnem polju pa smo v (2.19) definirali porazdelitveno funkcijo za isto slučajno spremenljivko v dveh različnih točkah. **Gostoto verjetnosti slučajnega polja** definiramo kot odvod porazdelitvene funkcije:

$$f_{Z(\mathbf{s})}(z) = \frac{dF_{Z(\mathbf{s})}(z)}{dz} \quad (2.20)$$

oziroma:

$$f_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)}(z_1, z_2) = \frac{\partial^2 F_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)}(z_1, z_2)}{\partial z_1 \partial z_2}. \quad (2.21)$$

Podobno kot za slučajne spremenljivke lahko določimo še nekaj osnovnih momentov, s katerimi opišemo osnovne lastnosti slučajnega polja. **Srednja vrednost slučajnega polja** je:

$$m_{Z(\mathbf{s})} = E[Z(\mathbf{s})] = \int_{-\infty}^{\infty} z f_{Z(\mathbf{s})}(z) dz, \quad (2.22)$$

**varianca slučajnega polja** pa je:

$$\sigma_{Z(\mathbf{s})}^2 = \text{var}[Z(\mathbf{s})] = \int_{-\infty}^{\infty} (z - m_{Z(\mathbf{s})})^2 f_{Z(\mathbf{s})}(z) dz = E[Z(\mathbf{s})^2] - E[Z(\mathbf{s})]^2 = C(\mathbf{s}). \quad (2.23)$$

V primeru slučajnega polja lahko določimo še funkcije, ki jih pri slučajnih spremenljivkah nismo imeli. To so **kovariančna funkcija**, **korelacijska funkcija** in **variogram**. Kovariančna funkcija predstavlja kovarianco med slučajnim poljem v dveh različnih točkah  $\mathbf{s}_1$  in  $\mathbf{s}_2$ :

$$\begin{aligned} \sigma_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)} &= \text{cov}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z(\mathbf{s}_1)}) (z_2 - m_{Z(\mathbf{s}_2)}) f_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)}(z_1, z_2) dz_1 dz_2 = \\ &= E[Z(\mathbf{s}_1)Z(\mathbf{s}_2)] - E[Z(\mathbf{s}_1)] E[Z(\mathbf{s}_2)] = C(\mathbf{s}_1, \mathbf{s}_2). \end{aligned} \quad (2.24)$$

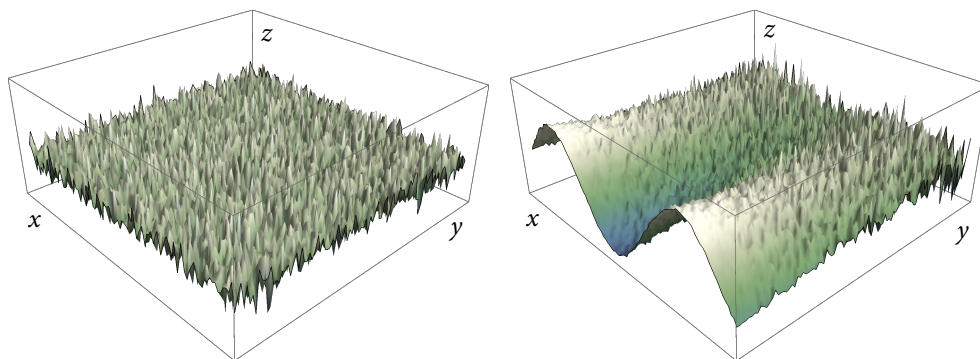
V zadnjih dveh enačbah smo vpeljali oznaki za varianco  $C(\mathbf{s})$  in kovariančno funkcijo  $C(\mathbf{s}_1, \mathbf{s}_2)$ , ki sta običajni pri prostorski statistiki. Korelacijsko funkcijo določimo na enak način kot korelacijski koeficient pri slučajnih spremenljivkah:

$$\rho(\mathbf{s}_1, \mathbf{s}_2) = \frac{\text{cov}[Z(\mathbf{s}_1), Z(\mathbf{s}_2)]}{\sqrt{\text{var}[Z(\mathbf{s}_1)] \text{var}[Z(\mathbf{s}_2)]}} = \frac{C(\mathbf{s}_1, \mathbf{s}_2)}{\sqrt{C(\mathbf{s}_1) C(\mathbf{s}_2)}}. \quad (2.25)$$

**Variogram** oziroma **semivariogram** je definiran kot polovica variance razlike vrednosti slučajnega polja v dveh različnih točkah. Zaradi take definicije (polovica variance) sta se uveljavili obe poimenovanji, a je prvo bolj pogosto in bo dosledno uporabljeno v nadaljevanju učbenika.

$$2\gamma(\mathbf{s}_1, \mathbf{s}_2) = \text{var}[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)] = E\left[(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2\right] - E[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)]^2. \quad (2.26)$$

Če se lastnosti slučajnega polja po prostoru ne spreminjajo, pomeni, da obravnavamo **stacionarno slučajno polje**. Primera stacionarnega in nestacionarnega slučajnega polja prikazujemo na sliki 2.1, kjer s koordinatama  $x, y$  opišemo lego točke,  $z$  pa opisuje vrednost slučajnega polja.



SLIKA 2.1: Stacionarno (levo) in nestacionarno (desno) slučajno polje

**Stacionarnost** slučajnega polja opredelimo s pogojem, da je porazdelitvena funkcija za eno ali dve točki neodvisna od lege točke, določene s krajevnim vektorjem  $\mathbf{s}$ :

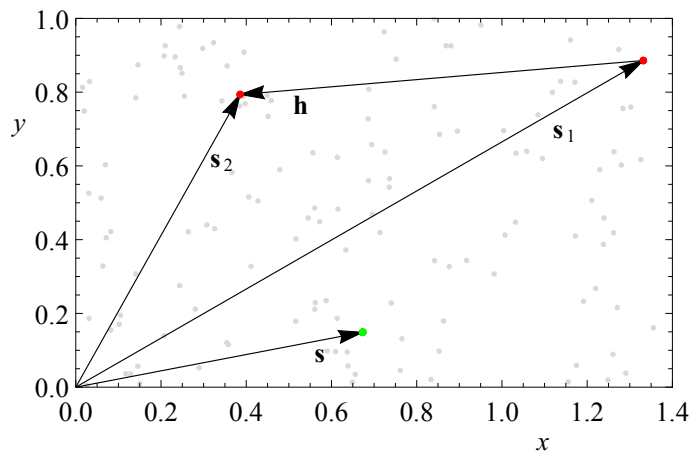
$$\begin{aligned} F_{Z(\mathbf{s})}(z) &= P[Z(\mathbf{s}) \leq z] = F_Z(z), \\ F_{Z(\mathbf{s}_1), Z(\mathbf{s}_2)}(z_1, z_2) &= P[Z(\mathbf{s}) \leq z_1 \cap Z(\mathbf{s} + \mathbf{h}) \leq z_2] = F_{Z(\mathbf{h})}(z_1, z_2). \end{aligned} \quad (2.27)$$

Če veljata pogoja (2.27), sta od lege neodvisni tudi srednja vrednost in varianca, medtem ko sta kovariančna funkcija in variogram odvisni le od vektorja  $\mathbf{h}$ , ki predstavlja vektor med dvema točkama (glej sliko 2.2):

$$\mathbf{s}_2 = \mathbf{s}_1 + \mathbf{h}, \quad \mathbf{h} = \mathbf{s}_2 - \mathbf{s}_1. \quad (2.28)$$

Zapišemo enačbe za srednjo vrednost, varianco, kovariančno funkcijo in variogram za stacionarna polja:

$$\begin{aligned} m_{Z(\mathbf{s})} &= E[Z(\mathbf{s})] = m_Z, \\ \sigma_{Z(\mathbf{s})}^2 &= \text{var}[Z(\mathbf{s})] = \sigma_Z^2 = C(\mathbf{0}), \\ C(\mathbf{s}_1, \mathbf{s}_2) &= C(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = C(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(\mathbf{h}), \\ \gamma(\mathbf{s}_1, \mathbf{s}_2) &= \gamma(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = \gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \gamma(\mathbf{h}), \\ \rho(\mathbf{h}) &= \frac{C(\mathbf{h})}{\sqrt{C(\mathbf{0})C(\mathbf{0})}} = \frac{C(\mathbf{h})}{C(\mathbf{0})}. \end{aligned} \quad (2.29)$$



SLIKA 2.2: Krajevni vektorji do nekaterih točk v prostoru

Varianco  $\sigma_z^2$  smo zapisali kot  $C(0)$ , saj s primerjavo enačb (2.22) in (2.24) vidimo, da lahko varianco zapišemo kot kovariančno funkcijo, zapisano za isto točko:

$$\sigma_{Z(\mathbf{s})}^2 = E[Z(\mathbf{s})Z(\mathbf{s})] - E[Z(\mathbf{s})]E[Z(\mathbf{s})] = C(\mathbf{s}, \mathbf{s}) = C(\mathbf{s}, \mathbf{s} + \mathbf{0}) = C(\mathbf{0}). \quad (2.30)$$

Za stacionarno polje lahko izpeljemo tudi zvezo med variogramom in kovariančno funkcijo. Zapišemo enačbo variograma, izraženo s pričakovanimi vrednostmi:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E\left[(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2\right] - E[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})]^2 = \\ &\quad \text{(v desnem izrazu upoštevamo linearnost operatorja } E[\cdot]) \\ &= E[Z(\mathbf{s})^2 - 2Z(\mathbf{s})Z(\mathbf{s} + \mathbf{h}) + Z(\mathbf{s} + \mathbf{h})^2] - (E[Z(\mathbf{s})] - E[Z(\mathbf{s} + \mathbf{h})])^2 = \\ &\quad \text{(v levem izrazu upoštevamo linearnost operatorja } E[\cdot], \\ &\quad \text{zaradi stacionarnosti polja je } E[Z(\mathbf{s})] = E[Z(\mathbf{s} + \mathbf{h})] = m_Z, \\ &\quad \text{zato se zadnja dva člena odštejeta)} \\ &= E[Z(\mathbf{s})^2] - m_Z^2 - 2E[Z(\mathbf{s})Z(\mathbf{s} + \mathbf{h})] + 2m_Z^2 + E[Z(\mathbf{s} + \mathbf{h})^2] - m_Z^2 = \\ &\quad \text{(prišteli in odšteli smo } 2m_Z^2, \\ &\quad \text{da bomo lahko tvorili varianco in kovariančno funkcijo)} \\ &= C(\mathbf{0}) - 2C(\mathbf{h}) + C(\mathbf{0}) = 2C(\mathbf{0}) - 2C(\mathbf{h}). \end{aligned}$$

Sedaj lahko zapišemo končni rezultat:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad \text{in} \quad C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h}). \quad (2.31)$$



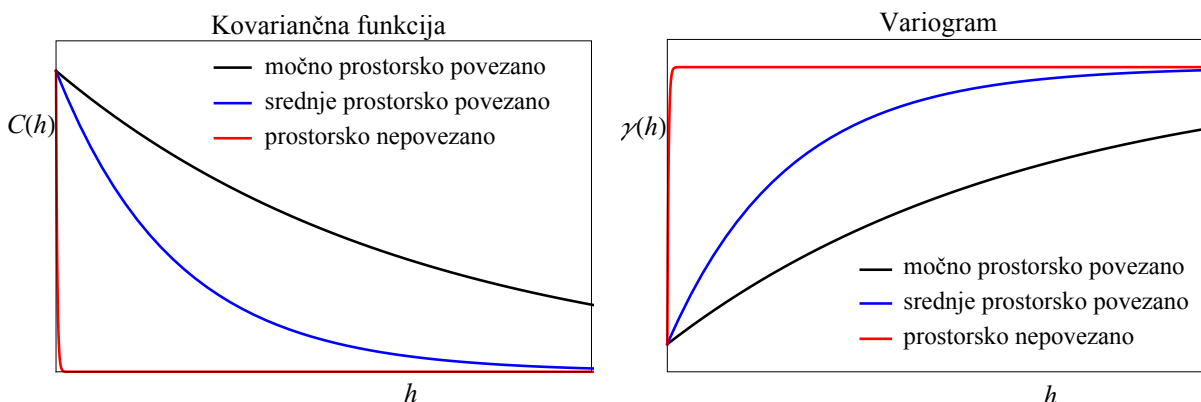
Zapišemo lahko tudi zvezo med korelacijsko funkcijo in variogramom:

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})} = \frac{C(\mathbf{0}) - \gamma(\mathbf{h})}{C(\mathbf{0})} = 1 - \frac{\gamma(\mathbf{h})}{C(\mathbf{0})}. \quad (2.32)$$

Za običajna slučajna polja velja, da kovariančna funkcija pada z naraščanjem razdalje med točkami, obratno pa variogram z naraščanjem razdalje med točkama običajno narašča in se približuje vrednosti  $C(0)$ . Zapišemo lahko:

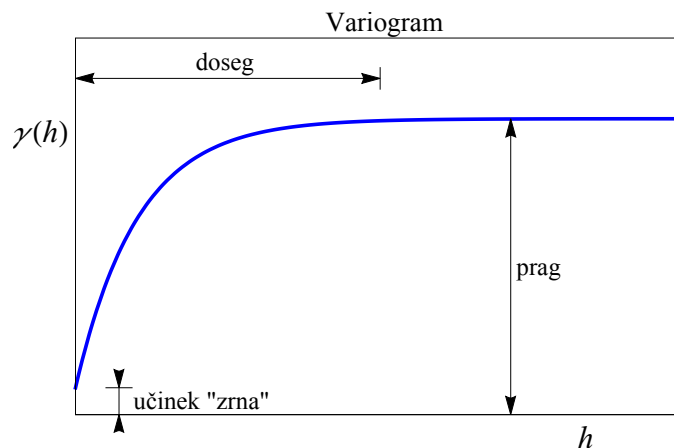
$$\lim_{|\mathbf{h}| \rightarrow \infty} C(\mathbf{h}) = 0 \quad \text{in} \quad \lim_{|\mathbf{h}| \rightarrow \infty} \gamma(\mathbf{h}) = C(\mathbf{0}). \quad (2.33)$$

Nekaj značilnih kovariančnih funkcij in variogramov prikazujemo na sliki 2.3.



SLIKA 2.3: Kovariančna funkcija in variogram

V grafu variograma na sliki 2.4 lahko prepoznamo nekaj značilnih lastnosti. Variogram običajno narašča, saj pričakujemo, da bo varianca razlike med vrednostima slučajnega polja z razdaljo naraščala do neke meje, ki jo imenujemo **prag** (ang. sill). Opazimo lahko tudi, da je prag dosežen na neki določeni razdalji, do katere lahko prepoznamo prostorsko odvisnost. Tej razdalji rečemo **doseg** (ang. range) slučajnega polja in predstavlja razdaljo med točkama, kjer vrednost v eni točki še opazno vpliva na vrednost druge točke. Zadnja in morda najbolj presenetljiva lastnost variograma je tako imenovan **učinek zrna** (ang. nugget effect), ki odraža variabilnost vrednosti slučajnega polja v točki. Pričakovali bi, da je varianca razlike med vrednostima slučajnega polja v isti točki enaka nič, a je zaradi merskih napak in možnih nezveznosti le-ta od nič različna. Ime tega učinka izvira iz rudarstva, ki je bilo eno od področij, v okviru katerega se je prostorska statistika hitro razvijala. Predstavljamo si zrno zlata, ki ima seveda povsem druge lastnosti kot kamnina, v kateri zrno leži. Če bi torej merili lastnosti na meji med zrnem in kamnino, bi morali ugotoviti nezveznost oziroma variabilnost vrednosti v isti točki.



SLIKA 2.4: Variogram in njegove značilnosti

## 2.5 Vektorske slučajne funkcije

Kot je omenjeno v razdelku [Geostatistični podatki](#) in bomo podrobneje spoznali v [naslednjem poglavju](#), lahko podatki obsegajo vrednosti več slučajnih spremenljivk. Tako na primer geostatistični podatki s področja Jura v Švici, ki jih bomo uporabili v različnih analizah prostorske statistike, poleg opisnih podatkov, kot so raba tal in tip kamnine, vsebuje tudi numerične podatke o koncentraciji različnih kovin (Cd, Cu, Pb, Co, Cr, Ni in Zn) v izbranih točkah. To je zelo značilen primer slučajnega vektorskega polja oziroma slučajne vektorske funkcije.

Vektorsko slučajno polje zapišemo kot:

$$\{Z_1(\mathbf{s}), Z_2(\mathbf{s}), Z_3(\mathbf{s}), \dots, Z_N(\mathbf{s})\} \in \mathbb{R}^N, \mathbf{s} \in \mathbb{R}^n, \quad (2.34)$$

vektorski slučajni proces pa kot:

$$\begin{aligned} \{Z_1(\mathbf{s}, t), Z_2(\mathbf{s}, t), Z_3(\mathbf{s}, t), \dots, Z_N(\mathbf{s}, t)\} &\in \mathbb{R}^N, \mathbf{s} \in \mathbb{R}^n, t \in \mathbb{R}, t > 0, \quad \text{oziroma} \\ \{Z_1(t), Z_2(t), Z_3(t), \dots, Z_N(t)\} &\in \mathbb{R}^N, t \in \mathbb{R}, t > 0, \end{aligned} \quad (2.35)$$

kjer  $\mathbf{s}$  predstavlja krajevni vektor v  $n$ -razsežnem prostoru do točke, kjer zapišemo obravnavano slučajno polje,  $t$  pa čas, pri katerem obravnavamo slučajni proces. Število  $N$  predstavlja razsežnost slučajnega vektorskega polja, torej število slučajnih spremenljivk, ki so vključene v slučajni vektor oziroma slučajno vektorsko funkcijo. V primeru podatkov iz Jure obravnavamo sedem različnih slučajnih spremenljivk ( $N = 7$ ) v dvorazsežnem prostoru geografske dolžine in širine ( $n = 2$ ).

Zaradi preprostejšega zapisa bomo v nadaljevanju obravnavali le vektorska slučajna polja z dvema slučajnima spremenljivkama ( $N = 2$ ), saj v takem polju lahko opišemo vse pomembne značilnosti

analize vektorskih polj in je posplošitev na večje število slučajnih spremenljivk relativno preprosta. Zapišimo vektorsko polje dveh slučajnih spremenljivk:

$$\{Z_1(\mathbf{s}), Z_2(\mathbf{s})\}, \mathbf{s} \in \mathbb{R}^n. \quad (2.36)$$

Ker obravnavamo zvezno vektorsko slučajno polje, je zaloga vrednosti neko zvezno območje na  $\mathbb{R}^2$ .

**Porazdelitvena funkcija vektorskega slučajnega polja**  $F_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2)$  je:

$$F_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) = P[Z_1(\mathbf{s}) \leq z_1 \cap Z_2(\mathbf{s}) \leq z_2] \quad (2.37)$$

in pomeni verjetnost, da slučajno vektorsko polje v točki  $\mathbf{s}$ , zavzame vrednost na območju  $Z_1 \leq z_1 \cap Z_2 \leq z_2$ . Podobno kot pri skalarnem slučajnem polju lahko v analizo vpeljemo dve točki, določeni z vektorjema  $\mathbf{s}_1$  in  $\mathbf{s}_2$ , in definiramo novo porazdelitveno funkcijo:

$$F_{Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)}(z_1, z_2) = P[Z_1(\mathbf{s}_1) \leq z_1 \cap Z_2(\mathbf{s}_2) \leq z_2], \quad (2.38)$$

s čimer postavljamo osnovo za obravnavo dveh različnih slučajnih spremenljivk  $Z_1$  in  $Z_2$  v dveh različnih točkah  $\mathbf{s}_1$  in  $\mathbf{s}_2$ .

**Gostoto verjetnosti vektorskega slučajnega polja** definiramo kot odvod porazdelitvene funkcije:

$$f_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) = \frac{\partial^2 F_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2)}{\partial z_1 \partial z_2}, \quad (2.39)$$

oziroma

$$f_{Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)}(z_1, z_2) = \frac{\partial^2 F_{Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)}(z_1, z_2)}{\partial z_1 \partial z_2}. \quad (2.40)$$

Podobno kot za skalarno slučajno polje lahko za vektorsko slučajno polje določimo osnovne momente, s katerimi opišemo osnovne lastnosti vektorskega slučajnega polja. **Srednji vrednosti komponent vektorskega slučajnega polja** sta

$$\begin{aligned} m_{Z_1(\mathbf{s})} &= E[Z_1(\mathbf{s})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_1 f_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) dz_1 dz_2, \\ m_{Z_2(\mathbf{s})} &= E[Z_2(\mathbf{s})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_2 f_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) dz_1 dz_2, \end{aligned} \quad (2.41)$$

**varianci komponent vektorskega slučajnega polja** pa sta

$$\begin{aligned} \sigma_{Z_1(\mathbf{s})}^2 &= \text{var}[Z_1(\mathbf{s})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z_1(\mathbf{s})})^2 f_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) dz_1 dz_2, \\ \sigma_{Z_2(\mathbf{s})}^2 &= \text{var}[Z_2(\mathbf{s})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_2 - m_{Z_2(\mathbf{s})})^2 f_{Z_1(\mathbf{s}), Z_2(\mathbf{s})}(z_1, z_2) dz_1 dz_2, \end{aligned} \quad (2.42)$$

ki ju lahko zapišemo tudi takole:

$$\begin{aligned}\sigma_{Z_1(\mathbf{s})}^2 &= E[Z_1(\mathbf{s})^2] - E[Z_1(\mathbf{s})]^2 = C_1(\mathbf{s}), \\ \sigma_{Z_2(\mathbf{s})}^2 &= E[Z_2(\mathbf{s})^2] - E[Z_2(\mathbf{s})]^2 = C_2(\mathbf{s}).\end{aligned}\quad (2.43)$$

Analogno kot pri skalarnem slučajnem polju določimo tudi kovariančni funkciji in variograma za komponenti vektorskega slučajnega polja:

$$\begin{aligned}\sigma_{Z_1(\mathbf{s}_1), Z_1(\mathbf{s}_2)} &= cov[Z_1(\mathbf{s}_1), Z_1(\mathbf{s}_2)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z_1(\mathbf{s}_1)}) (z_2 - m_{Z_1(\mathbf{s}_2)}) f_{Z_1(\mathbf{s}_1), Z_1(\mathbf{s}_2)}(z_1, z_2) dz_1 dz_2 = \\ &= E[Z_1(\mathbf{s}_1)Z_1(\mathbf{s}_2)] - E[Z_1(\mathbf{s}_1)] E[Z_1(\mathbf{s}_2)] = C_1(\mathbf{s}_1, \mathbf{s}_2),\end{aligned}\quad (2.44)$$

$$\begin{aligned}\sigma_{Z_2(\mathbf{s}_1), Z_2(\mathbf{s}_2)} &= cov[Z_2(\mathbf{s}_1), Z_2(\mathbf{s}_2)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z_2(\mathbf{s}_1)}) (z_2 - m_{Z_2(\mathbf{s}_2)}) f_{Z_2(\mathbf{s}_1), Z_2(\mathbf{s}_2)}(z_1, z_2) dz_1 dz_2 = \\ &= E[Z_2(\mathbf{s}_1)Z_2(\mathbf{s}_2)] - E[Z_2(\mathbf{s}_1)] E[Z_2(\mathbf{s}_2)] = C_2(\mathbf{s}_1, \mathbf{s}_2),\end{aligned}\quad (2.45)$$

$$2\gamma_1(\mathbf{s}_1, \mathbf{s}_2) = var[Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2)] = E[(Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2))^2] - E[Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2)]^2, \quad (2.46)$$

$$2\gamma_2(\mathbf{s}_1, \mathbf{s}_2) = var[Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2)] = E[(Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2))^2] - E[Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2)]^2. \quad (2.47)$$

Pri vektorskem slučajnem polju lahko definiramo še **križno kovariančno funkcijo**, ki je definirana kot kovarianca slučajne spremenljivke  $Z_1$  v točki  $\mathbf{s}_1$  s slučajno spremenljivko  $Z_2$  v točki  $\mathbf{s}_2$ :

$$\begin{aligned}\sigma_{Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)} &= cov[Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1 - m_{Z_1(\mathbf{s}_1)}) (z_2 - m_{Z_2(\mathbf{s}_2)}) f_{Z_1(\mathbf{s}_1), Z_2(\mathbf{s}_2)}(z_1, z_2) dz_1 dz_2 = \\ &= E[Z_1(\mathbf{s}_1)Z_2(\mathbf{s}_2)] - E[Z_1(\mathbf{s}_1)] E[Z_2(\mathbf{s}_2)] = C_{12}(\mathbf{s}_1, \mathbf{s}_2)\end{aligned}\quad (2.48)$$

in **križni variogram** oziroma **kovariogram**, ki je definiran kot kovarianca razlik  $(Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2))$  in  $(Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2))$ :

$$\begin{aligned}2\gamma_{12}(\mathbf{s}_1, \mathbf{s}_2) &= cov[(Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2)), (Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2))] = \\ &= E[(Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2))(Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2))] - \\ &= E[Z_1(\mathbf{s}_1) - Z_1(\mathbf{s}_2)] E[Z_2(\mathbf{s}_1) - Z_2(\mathbf{s}_2)].\end{aligned}\quad (2.49)$$

Podobno kot pri skalarnih slučajnih poljih lahko tudi pri vektorskih slučajnih poljih opredelimo stacionarnost, za taka vektorska slučajna polja velja:

$$\begin{aligned}
m_{Z_1(\mathbf{s})} &= E[Z_1(\mathbf{s})] = m_{Z_1}, & m_{Z_2(\mathbf{s})} &= E[Z_2(\mathbf{s})] = m_{Z_2}, \\
\sigma_{Z_1(\mathbf{s})}^2 &= \text{var}[Z_1(\mathbf{s})] = \sigma_{Z_1}^2 = C_1(\mathbf{0}), & \sigma_{Z_2(\mathbf{s})}^2 &= \text{var}[Z_2(\mathbf{s})] = \sigma_{Z_2}^2 = C_2(\mathbf{0}), \\
C_1(\mathbf{s}_1, \mathbf{s}_2) &= C_1(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = C_1(\mathbf{h}), & C_2(\mathbf{s}_1, \mathbf{s}_2) &= C_2(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = C_2(\mathbf{h}) \\
\gamma_1(\mathbf{s}_1, \mathbf{s}_2) &= \gamma_1(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = \gamma_1(\mathbf{h}), & \gamma_2(\mathbf{s}_1, \mathbf{s}_2) &= \gamma_2(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = \gamma_2(\mathbf{h}), \\
C_{12}(\mathbf{s}_1, \mathbf{s}_2) &= C_{12}(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = C_{12}(\mathbf{h}), & \gamma_{12}(\mathbf{s}_1, \mathbf{s}_2) &= \gamma_{12}(\mathbf{s}_1, \mathbf{s}_1 + \mathbf{h}) = \gamma_{12}(\mathbf{h}).
\end{aligned} \tag{2.50}$$

Z upoštevanjem enačbe (2.31) lahko zapišemo zveze med variogramoma  $\gamma_1(\mathbf{h})$ ,  $\gamma_2(\mathbf{h})$  in korelacijskima funkcijama  $C_1(\mathbf{h})$ ,  $C_2(\mathbf{h})$ :

$$\gamma_1(\mathbf{h}) = C_1(\mathbf{0}) - C_1(\mathbf{h}) \quad \text{in} \quad \gamma_2(\mathbf{h}) = C_2(\mathbf{0}) - C_2(\mathbf{h}). \tag{2.51}$$

Za stacionarno vektorsko polje lahko izpeljemo tudi zvezo med kovariogramom in križno kovariančno funkcijo. Zapišemo enačbo kovariograma, izraženo s pričakovanimi vrednostmi:

$$\begin{aligned}
2\gamma_{12}(\mathbf{h}) &= E\left[(Z_1(\mathbf{s}) - Z_1(\mathbf{s} + \mathbf{h}))(Z_2(\mathbf{s}) - Z_2(\mathbf{s} + \mathbf{h}))\right] - \\
&\quad - E[Z_1(\mathbf{s}) - Z_1(\mathbf{s} + \mathbf{h})]E[Z_2(\mathbf{s}) - Z_2(\mathbf{s} + \mathbf{h})] = \\
&\quad \text{(upoštevamo linearnost operatorja } E[\cdot]) \\
&= E[Z_1(\mathbf{s})Z_2(\mathbf{s}) - Z_1(\mathbf{s})Z_2(\mathbf{s} + \mathbf{h}) - Z_1(\mathbf{s} + \mathbf{h})Z_2(\mathbf{s}) + Z_1(\mathbf{s} + \mathbf{h})Z_2(\mathbf{s} + \mathbf{h})] - \\
&\quad - (E[Z_1(\mathbf{s})] - E[Z_1(\mathbf{s} + \mathbf{h})])(E[Z_2(\mathbf{s})] - E[Z_2(\mathbf{s} + \mathbf{h})]) = \\
&\quad \text{(upoštevamo linearnost operatorja } E[\cdot]), \\
&\quad \text{zaradi stacionarnosti polja je } E[Z_1(\mathbf{s})] = E[Z_1(\mathbf{s} + \mathbf{h})] \text{ in} \\
&\quad E[Z_2(\mathbf{s})] = E[Z_2(\mathbf{s} + \mathbf{h})], \text{ zato je drugi del izraza enak nič} \\
&= E[Z_1(\mathbf{s})Z_2(\mathbf{s})] - m_{Z_1}m_{Z_2} - E[Z_1(\mathbf{s})Z_2(\mathbf{s} + \mathbf{h})] + m_{Z_1}m_{Z_2} - \\
&\quad - E[Z_1(\mathbf{s} + \mathbf{h})Z_2(\mathbf{s})] + m_{Z_1}m_{Z_2} + E[Z_1(\mathbf{s} + \mathbf{h})Z_2(\mathbf{s} + \mathbf{h})] - m_{Z_1}m_{Z_2} = \\
&\quad \text{(prišteli in odšteli smo } 2m_{Z_1}m_{Z_2}, \\
&\quad \text{da bomo lahko tvorili kovariance in križno kovariančno funkcijo)} \\
&= C_{12}(\mathbf{0}) - C_{12}(\mathbf{h}) - C_{21}(\mathbf{h}) + C_{12}(\mathbf{0}).
\end{aligned}$$

Sedaj lahko zapišemo končni rezultat:

$$\gamma_{12}(\mathbf{h}) = C_{12}(\mathbf{0}) - \frac{1}{2}(C_{12}(\mathbf{h}) + C_{21}(\mathbf{h})). \tag{2.52}$$

Običajno predpostavimo simetričnost  $C_{12}(\mathbf{h}) = C_{21}(\mathbf{h})$ , zato sledi enaka zveza kot pri zvezi med variogrami in korelacijskimi funkcijami (2.51)

$$\gamma_{12}(\mathbf{h}) = C_{12}(\mathbf{0}) - C_{12}(\mathbf{h}). \tag{2.53}$$

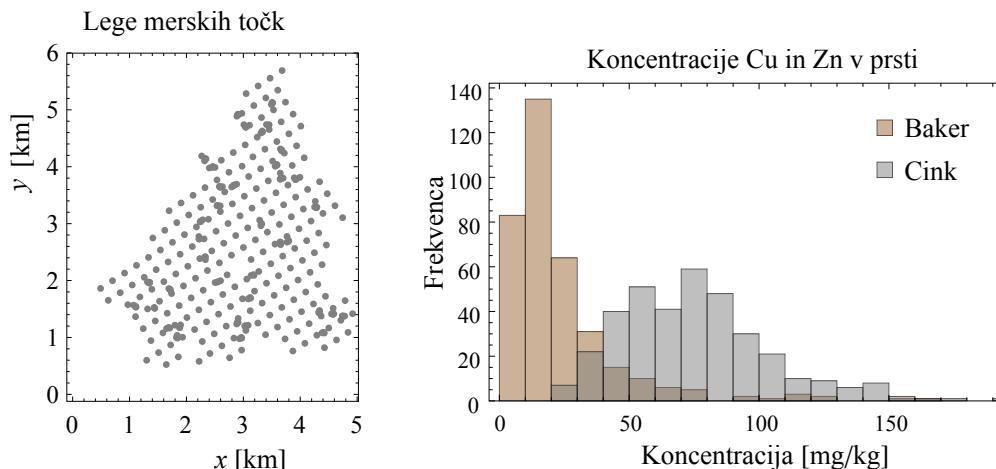


## 3 Geostatistika

V prejšnjem poglavju smo opisali slučajno polje oziroma slučajni proces in njihove osnovne značilnosti, porazdelitve, momente, mere prostorske povezanosti. V tem poglavju prehajamo na statistično analizo, obravnavali bomo vzorce slučajnih polj oziroma procesov in opisali ustrezne statistične metode.

### 3.1 Geostatistični podatki

Geostatistične podatke sestavljajo podatki o legi točke, kjer je bila opravljena meritev količine, ki jo obravnavamo, in vrednosti meritev te količine. Primer, ki ga bomo uporabili za prikaz različnih geostatističnih metod, so podatki iz severozahodnega švicarskega kantona Jura na meji s Francijo. Tam je Švicarski zvezni inštitut za tehnologijo iz Lozane (EPFL) na približno  $14 \text{ km}^2$  v 359 točkah ocenil tip kamnine, rabo tal in izmeril koncentracijo težkih kovin v tleh [mg/kg]. Geostatistični podatki so v tem primeru sestavljeni iz koordinat merskih točk, podatkov o tipu kamnine in rabi tal ter rezultatov meritev koncentracij težkih kovin v tleh.



SLIKA 3.1: Razpored merskih točk in histogram koncentracije dveh kovin

Na sliki 3.1 je prikazan razpored merskih točk. Vidimo, da je večina točk razporejenih v pravilni mreži, poleg tega pa je dodanih še nekaj merskih mest, ki so dodana v neurejeni obliki. Lahko predvidevamo, da so bile te dodatne merske točke postavljene v bližino potencialnih onesnaževalcev ali pa v bližino zajetja

pitne vode, bližino vrtcev ali podobno. Zanimiva je tudi primerjava v porazdelitvi koncentracije bakra in cinka, saj vidimo, da so koncentracije cinka bistveno višje, njihova porazdelitev pa je bolj simetrična od porazdelitve koncentracije bakra. Lastnosti slučajnih polj koncentracije bakra in cinka bomo prikazali v nadaljevanju.

Začetnih nekaj vrstic podatkov prikazujemo v preglednici 3.1. Razlikovali so med petimi tipi kamnine: prvi štirje (*Argovian, Kimmeridgian, Sequanian in Portlandian*) pripadajo jurskim apnencem, peta (*Quartarnary*) pa je mlajša kamnina iz obdobja kvartarja. Rabo tal so klasificirali v štiri razrede: gozd, pašnik, travnik in polja.

PREGLEDNICA 3.1: Prvih nekaj vrstic podatkov za območje Jure

x [km]	y [km]	Tip kamnine	Raba tal	Cd [mg/kg]	Cu [mg/kg]	Pb [mg/kg]	Co [mg/kg]	Cr [mg/kg]	Ni [mg/kg]	Zn [mg/kg]
2.386	3.077	3	3	1.740	25.72	77.36	9.32	38.32	21.32	92.56
2.544	1.972	2	2	1.335	24.76	77.88	10.00	40.20	29.72	73.56
2.807	3.347	3	2	1.610	8.88	30.80	10.60	47.00	21.40	64.80
4.308	1.933	2	3	2.150	22.70	56.40	11.92	43.52	29.72	90.00
4.383	1.081	5	3	1.565	34.32	66.40	16.32	38.52	26.20	88.40
3.244	4.519	5	3	1.145	31.28	72.40	3.50	40.40	22.04	75.20
3.925	3.785	5	3	0.894	27.44	60.00	15.08	30.52	21.76	72.40
2.116	3.498	1	3	0.525	66.12	141.00	4.20	25.40	9.72	72.08
1.842	0.989	1	3	0.240	22.32	52.40	4.52	27.96	11.32	56.40
1.709	1.843	3	3	0.625	18.72	41.60	12.08	33.32	16.88	75.60
...	...	...	...	...	...	...	...	...	...	...

Iz podatkov lahko določimo ocene srednjih vrednosti (povprečja) in variance (vzorčne variance) ter nekaj drugih statističnih podatkov, kar prikazujemo v preglednici 3.2. Iz te lahko preberemo, da najvišje povprečne koncentracije izkazuje svinec in cink, ki imata tudi največjo razpršenost. Verjetnostne porazdelitve koncentracije kobalta, kroma in niklja so najbližje normalni porazdelitvi, saj so te porazdelitve precej simetrične (koeficient simetričnosti je blizu nič) po obliki pa niso niti špičaste niti sploščene (kurtosis je blizu vrednosti tri), kar je značilnost normalne porazdelitve.



PREGLEDNICA 3.2: Nekaj osnovnih statističnih podatkov

Statistika	Slučajne spremenljivke - elementi						
	Cd	Cu	Pb	Co	Cr	Ni	Zn
Povprečje	1.30	23.6	54.6	9.48	35.0	20.0	75.9
Standardni odklon	0.873	22.3	33.1	3.63	10.7	8.14	30.8
Mediana	1.11	17.2	46.8	9.84	34.8	20.6	73.6
Simetričnost	1.49	3.01	3.34	-0.0705	0.295	0.104	1.49
Kurtosis	5.66	14.7	18.6	2.72	3.32	3.17	7.70
5. centil	0.32	5.72	26.3	3.55	18.5	6.08	35.1
95. centil	3.15	65.2	116.	14.6	52.8	32.6	136.
Minimum	0.135	3.55	18.7	1.55	3.32	1.98	25.0
Maksimum	5.13	166.	300.	22.7	70.0	53.2	260.

Pomembna je tudi informacija, v kolikšni meri so slučajne spremenljivke povezane. Mera za linearno povezanost je kovarianca oziroma korelacijski koeficient. V preglednici 3.3 so prikazani korelacijski koeficienti, ki prikazujejo linearno povezanost med spremenljivkami. Korelacijski koeficient, ki se po absolutni vrednosti približuje vrednosti 1, predstavlja močno linearno povezani slučajni spremenljivki. Vidimo lahko, da sta močno linearno povezani koncentraciji bakra in svinca, zelo šibko pa sta povezni koncentraciji svinca in kadmija.

PREGLEDNICA 3.3: Korelacijski koeficienti med obravnavanimi slučajnimi spremenljivkami

	Cd	Cu	Pb	Co	Cr	Ni	Zn
Cd	1.	0.139	0.221	0.259	0.566	0.475	0.601
Cu	0.139	1.	0.824	0.183	0.215	0.222	0.657
Pb	0.221	0.824	1.	0.148	0.256	0.271	0.665
Co	0.259	0.183	0.148	1.	0.472	0.703	0.441
Cr	0.566	0.215	0.256	0.472	1.	0.705	0.61
Ni	0.475	0.222	0.271	0.703	0.705	1.	0.591
Zn	0.601	0.657	0.665	0.441	0.61	0.591	1.

## 3.2 Priprava podatkov

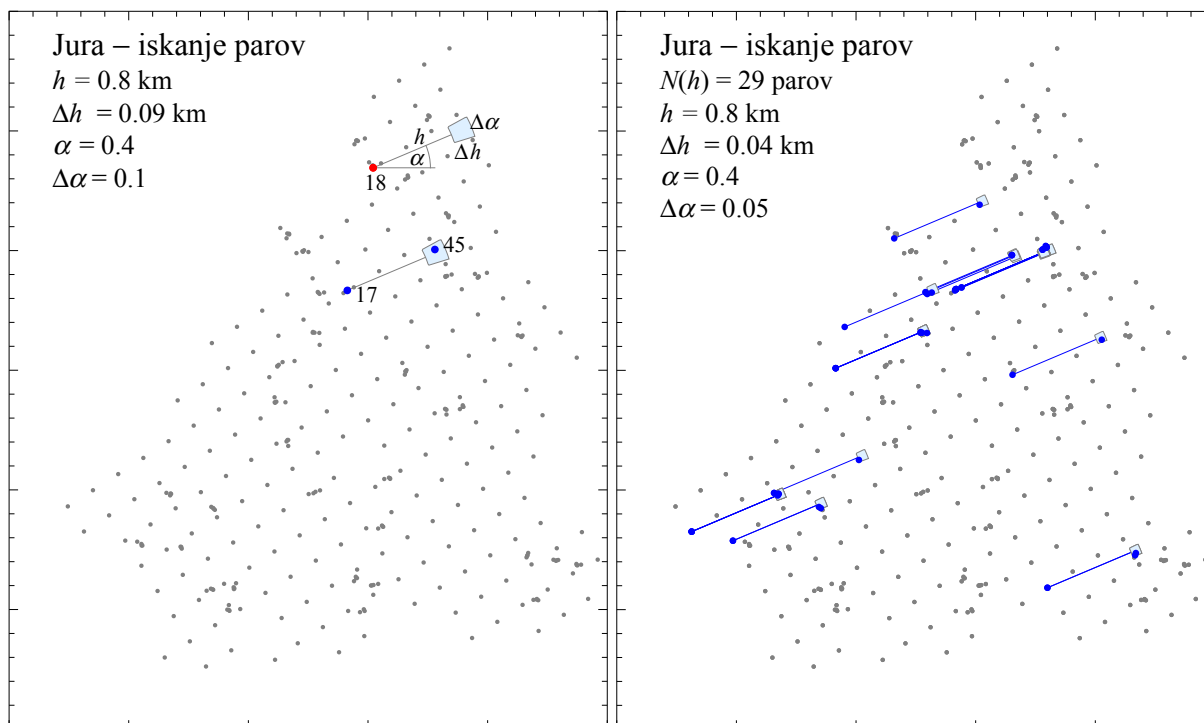
Ob predpostavki, da je polje stacionarno, lahko iz podatkov določimo ocene srednjih vrednosti in varianc za vse merjene količine. Pri določitvi ocene kovariančne funkcije in variograma pa bomo morali poiskati pare točk z enako ali približno enako medsebojno oddaljenostjo. Zato izberemo vektor  $\mathbf{h}$  z dolžino  $h = |\mathbf{h}|$  in smernim kotom  $\alpha$ . Za ta vektor poiščemo vse tiste pare točk, za katere velja, da je medsebojna

razdalja znotraj intervala  $h \pm \Delta h$  in da je smerni kot med prvo in drugo točko znotraj intervala  $\alpha \pm \Delta\alpha$ . Število teh parov označimo z  $N(\mathbf{h})$ . Na sliki 3.2 prikazujemo množico 359 točk in lahko ugotovimo, da za izbrani vektor oddaljenosti med točkama  $\mathbf{h}$  točko 17 lahko povežemo s točko 45 (obe sta obarvani v modro), medtem ko točka 18 za izbrani vektor  $\mathbf{h}$  nima para (obarvana je rdeče). Na desni strani slike 3.2 je prikazanih vseh  $N(\mathbf{h}) = 29$  parov točk s približno enako medsebojno oddaljenostjo in približno enakim smernim kotom med točkama. Vidimo, da je lahko ena točka povezana z več točkami.

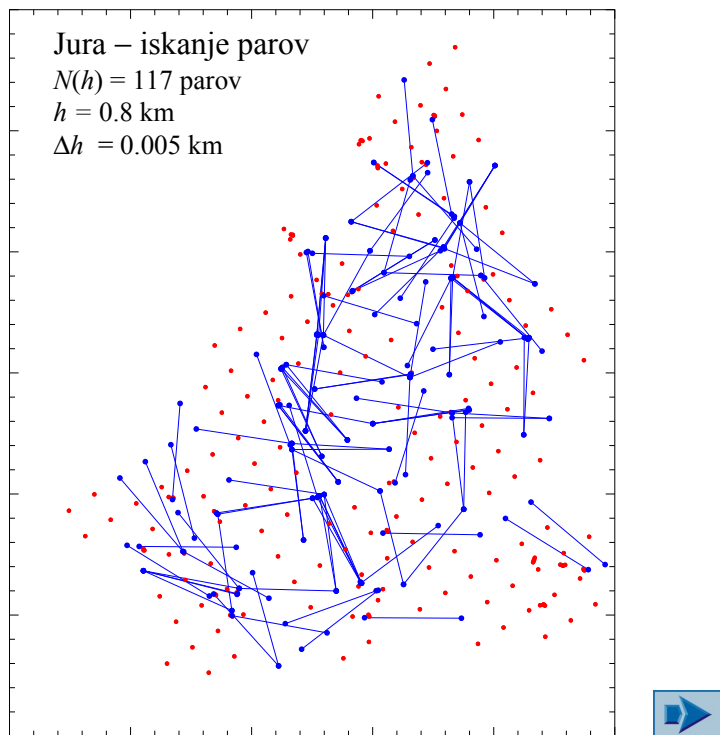
V prikazanem primeru smo iskali le tiste pare točk z medsebojno oddaljenostjo  $h \pm \Delta h$  in s smernim kotom daljice od prve do druge točke v intervalu  $\alpha \pm \Delta\alpha$ . Pogosto predpostavimo, da je polje izotropno, torej da se lastnosti v različnih smereh ne razlikujejo. Tedaj pare točk s podobno oddaljenostjo iščemo ne glede na smerni kot daljice med dvema točkama. Tak izbor parov za istih 359 točk je prikazan na sliki 3.3. Čeprav smo velikost intervala  $\Delta h$  zmanjšali za 8-krat, je število parov sedaj bistveno večje,  $N(\mathbf{h}) = 117$ . Tudi na sliki 3.3 so z modro označene točke, ki so v parih, z rdečo pa tiste, ki nimajo para.

Najpogosteje bomo obravnavali primere, ki niso odvisni od smeri vektorja  $\mathbf{h}$ , zato bomo v nadaljevanju v teh primerih namesto vektorja  $\mathbf{h}$  pri zapisu  $N(\mathbf{h})$  upoštevali le dolžino vektorja  $\mathbf{h}$  in zapisali

$$N(|\mathbf{h}|) = N(h.)$$



SLIKA 3.2: Iskanje parov točk, ki so oddaljene za približno  $\mathbf{h}$



SLIKA 3.3: Pari točk s približno oddaljenostjo  $h$  ne glede na smer daljice

V vsaki točki  $s_k$  poznamo tudi vrednost  $Z_k$  slučajnega polja  $Z(s)$ . Zato lahko ob iskanju parov za nek izbran  $\mathbf{h}$  za vsak par točk na razdalji  $h$  zapišemo par vrednosti  $(Z_{1,i}, Z_{2,i})$ , kjer je  $Z_{1,i}$  vrednost slučajnega polja v prvi točki  $i$ -tega para,  $Z_{2,i}$  pa vrednost polja v drugi točki tega para. Na primer, če smo na sliki 3.2 ugotovili, da točki 17 in 45 predstavljata par dveh točk na ustrezni oddaljenosti glede na izbrani vektor  $\mathbf{h}$ , potem je vrednost  $Z_{1,1} = Z_{17}$  (vrednost  $Z$  v prvi točki prvega para je enaka vrednosti  $Z_{17}$ ), vrednost v drugi točki pa je  $Z_{2,1} = Z_{45}$  (vrednost  $Z$  v drugi točki prvega para je enaka vrednosti  $Z_{45}$ ). S tem smo podatke na novo uredili tako, da smo iz geostatističnih podatkov  $\{x_k, y_k, Z_k\}$ ,  $k = 1, \dots, n$  izluščili pare vrednosti  $\{Z_{1,i}, Z_{2,i}\}$ ,  $i = 1, \dots, N(\mathbf{h})$  za določen vektor  $\mathbf{h}$ , kot je prikazano v preglednici 3.4. V primeru s slike 3.2 je bilo vseh točk  $n = 359$ , ustreznih parov točk za izbrani vektor z dolžino  $h = 0.8$  km in smerjo  $\alpha = 0.4$  pa je  $N(\mathbf{h}) = 29$ .

Zaradi preglednosti bomo v nadaljevanju vrednost slučajnega polja za določen par  $i$  pisali brez vejic med indeksi

$$Z_{1,i} \rightarrow Z_{1i} \quad \text{in} \quad Z_{2,i} \rightarrow Z_{2i}.$$

PREGLEDNICA 3.4: Preurejevanje podatkov za geostatistično analizo

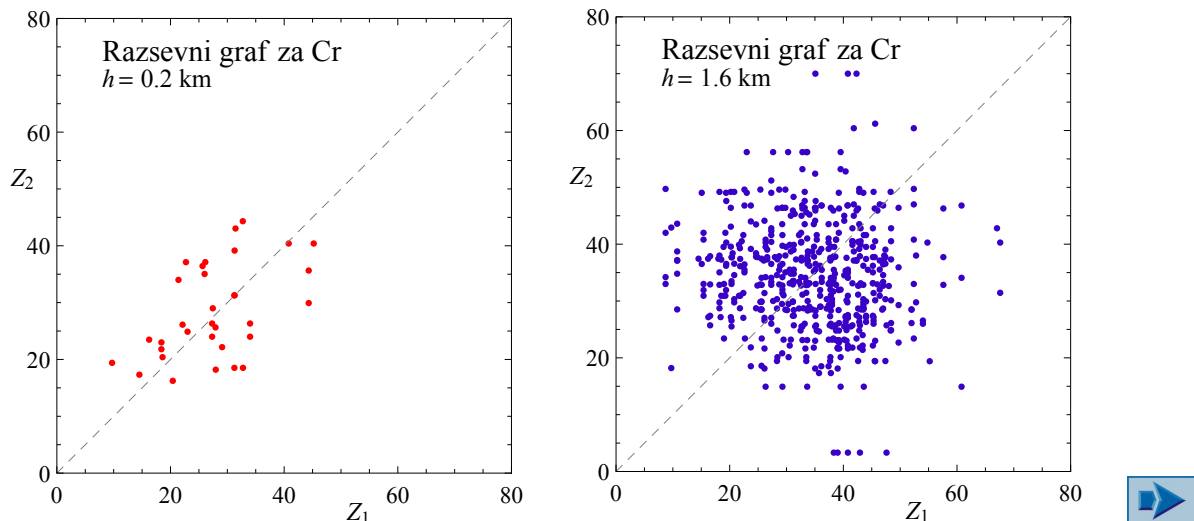
Geostatistični podatki			→	Pari vrednosti	
$x_1$	$y_1$	$Z_1$		$Z_{11}$	$Z_{21}$
$x_2$	$y_2$	$Z_2$	$Z_{12}$	$Z_{22}$	
...	...	...	...	...	
$x_k$	$y_k$	$Z_k$	$Z_{1i}$	$Z_{2i}$	
...	...	...	...	...	
$x_n$	$y_n$	$Z_n$	$Z_{1N(h)}$	$Z_{2N(h)}$	

S tako urejenimi podatki lahko grafično ali numerično prikažemo prostorsko povezanost. Najpreprostejši grafični prikaz je **razsevni graf**, ki ga lahko narišemo za različne razdalje  $h$ . Za vsako razdaljo  $h$  lahko izračunamo tudi oceno kovariančne in korelacijske funkcije ( $\hat{C}(h)$  in  $\hat{\rho}(h)$ ) ter oceno variograma  $\hat{\gamma}(h)$ .

### 3.3 Razsevni graf

Z razsevnim grafom narišemo pare točk  $\{Z_{1i}, Z_{2i}\}$  za izbrane razdalje  $h$  med točkami. Na sliki 3.4 prikazujemo razsevni graf za krom za dve različni razdalji  $h$  med točkami.

Če so točke razsevnega grafa bližje simetrijski osi prvega kvadranta, pomeni, da je slučajno polje za izbrani  $h$  prostorsko bolj povezano (glej sliko 3.4 za  $h = 0.2$  km). Za prostorsko neodvisno slučajno polje je značilno, da so točke razsevnega grafa razporejene po celotnem območju in niso zbrane v bližini simetrijske osi (glej sliko 3.4 za  $h = 1.6$  km).

SLIKA 3.4: Razsevni graf za Cr pri  $h = 0.2$  km in  $h = 1.6$  km

### 3.4 Kovariančna funkcija

Za določitev ocene kovariančne in korelacijske funkcije moramo najprej oceniti srednje vrednosti in variance slučajnega polja. Te določimo z naslednjimi enačbami:

$$\begin{aligned}\hat{m}_1 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z_{1i}, \\ \hat{m}_2 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z_{2i}, \\ \hat{\sigma}_1^2 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Z_{1i} - \hat{m}_1)^2 = \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z_{1i}^2 - \hat{m}_1^2, \\ \hat{\sigma}_2^2 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Z_{2i} - \hat{m}_2)^2 = \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z_{2i}^2 - \hat{m}_2^2.\end{aligned}\tag{3.1}$$

Predpostavka o stacionarnih poljih sicer zahteva, da sta enaki srednji vrednosti in varianci  $m_1 = m_2$  in  $\sigma_1^2 = \sigma_2^2$ , a zaradi naključne narave slučajnih procesov ne moremo pričakovati, da bosta enaki tudi oceni srednjih vrednosti in varianc v prvih oziroma drugih točkah parov. Velja torej:

$$\begin{aligned}\hat{m}_1 &\neq \hat{m}_2, \\ \hat{\sigma}_1^2 &\neq \hat{\sigma}_2^2.\end{aligned}$$

V primeru, da sta oceni srednje vrednosti in variance za prve točke parov zelo različni, predpostavka o stacionarnosti ne velja in moramo slučajno polje obravnavati kot nestacionarno.

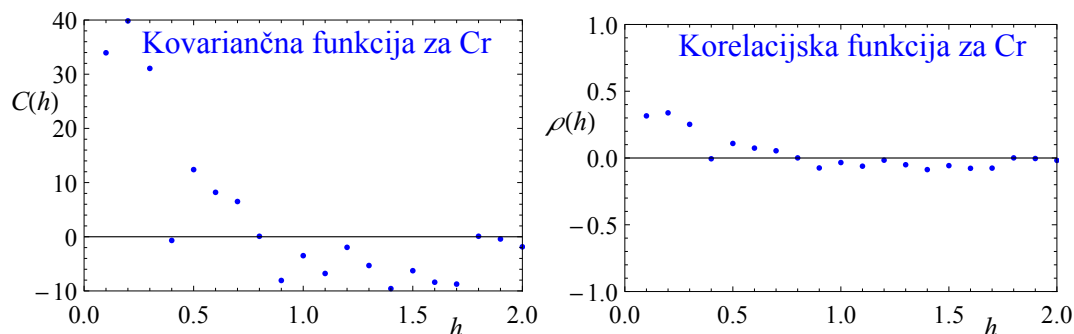
Oceno kovariančne funkcije oziroma vzorčno kovariančno funkcijo izračunamo po naslednji enačbi:

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Z_{1i} - \hat{m}_1)(Z_{2i} - \hat{m}_2) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} Z_{1i}Z_{2i} - \hat{m}_1\hat{m}_2,\tag{3.2}$$

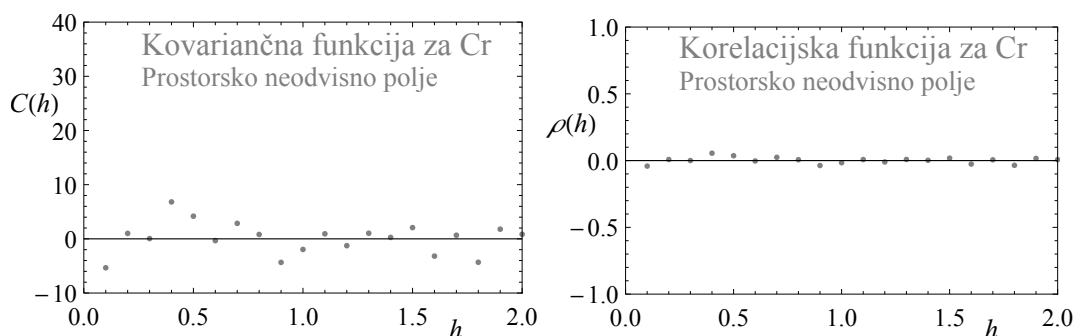
oceno korelacijske funkcije pa po enačbi:

$$\hat{\rho}(h) = \frac{\hat{C}(h)}{\hat{\sigma}_1\hat{\sigma}_2}, \quad \text{kjer je} \quad -1 \leq \hat{\rho}(h) \leq 1.\tag{3.3}$$

Ocene kovariančne in korelacijske funkcije za  $h = 0.1, \dots, 2.0$  km prikazujemo na sliki 3.5. Vidimo lahko, da je povezava nekoliko bolj izrazita za razdalje do 0.4 km, nato pa vrednost funkcije pade in prostorska povezanost slučajnega polja ni več značilna. Če primerjamo sliko 3.5 s sliko 3.6, kjer sta prikazani kovariančna in korelacijska funkcija za prostorsko neodvisno polje z enakima srednjo vrednostjo in varianco kot prostorsko povezano polje, vidimo, da imata v primeru neodvisnih polj tako kovariančna kot korelacijska funkcija za vse vrednosti  $h$  nizke vrednosti, ki so od nič različne le zaradi naključne narave slučajnega polja.



SLIKA 3.5: Oceni kovariančne in korelacijske funkcije za Cr



SLIKA 3.6: Oceni kovariančne in korelacijske funkcije za Cr (prostorsko neodvisno polje)

### 3.5 Variogram

Oceno variograma oziroma vzorčni variogram izračunamo po naslednji enačbi:

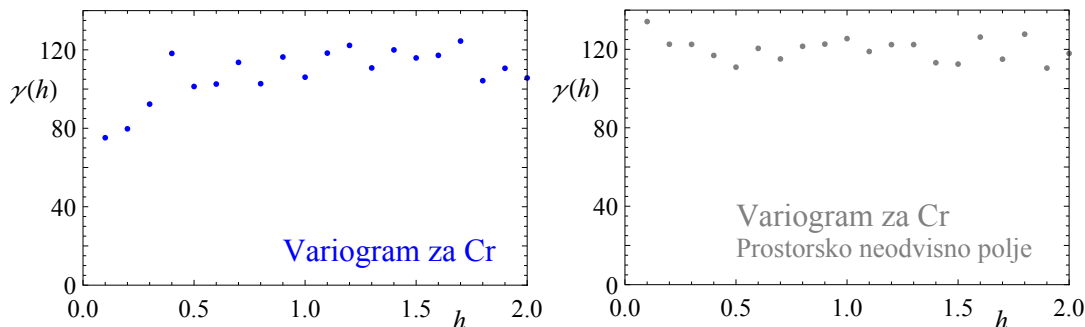
$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Z_{1i} - Z_{2i} - (\hat{m}_1 - \hat{m}_2))^2,$$

(zaradi stacionarnosti polja vzamemo, da je  $(\hat{m}_1 - \hat{m}_2) = 0$ )

(3.4)

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Z_{1i} - Z_{2i})^2.$$

Variogram predstavlja polovico variance razlike med vrednostjo slučajnega polja v dveh točkah na oddaljenosti  $h$ . Pričakujemo lahko, da bodo razlike med vrednostmi polja pri večjih razdaljah večje, torej da bo variogram naraščajoča funkcija  $h$ . Če pa je polje prostorsko neodvisno, lahko pričakujemo, da bo variogram konstanten. Na sliki 3.7 vidimo, da variogram pri dejanskih podatkih o koncentraciji kroma na področju Jure narašča, medtem ko je pri podatkih, ki so bili generirani kot prostorsko neodvisno polje, variogram konstanten v okviru nezanesljivosti zaradi naključne narave slučajnih polj.

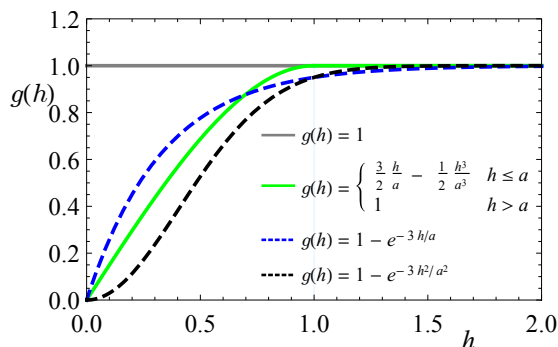


SLIKA 3.7: Oceni variograma za Cr (prostorsko odvisno in neodvisno polje)

### 3.6 Določitev empirične funkcije variograma

V naslednjem poglavju bomo obravnavali krigiranje, pri katerem bomo potrebovali vrednosti variograma oziroma kovariančne funkcije v odvisnosti od razdalje med točkami  $h$ . Zato bomo iz vzorčnih vrednosti variograma, ki so podane po točkah, določili približno funkcijsko zvezo. Naštejmo nekaj značilnih funkcij, ki so glede na običajno obliko variograma primerne za njegovo aproksimacijo:

$$\begin{aligned}
 g_1(h) &= 1 && \text{konstanta} \\
 g_2(h) &= \begin{cases} \frac{3h}{2a} - \frac{h^3}{2a^3} & h \leq a \\ 1 & h > a \end{cases} && \text{sferična} \\
 g_3(h) &= 1 - e^{-3h/a} && \text{eksponentna} \\
 g_4(h) &= 1 - e^{-3h^2/a^2} && \text{Gaussova}
 \end{aligned} \tag{3.5}$$



SLIKA 3.8: Nekaj funkcij, primernih za aproksimacijo variograma

Empirično funkcijo variograma izračunamo s seštevanjem funkcij iz (3.5) po naslednji enačbi:

$$g(h) = b_1 g_1(h) + b_2 g_2(h, a_2) + b_3 g_3(h, a_3) + \dots, \tag{3.6}$$

pri čemer neznane koeficiente  $b_1, b_2, a_2, \dots$  določimo po metodi najmanjših kvadratov. Če se na primer odločimo za kombinacijo konstante in eksponentne funkcije, je empirična funkcija variograma:

$$g(h) = b_1 + b_2(1 - e^{-3h/a_2}), \quad (3.7)$$

če pa izberemo kombinacijo konstante in Gaussove funkcije, je empirična funkcija:

$$g(h) = b_1 + b_2(1 - e^{-3h^2/a_2^2}), \quad (3.8)$$

pri čemer moramo določiti še parametre  $b_1, b_2$  in  $a_2$  tako, da se empirična funkcija čim bolj prilega točkovnim vrednostim vzorčnega variograma. Pri metodi najmanjših kvadratov moramo poiskati take parametre empiričnih funkcij variograma, da je vsota kvadratov razlik med vrednostmi vzorčnega variograma in izbrano funkcijo čim manjša:

$$SS = \sum_{i=1}^{n_v} (\gamma_i - g(h_i))^2 = \sum_{i=1}^{n_v} \left( \gamma_i - b_1 - b_2(1 - e^{-3h_i/a_2}) \right)^2 \quad \text{za eksponentno funkcijo,} \quad (3.9)$$

kjer  $n_v$  predstavlja število točk vzorčnega variograma. Vrednosti  $h_i$  izberemo glede na lege točk, saj tako za premajhne kot za prevelike vrednosti  $h_i$  ne moremo najti ustreznega para, torej je  $N(h) = 0$  in kovariančne funkcije oziroma variograma ne moremo izračunati.

Za določitev empirične kovariančne funkcije uporabimo zvezo med kovariančno funkcijo in variogramom (2.31) in upoštevamo, da vsota vseh  $b_i$  predstavlja največjo vrednost variograma, ki je enaka varianci slučajnega polja  $C(0)$ . Empirično kovariančno funkcijo zapišemo z enačbo:

$$C(h) = (b_1 + b_2 + \dots) - (b_1 g_1(h) + b_2 g_2(h, a_2) + b_3 g_3(h, a_3) + \dots). \quad (3.10)$$

Učinek zrna je v empirični funkciji variograma opisan s parametrom  $b_1$ , ki predstavlja konstantni del variograma.

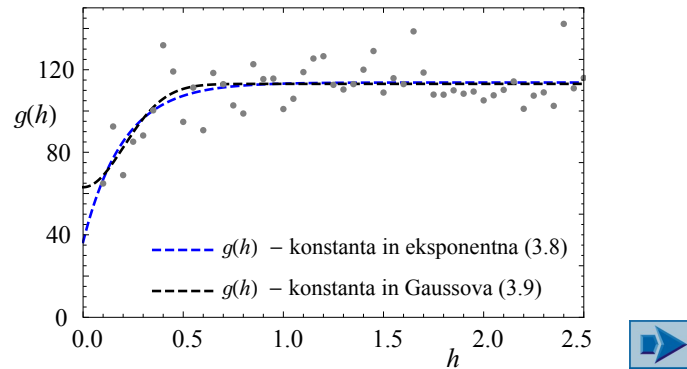
Sedaj določimo empirični variogram za krom za razdalje med točkami od 0 do 2.5 km ob uporabi dveh različnih nastavkov po enačbah (3.7) in (3.8). Rezultat aproksimacije so vrednosti parametrov za obe enačbi, ki jih podajamo v preglednici 3.5, aproksimacijo variograma z vzorčnim variogramom pa prikazujemo na sliki 3.9.

PREGLEDNICA 3.5: Parametri aproksimacije variograma za krom po enačbah (3.7) in (3.8)

Funkcija	$n_v$	$b_1$	$b_2$	$a_2$	Vsota kvadratov $SS$
(3.7)	49	36.5	77.3	0.604	5529
(3.8)	49	63.1	50.1	0.498	5496

Iz preglednice 3.5 in s slike 3.9 lahko razberemo, da sta prag in doseg obeh funkcij variograma podobni, saj je ta vrednost pri eksponentni funkciji 113.8, pri Gaussovi pa 113.1. Razlika v učinku zrna je bolj izrazita: v primeru eksponentne funkcije je enak 36.5, v primeru Gaussove pa 63.1. Vsota kvadratov odstopanj  $SS$  je v obeh primerih zelo podobna (5529 in 5496), Gaussova aproksimacija kaže nekoliko boljše prileganje, a je razlika zanemarljiva.





SLIKA 3.9: Aproximacija variograma

### 3.7 Geostatistična analiza z indikatorji

Pri analizi geostatističnih podatkov nas pogosto zanimajo le vrednosti, ki so manjše ali večje od izbrane kritične vrednosti. Tako nas na primer lahko zanima, ali so vrednosti, ki presegajo zakonsko dovoljene koncentracije onesnaženja, prostorsko odvisne ali ne. Podobno nas lahko zanima, ali je vsebnost rude v kamnini, manjša od določene kritične vrednosti, prostorsko odvisna in kako. Če želimo opraviti tako analizo, moramo iz osnovnih geostatističnih podatkov določiti takoimenovane indikatorje po naslednji enačbi:

$$i_k = \begin{cases} 1 & \text{za } Z_k \leq z_{krit}, \\ 0 & \text{za } Z_k > z_{krit}, \end{cases} \quad (3.11)$$

ali pa:

$$i_k = \begin{cases} 1 & \text{za } Z_k \geq z_{krit}, \\ 0 & \text{za } Z_k < z_{krit}. \end{cases} \quad (3.12)$$

Geostatistične podatke smo sedaj razširili za indikatorje  $i_k(z_{krit})$  glede na izbrano kritično vrednost  $z_{krit}$ . Ob določitvi parov za izbrano razdaljo med točkami  $h$  dobimo nove pare vrednosti  $\{i_{1i}, i_{2i}\}$ , kot prikazujemo v preglednici 3.6.

PREGLEDNICA 3.6: Preurejevanje podatkov za geostatistično analizo z indikatorji

Geostatistični podatki				→	Pari indikatorjev	
$x_1$	$y_1$	$Z_1$	$i_1$		$i_{11}$	$i_{21}$
$x_2$	$y_2$	$Z_2$	$i_2$	$i_{12}$	$i_{22}$	
...	...	...	...	...	...	
$x_k$	$y_k$	$Z_k$	$i_k$	$i_{1i}$	$i_{2i}$	
...	...	...	...	...	...	
$x_n$	$y_n$	$Z_n$	$i_n$	$i_{1N(h)}$	$i_{2N(h)}$	

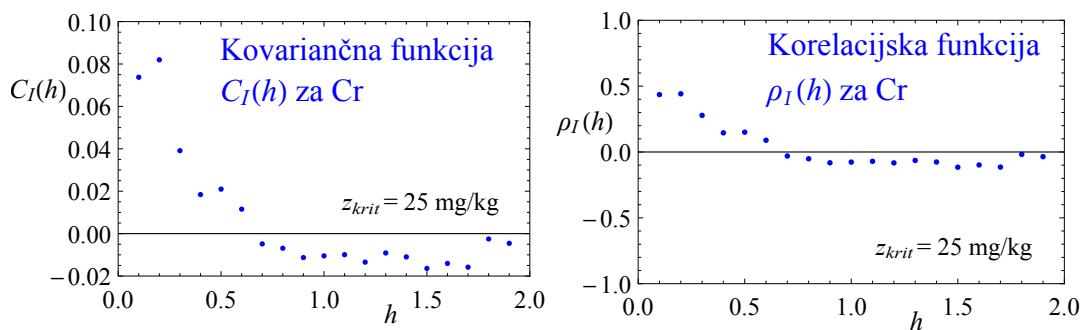
Glede na to, da indikatorji lahko dosežejo le dve vrednosti (ena ali nič), razsevnega grafa ne rišemo, saj

lahko vsebuje največ 4 točke ( $\{0, 0\}$ ,  $\{1, 0\}$ ,  $\{0, 1\}$ ,  $\{1, 1\}$ ). Zato pa lahko izračunamo vse druge mere: srednjo vrednost, varianco, kovariančno in korelacijsko funkcijo ter variogram:

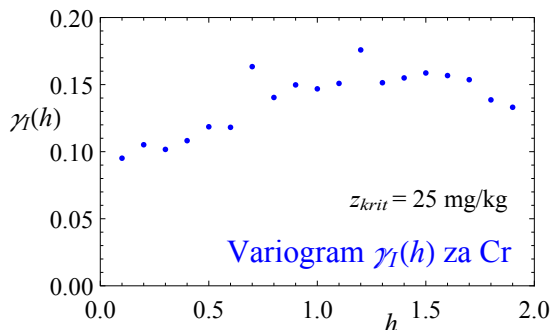
$$\begin{aligned}\hat{m}_{I1} &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} i_{1i}, \\ \hat{m}_{I2} &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} i_{2i},\end{aligned}\tag{3.13}$$

$$\begin{aligned}\hat{\sigma}_{I1}^2 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} i_{1i}^2 - \hat{m}_{I1}^2, \\ \hat{\sigma}_{I2}^2 &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} i_{2i}^2 - \hat{m}_{I2}^2, \\ \hat{C}_I(h) &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} i_{1i}i_{2i} - \hat{m}_{I1}\hat{m}_{I2}, \\ \hat{\rho}_I(h) &= \frac{\hat{C}_I(h)}{\hat{\sigma}_{I1}\hat{\sigma}_{I2}}, \\ 2\hat{\gamma}_I(h) &= \frac{1}{N(h)} \sum_{i=1}^{N(h)} (i_{1i} - i_{2i})^2.\end{aligned}\tag{3.14}$$

Na slikah 3.10 in 3.11 prikazujemo ocene za kovariančno in korelacijsko funkcijo ter variograma za kritično vrednost  $z_{krit} = 25$  mg/kg koncentracije kroma, pri čemer smo indikatorje določili po enačbi (3.12). Vidimo lahko, da je za izbrano kritično vrednost 25 mg/kg prostorska povezanost podobna kot, če obravnavamo vrednosti koncentracije.

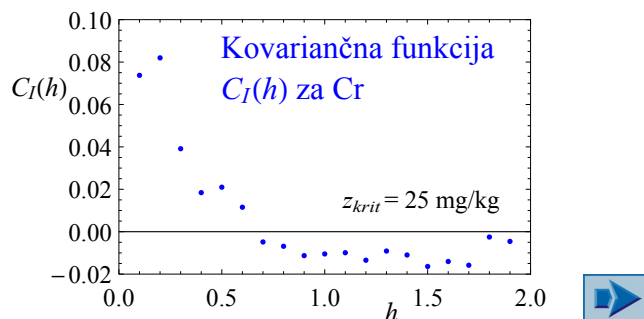


SLIKA 3.10: Oceni kovariančne in korelacijske funkcije z indikatorji za Cr



SLIKA 3.11: Ocena variograma z indikatorji za Cr

Na sliki 3.12 prikazujemo kovariančno funkcijo z indikatorji za različne kritične vrednosti. Opazimo lahko, da je prostorska povezanost izrazitejša za kritične vrednosti med 25 in 35mg/kg, za manjše ali večje kritične vrednosti pa je povezanost manjša.

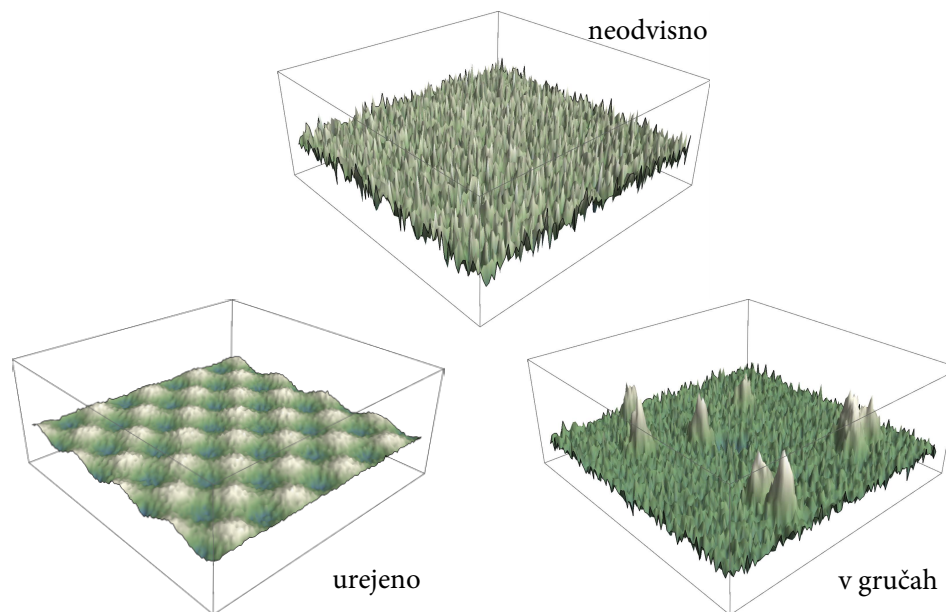
SLIKA 3.12: Kovariančna funkcija z indikatorji za Cr za različne vrednosti  $z_{krit}$ 

### 3.8 Preizkušanje domnev o prostorski povezanosti

Geostatistične analize, ki smo jih opisali do sedaj, so pokazale prostorsko odvisnost preko oblike raz-sevnega grafa, preko oblike kovariančne, korelacijske funkcije ali variograma. Mnenje o tem, ali je slučajno polje prostorsko povezano ali ne, pa je prepuščeno subjektivni presoji grafičnih prikazov. V nadaljevanju bomo pokazali, da lahko pri presoji o prostorski povezanosti slučajnega polja uporabimo običajno statistično orodje: **preizkušanje domnev** (tudi **testiranje hipotez**, angl. hypothesis testing). Postopek preizkušanja domnev o prostorski povezanosti slučajnega polja je povsem enak kot pri vseh drugih preizkušanjih statističnih domnev:

- 1) Postavitev ničelne  $H_0$  in alternativne domneve  $H_1$ . Alternativna domneva ne sme biti združljiva z ničelno.
- 2) Izbira največjega tveganja  $\alpha$ , ki ga dovolimo pri zavrnitvi ničelne domneve (običajne vrednosti so  $\alpha = 0.01$ ,  $\alpha = 0.05$  ali  $\alpha = 0.10$ ).
- 3) Zapis statistike, ki je skladna z ničelno domnevo.
- 4) Določitev mej kritičnega območja oziroma območja zavrnitve ničelne domneve.
- 5) Glede na izračunano statistiko v primerjavi z mejami kritičnega območja odločitev o zavrnitvi ničelne domneve in zapis končne izjave.

Ničelna domneva je pri vseh preizkusih o prostorski povezanosti enaka: „*Slučajno polje je prostorsko neodvisno.*“ Poleg prostorsko neodvisnega slučajnega polja so lahko vrednosti v slučajnem polju urejene po neki zakonitosti. Tedaj rečemo, da so vrednosti slučajnega polja urejene. Slučajno polje je lahko tudi tako, da so vrednosti na manjšem območju bistveno drugačne od vrednosti polja na drugih delih obravnavanega območja. V tem primeru pravimo, da so vrednosti polja zbrane v gručah. Na sliki 3.13 prikazujemo primere treh polj, prostorsko neodvisnega, urejenega in polja v gručah.



SLIKA 3.13: Prostorsko neodvisno, urejeno slučajno polje in slučajno polje z vrednostmi v gručah

Alternativna domneva je „*Slučajno polje je prostorsko odvisno.*“, če želimo izvesti dvostranski preizkus, ter „*Slučajno polje je urejeno.*“ ali „*Vrednosti slučajnega polja so v gručah.*“, če želimo izvesti enostranski preizkus. V geostatistiki sta se uveljavila dva preizkusa o prostorski povezanosti slučajnega polja **Moranov indeks**  $I$ , ki temelji na korelacijski funkciji, ter **Gearyjevo razmerje**  $C$ , ki temelji na variogramu.

### 3.8.1 Moranov indeks

Moranov indeks  $I$  je mera za oceno prostorske povezanosti vrednosti slučajnega polja. Pomembna razlika glede na razsevni graf, kovariančno funkcijo in variogram je, da Moranov indeks ni odvisen od razdalje med dvema točkama, temveč so v njem zajete vse razdalje med točkami. Definiran je z enačbo:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \left( \sum_{i=1}^n (Z_i - \bar{Z})^2 \right)}, \quad (3.15)$$

kjer so z  $w_{ij}$  označene uteži,  $Z_i$  in  $Z_j$  so vrednosti slučajnega polja v vseh  $n$  točkah,  $\bar{Z}$  pa je povprečje vrednosti slučajnega polja, ki ga izračunamo po enačbi:

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i. \quad (3.16)$$

Uteži  $w_{ij}$  so lahko binarne:

$$w_{ij} = \begin{cases} 1 & \text{če sta točki sosedni,} \\ 0 & \text{če točki nista sosedni.} \end{cases} \quad (3.17)$$

Določitev sosede je lahko vezana na razdaljo med točkami. Pogoj je torej, da je razdalja  $d_{ij}$  med sosednjima točkama manjša od neke določene razdalje  $d_k$ . Sosednjo točko lahko določimo tudi glede na skupno mejo območja, ki pripada določeni točki. Poleg tega naj velja, da točka sama sebi ni soseda, kar pomeni, da je  $w_{ii} = 0$ .

Uteži so lahko tudi zvezne funkcije razdalje med točkami:

$$w_{ij} = f(d_{ij}). \quad (3.18)$$

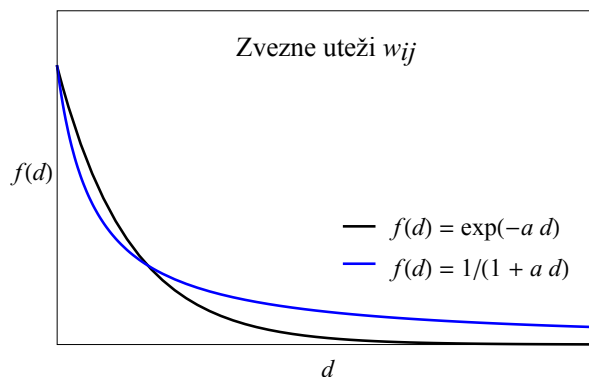
Običajno uporabimo padajoče eksponentne ali racionalne funkcije:

$$\begin{aligned} f(d) &= e^{-ad}, \\ f(d) &= \frac{1}{1 + ad}. \end{aligned} \quad (3.19)$$

Na sliki 3.14 prikazujemo dve pogosto uporabljene funkciji, s katerima določamo vrednosti uteži. Razdaljo  $d$  običajno računamo kot Evklidovo<sup>1</sup> razdaljo med dvema točkama, lahko pa namesto tega uporabimo tudi druge mere razdalje med točkami, kot so: razdalja Manhattan, razdalja po poteh ali cestah, čas, potreben za premik iz ene v drugo točko, stroški, potrebni za premik iz ene v drugo točko, razdalja Mahalanobis<sup>2</sup> in drugo.

<sup>1</sup> Evklid, grški matematik (približno 325–265 pred našim štetjem).

<sup>2</sup> Prasanta Chandra Mahalanobis, indijski (bengalski) statistik (1893–1972).



SLIKA 3.14: Dve funkciji, ki ju lahko uporabimo za določitev vrednosti uteži  $w_{ij}$

Uteži lahko zapišemo v matriko uteži:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & 0 & w_{23} & \cdots & w_{2n} \\ w_{31} & w_{32} & 0 & \cdots & w_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & 0 \end{bmatrix}. \quad (3.20)$$

Moranov indeks  $I$  izračunamo iz geostatističnih podatkov, torej iz slučajnega vzorca, zato je slučajna spremenljivka. Zaloga vrednosti je  $I \in [-1, 1]$ , pri čemer vrednosti okoli 0 predstavljajo prostorsko neodvisno polje, pozitiven  $I$  nakazuje na to, da so vrednosti polja zbrane v gručah, negativne vrednosti  $I$  pa nakazujejo na prostorsko pravilno urejeno polje. Srednja vrednost Moranovega indeksa je

$$E[I] = -\frac{1}{n-1} \approx 0, \quad (3.21)$$

enačba za varianco pa je bistveno bolj zapletena:

$$\text{var}[I] = \frac{nS_4 - S_3S_5}{(n-1)(n-2)(n-3)S_0^2} - \left(\frac{-1}{n-1}\right)^2, \quad (3.22)$$

kjer  $S_i$  pomenijo naslednje:

$$\begin{aligned} S_0 &= \sum_{i=1}^n \sum_{j=1}^n w_{ij}, \\ S_1 &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} - w_{ji})^2, \\ S_2 &= \sum_{i=1}^n \left( \sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji} \right)^2, \end{aligned} \quad (3.23)$$

$$S_3 = \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^4}{\left(\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2\right)^2}, \quad (3.24)$$

$$S_4 = (n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2,$$

$$S_5 = (n^2 - n)S_1 - 2nS_2 + 6S_0^2.$$

V nadaljevanju še predpostavimo, da so geostatistični podatki taki, da je  $I$  slučajna spremenljivka

$$Z = \frac{I - E[I]}{\sqrt{\text{var}[I]}} \sim N(0, 1), \quad (3.25)$$

porazdeljena standardizirano normalno. To je osnova za preizkus domneve o prostorski neodvisnosti na osnovi Moranovega indeksa.

$H_0$ : Slučajno polje je prostorsko neodvisno ( $I = -1/(n-1)$ ).

$H_1$ : Slučajno polje je prostorsko odvisno ( $I \neq -1/(n-1)$ ).

---

$\alpha = 5\%$ .

Zapišemo statistiko  $Z$ :

$$Z = \frac{I - E[I]}{\sqrt{\text{var}[I]}} \quad (3.26)$$

in postavimo meje kritičnega območja za dvostranski preizkus domneve:

$$Z > k_{1-\alpha/2} \vee Z < -k_{1-\alpha/2} \longrightarrow H_0 \text{ zavrnemo.} \quad (3.27)$$

Če je torej  $Z > k_{1-\alpha/2}$  ali  $Z < -k_{1-\alpha/2}$ , ničelno domnevo **zavrnemo**. Za  $\alpha = 0.05$  je kritična vrednost  $k_{0,975} = 1.96$ . V tem primeru izjavimo: „*Slučajno polje je statistično značilno prostorsko odvisno!*“ ali pa „*S tveganjem, manjšim od  $\alpha = 5\%$ , lahko trdimo, da je slučajno polje prostorsko odvisno!*“

Če statistika  $Z$  ne pade v kritično območje ( $-k_{1-\alpha/2} \leq Z \leq k_{1-\alpha/2}$ ), potem ničelne domneve **ne zavrnemo**. Tedaj izjavimo: „*Slučajno polje ni statistično značilno prostorsko odvisno!*“ ali pa „*S tveganjem, manjšim ali enakim od  $\alpha = 5\%$ , ne moremo trditi, da je slučajno polje prostorsko odvisno!*“

### 3.8.2 Gearyjevo razmerje

Gearyjevo<sup>3</sup> razmerje  $C$  je še ena mera za oceno prostorske povezanosti vrednosti slučajnega polja. Definirano je z naslednjo enačbo:

$$C = \frac{(n-1)}{2} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Z_i - Z_j)^2}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \left( \sum_{i=1}^n (Z_i - \bar{Z})^2 \right)}, \quad (3.28)$$

iz katere je razvidno, da Gearyjevo razmerje izhajajo iz definicije variograma. Tudi Gearyjevo razmerje je podobno kot Moranov indeks slučajna spremenljivka. Zaloga vrednosti je interval od nič do približno dve. Vrednosti  $C$  v bližini ena nakazujejo na prostorsko neodvisno slučajno polje, vrednosti  $C$ , manjše od ena, predstavljajo slučajno polje, katerega vrednosti so zbrane v grućah, medtem ko vrednosti  $C$ , večje od ena, predstavljajo prostorsko odvisna urejena polja. Srednja vrednost in varianca slučajne spremenljivke  $C$  pa sta:

$$E[C] = 1 \quad (3.29)$$

in:

$$\text{var}[C] = \frac{S_6 - S_7 + S_8}{n(n-2)(n-3)S_0^2}, \quad (3.30)$$

kjer so

$$\begin{aligned} S_6 &= S_1(n-1)(n^2 - 3n + 3 - (n-1)S_3), \\ S_7 &= \frac{1}{4}S_2(n-1)(n^2 + 3n - 6 - (n^2 - n + 2)S_3), \\ S_8 &= S_0^2(n^2 - 3 - (n-1)^2S_3), \end{aligned} \quad (3.31)$$

pomen členov od  $S_0$  do  $S_3$  pa je enak kot pri Moranovem indeksu. Podobno kot pri Moranovem indeksu predpostavimo, da so geostatistićni podatki taki, da lahko tvorimo standardizirano normalno statistiko  $Z$  po enaćbi:

$$Z = \frac{C - E[C]}{\sqrt{\text{var}[C]}} \sim N(0, 1). \quad (3.32)$$

To je osnova za preizkus domneve o prostorski neodvisnosti na osnovi Gearyjevega razmerja. PokaŇimo sedaj primer enostranskega preizkuŇanja domneve o prostorski neodvisnosti. Nićelna domneva je enaka, kot pri dvostranskem preizkus, ki smo ga pokazali ob uporabi Moranovega indeksa. Pri enostranskem preizkusu v alternativni domnevi zapiŇemo, da so vrednosti zbrane v grućah ali pa da so vrednosti urejene po doloćenem vzorcu. V nadaljevanju bomo prikazali primer, ko so vrednosti zbrane v grućah.

<sup>3</sup> Robert Charles Geary, irski statistik (1896–1983).



$H_0$ : Slučajno polje je prostorsko neodvisno ( $C = 1$ ).

$H_1$ : Vrednosti slučajnega polja so zbrane v gručah ( $C < 1$ ).

$\alpha = 5\%$ .

Zapišemo statistiko  $Z$ :

$$Z = \frac{C - 1}{\sqrt{\text{var}[C]}} \quad (3.33)$$

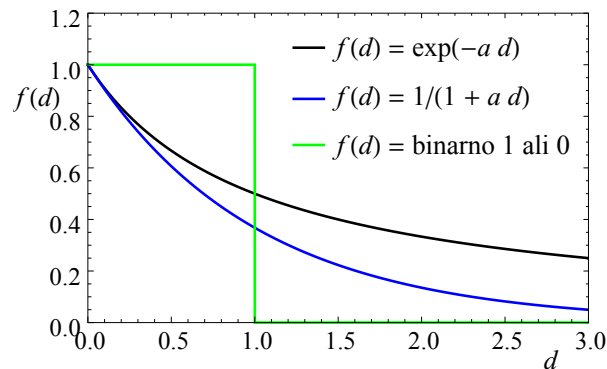
in postavimo meje kritičnega območja za ta enostranski preizkus domneve:

$$Z < -k_{1-\alpha} \longrightarrow H_0 \text{ zavrnemo.} \quad (3.34)$$

Če je torej  $Z < -k_{1-\alpha}$ , ničelno domnevo **zavrnemo**. Za  $\alpha = 0.05$  je kritična vrednost  $k_{0.95} = 1.645$ . V tem primeru izjavimo: „Vrednosti so statistično značilno zbrane v gručah!“ ali pa „S tveganjem, manjšim od  $\alpha = 5\%$ , lahko trdimo, da so vrednosti slučajnega polja zbrane v gručah!“

Če statistika  $Z$  ne pade v kritično območje ( $-k_{1-\alpha} \leq Z$ ), potem ničelne domneve **ne zavrnemo**. Tedaj izjavimo: „Vrednosti niso statistično značilno zbrane v gručah!“ ali pa „S tveganjem, manjšim ali enakim  $\alpha = 5\%$ , ne moremo trditi, da so vrednosti slučajnega polja zbrane v gručah!“

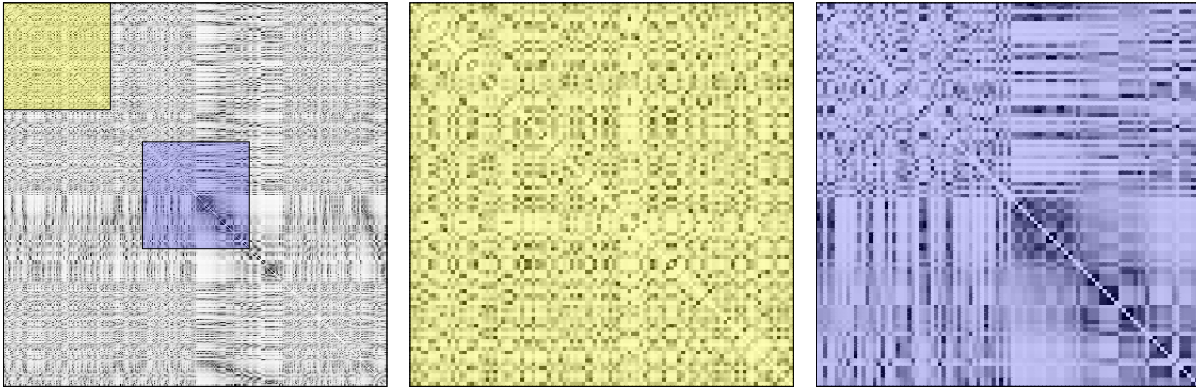
Določimo vrednost Moranovega indeksa  $I$  in Gearyjevega razmerja  $C$  ter naredimo preizkus domneve o prostorski neodvisnosti za koncentracijo kroma (Cr). Nalogo bomo opravili z različno definiranimi utežmi: v primeru binarnih uteži za kritično razdaljo vzamemo  $d_k = 1$  km. V primeru zveznih uteži po enačbi (3.19) pa vzamemo, da je  $a = 1 \text{ km}^{-1}$ . Vse tri funkcije so prikazane na sliki 3.15.



SLIKA 3.15: Grafi funkcij, s katerimi smo določili uteži  $w_{ij}$

Če uporabimo binarne uteži, lahko le-te prikažemo v mreži tako, da povezavo ( $w_{ij} = 1$ ) označimo s črnim kvadratom. V enačbi (3.35) in na sliki 3.16 zapišemo matriko binarnih uteži. Na sliki prikazujemo celotno matriko za vseh 359 točk ter v povečanem merilu dva dela te matrike od točke 1 do 100 in od točke 131 do 230.





SLIKA 3.17: Matrika uteži za eksponentno funkcijo (3.19) za vseh 359 točk ter za dva dela

Za vajo lahko eno vrednost uteži, prikazano v enačbah (3.35) in (3.36), izračunamo iz podatkov v preglednici 3.1. Utež med 1. in 2. točko  $w_{12}$  določimo takole:

koordinati 1. točke:  $\{2.386, 3.077\}$ ,

koordinati 2. točke:  $\{2.544, 1.972\}$ ,

razdalja med točkama:  $d = \sqrt{(2.544 - 2.386)^2 + (1.972 - 3.077)^2} = 1.116$  km,

binarna utež: ker je  $d > 1 \rightarrow w_{12} = 0$ ,

zvezna utež (eksponentna funkcija):  $w_{12} = e^{-1 \cdot 1.116} = 0.328$ .

Vidimo, da sta vrednosti uteži enaki kot v izrazu (3.35) oziroma (3.36).

V naslednji preglednici 3.7 prikazujemo povprečje  $\bar{Z}$ , vzorčno varianco  $S^2$  in statistike  $S_0$  do  $S_8$  za krom za tri različne načine določitve uteži.

PREGLEDNICA 3.7: Izračun statistik za krom

Statistika	$\bar{Z}$	$S^2$	$S_0$	$S_1$	$S_2$	$S_3$
Cr /binarna	35.02	113.4	22080	44160	$5.75 \cdot 10^6$	3.316
Cr /eksponentna	35.02	113.4	26701	20551	$8.22 \cdot 10^6$	3.316
Cr /racionalna	35.02	113.4	48340	41937	$2.62 \cdot 10^7$	3.316
Statistika	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	
Cr /binarna	$5.04 \cdot 10^9$	$4.47 \cdot 10^9$	$2.00 \cdot 10^{12}$	$-1.53 \cdot 10^{14}$	$-1.44 \cdot 10^{14}$	
Cr /eksponentna	$1.82 \cdot 10^9$	$1.02 \cdot 10^9$	$9.32 \cdot 10^{11}$	$-2.18 \cdot 10^{14}$	$-2.11 \cdot 10^{14}$	
Cr /racionalna	$2.95 \cdot 10^9$	$5.66 \cdot 10^8$	$1.90 \cdot 10^{12}$	$-6.96 \cdot 10^{14}$	$-6.92 \cdot 10^{14}$	

Iz zapisanih statistik lahko izračunamo Moranov indeks  $I$  in Gearyjevo razmerje  $C$  ter oceni standardnih

deviacij  $\bar{\sigma}_I$  in  $\bar{\sigma}_C$  ter statistiki  $Z_I$  in  $Z_C$ . Te rezultate prikazujemo v preglednici 3.8.

PREGLEDNICA 3.8: Izračun Moranovega indeksa in Gearyjevega razmerja za koncentracijo kroma

Statistika	$I$	$\bar{\sigma}_I$	$Z_I$	$\alpha_{dej}$
Cr /binarne	0.0732	0.00855	8.88	$6.49 \cdot 10^{-19}$
Cr /eksponentna	0.0329	0.00349	10.21	$1.73 \cdot 10^{-24}$
Cr /racionalna	0.0113	0.00146	9.62	$6.86 \cdot 10^{-22}$
Statistika	$C$	$\bar{\sigma}_C$	$Z_C$	$\alpha_{dej}$
Cr /binarne	0.918	0.0214	-3.821	0.00013
Cr /eksponentna	0.952	0.0154	-3.148	0.00164
Cr /racionalna	0.979	0.00739	-2.885	0.00391

Ugotovimo lahko, da je ne glede na tip uteži  $w_{ij}$  vrednost Moranovega indeksa  $I$  mnogo višja od kritične vrednosti  $k_{1-\alpha/2} = 1.96$ , zato moramo ničelno domnevo zavrnilo. Dejansko tveganje, da smo sprejeli napačno odločitev, je zanemarljivo majhno (manj od  $10^{-18}$ ). Izjavimo lahko, da je koncentracija kroma statistično značilno prostorsko odvisna. Pri preizkusu s Gearyjevim razmerjem pridemo do podobnih zaključkov, ničelno domnevo moramo zavrnilo, saj je statistika  $Z_C$  po absolutni vrednosti vedno večja od  $k_{1-\alpha/2} = 1.96$ . Dejansko tveganje, da smo sprejeli napačno odločitev, pa je nekoliko višje kot pri preizkusu z Moranovim indeksom (od 0.0039 do 0.016), a je vseeno manjše od določenega tveganja  $\alpha = 5\%$ . Zato tudi v tem primeru zaključimo, da je koncentracija kroma statistično značilno prostorsko odvisna.

## 4 Krigiranje

Krigiranje je ena od pomembnejših metod, ki jih uporabljamo v prostorski statistiki. Imenuje se po južnoafriškem rudarskem inženirju D. G. Krigeju<sup>1</sup>. V različnih disciplinah se pogosto najdemo pred problemom, kjer poznamo vrednosti slučajnega polja  $Z(\mathbf{s}_i)$  v končno mnogo točkah  $\mathbf{s}_i$ . Naša naloga je, da ocenimo oziroma interpoliramo vrednosti polja v poljubi točki  $\mathbf{s}$  obravnavanega območja, v kateri nimamo meritev. Interpolacijo lahko opravimo na različne načine: z linearno regresijo oziroma interpolacijo, nelinearno interpolacijo, interpolacijo z zlepkami (angl. spline), interpolacijo z radialnimi baznimi funkcijami, interpolacijo z umetnimi nevronskimi mrežami in seveda, s **krigiranjem**.

Ideja interpolacije s krigiranjem je dokaj preprosta, saj izhaja iz enačbe za uteženo povprečje vrednosti polja  $Z(\mathbf{s}_i)$  v določenem številu  $n(\mathbf{s})$  točk v okolici točke  $\mathbf{s}$ :

$$Z^*(\mathbf{s}) = m(\mathbf{s}) + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) (Z(\mathbf{s}_i) - m(\mathbf{s}_i)), \quad (4.1)$$

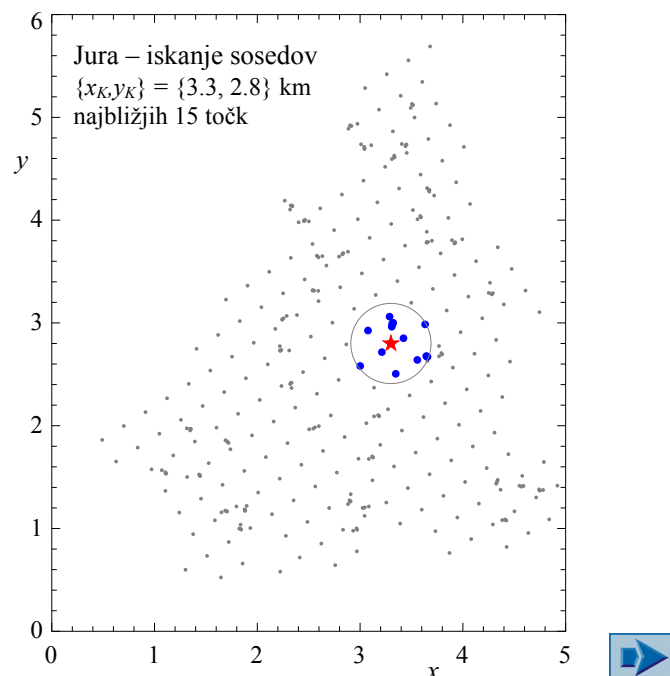
kjer je  $Z^*(\mathbf{s})$  interpolirana oziroma krigirana vrednost polja v točki  $\mathbf{s}$ ,  $\lambda_i(\mathbf{s})$  so uteži, ki jih moramo določiti za vsako interpolacijo posebej,  $Z(\mathbf{s}_i)$  so znane vrednosti polja v točkah  $\mathbf{s}_i$ ,  $m(\mathbf{s})$  in  $m(\mathbf{s}_i)$  so srednje vrednosti polja za točke  $\mathbf{s}$  in  $\mathbf{s}_i$ . Različni načini krigiranja se ločijo po načinu določitve srednje vrednosti  $m(\mathbf{s})$ : **preprosto**, **običajno krigiranje** in **krigiranje s trendom**. Podrobnosti različnih načinov krigiranja so podane v nadaljevanju.

Celoten postopek krigiranja lahko zajamemo v naslednjih točkah:

- 1) Iz geostatističnih podatkov  $Z(\mathbf{s}_i)$  določimo vzorčni variogram (glej sliko 3.7).
- 2) Na osnovi vzorčnega variograma zapišemo empirično funkcijo variograma, njene koeficiente določimo po metodi najmanjših kvadratov (glej sliko 4.4).
- 3) Za izbrano točko  $\mathbf{s} = \{x_K, y_K\}$  naredimo:
  - 3.1) Poiščemo  $n(\mathbf{s})$  sosedov (glej sliko 4.1).
  - 3.2) Določimo vrednosti uteži in srednjih vrednosti.
  - 3.3) Izračunamo krigirano (interpolirano) vrednost.
  - 3.4) Se vrnemo na 3), dokler ne interpoliramo v vseh izbranih točkah.
- 4) Grafično ali numerično prikažemo rezultate interpolacije.

<sup>1</sup> Daniel Gerhardus Krige, južnoafriški rudarski inženir, avtor interpolacijske metode – krigiranja (1919–2013).

Na sliki 4.1 prikazujemo točko ali točke, v katerih želimo opraviti interpolacijo, ter sosedo, ki so v tem primeru definirane kot prvih  $n(s)$  najbližjih točk.



SLIKA 4.1: Iskanje najbližjih točk v okolici točke, kjer želimo interpolirati vrednost polja

## 4.1 Preprosto krigiranje

Pri preprostem krigiranju predpostavimo, da je srednja vrednost enaka na celotnem območju krigiranja:  $m(\mathbf{s}) = m(\mathbf{s}_i) = m$ . Oceno srednje vrednosti  $\hat{m}$  določimo vnaprej s preprostim povprečjem vseh merjenih vrednosti slučajnega polja:

$$\hat{m} = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i. \quad (4.2)$$

Enačba krigiranja se sedaj nekoliko poenostavi:

$$Z^*(\mathbf{s}) = m + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) (Z(\mathbf{s}_i) - m). \quad (4.3)$$

Cilj postopka krigiranja je, da poiščemo take uteži  $\lambda_i(\mathbf{s})$ , da bodo razlike med  $Z^*(\mathbf{s})$  in  $Z(\mathbf{s})$  kar najmanjše. Ta pogoj izpolnimo tako, da določimo uteži  $\lambda_i(\mathbf{s})$  tako, da je varianca razlike med krigirano in pravo vrednostjo najmanjša:

$$\text{var}[Z^*(\mathbf{s}) - Z(\mathbf{s})] \dots \text{čim manjša}. \quad (4.4)$$

Zato da bomo imeli pri izpeljavi enostavnejše zapise, vpeljemo novo količino  $R(\mathbf{s})$ , ki jo imenujemo ostanek ali rezidual in predstavlja razliko med vrednostjo funkcije in njeno srednjo vrednostjo  $m$ :

$$\begin{aligned} R(\mathbf{s}) &= Z(\mathbf{s}) - m, \\ R^*(\mathbf{s}) &= Z^*(\mathbf{s}) - m, \\ R(\mathbf{s}_i) &= Z(\mathbf{s}_i) - m. \end{aligned} \quad (4.5)$$

Iz definicije reziduala sledi, da je pričakovana vrednost reziduala enaka nič:

$$E[R(\mathbf{s})] = E[Z(\mathbf{s}) - m] = E[Z(\mathbf{s})] - m = m - m = 0. \quad (4.6)$$

Hitro lahko tudi ugotovimo, da je razlika med krigirano in pravo vrednostjo slučajnega polja enaka razliki med krigirano in pravo vrednostjo reziduala:

$$Z^*(\mathbf{s}) - Z(\mathbf{s}) = (Z^*(\mathbf{s}) - m) - (Z(\mathbf{s}) - m) = R^*(\mathbf{s}) - R(\mathbf{s}), \quad (4.7)$$

zaradi česar je enaka tudi varianca teh razlik:

$$\text{var}[Z^*(\mathbf{s}) - Z(\mathbf{s})] = \text{var}[R^*(\mathbf{s}) - R(\mathbf{s})]. \quad (4.8)$$

Ker so vrednosti reziduala le za konstanto spremenjene vrednosti slučajnega polja, lahko pričakujemo, da so vse lastnosti reziduala  $R$ , povezane z variabilnostjo in prostorsko odvisnostjo, enake kot za slučajno polje  $Z$ . To pomeni, da so kovariančna funkcija  $C_R(\mathbf{h})$ , variogram  $\gamma_R(\mathbf{h})$  in varianca  $C_R(0)$  reziduala enake kovariančni funkciji  $C(\mathbf{h})$ , variogramu  $\gamma(\mathbf{h})$  in varianci  $C(0)$  slučajnega polja:

$$C_R(\mathbf{h}) = C(\mathbf{h}), \quad \gamma_R(\mathbf{h}) = \gamma(\mathbf{h}), \quad C_R(0) = C(0). \quad (4.9)$$

Zadnjo trditev preprosto dokažemo tako, da primerjamo kovariančni funkciji slučajnega polja in reziduala med dvema točkama  $\mathbf{s}_i$  in  $\mathbf{s}_j$ . Zapišimo najprej kovariančno funkcijo slučajnega polja  $Z$ :

$$C(\mathbf{s}_i - \mathbf{s}_j) = \text{cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] = E[(Z(\mathbf{s}_i) - m)(Z(\mathbf{s}_j) - m)] = E[Z(\mathbf{s}_i)Z(\mathbf{s}_j)] - m^2, \quad (4.10)$$

podobno zapišemo še za kovariančno funkcijo reziduala  $R$ :

$$C_R(\mathbf{s}_i - \mathbf{s}_j) = \text{cov}[R(\mathbf{s}_i), R(\mathbf{s}_j)] = E[(R(\mathbf{s}_i) - 0)(R(\mathbf{s}_j) - 0)] = E[R(\mathbf{s}_i)R(\mathbf{s}_j)]. \quad (4.11)$$

Če v enačbo (4.11) vstavimo tretjega od izrazov (4.5), dobimo:

$$C_R(\mathbf{s}_i - \mathbf{s}_j) = E[(Z(\mathbf{s}_i) - m)(Z(\mathbf{s}_j) - m)] = E[Z(\mathbf{s}_i)Z(\mathbf{s}_j)] - m^2 = C(\mathbf{s}_i - \mathbf{s}_j). \quad (4.12)$$

Zato bomo v nadaljevanju tudi kovariančno funkcijo reziduala označili kar s  $C(\mathbf{h})$ . Na enak način lahko pokažemo enakost tudi za variogram in varianco.

S primerjavo enačb (4.3) in (4.5) izpeljemo enačbo krigiranja za rezidual:

$$R^*(\mathbf{s}) = \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s})R(\mathbf{s}_i). \quad (4.13)$$

Zapišimo varianco razlike med krigirano in pravo vrednostjo reziduala v odvisnosti od neznanih uteži  $\lambda_i(\mathbf{s})$  in znanih vrednosti kovariančne funkcije  $C(\mathbf{h})$ :

$$\text{var}[R^*(\mathbf{s}) - R(\mathbf{s})] = \text{var}[R^*(\mathbf{s})] + \text{var}[R(\mathbf{s})] - 2 \text{cov}[R^*(\mathbf{s}), R(\mathbf{s})], \quad (4.14)$$

kjer smo uporabili znano zvezo za izračun variance linearne kombinacije slučajnih spremenljivk ( $\text{var}[a Z_1 + b Z_2] = a^2 \text{var}[Z_1] + b^2 \text{var}[Z_2] + 2 ab \text{cov}[Z_1, Z_2]$ ). Po vrsti bomo zapisali člene iz enačbe (4.14):

$$\begin{aligned} \text{var}[R^*(\mathbf{s})] &= E[R^*(\mathbf{s}) R^*(\mathbf{s})] = \\ &\text{vključimo enačbo (4.13)} \\ &= E \left[ \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) R(\mathbf{s}_i) \sum_{j=1}^{n(\mathbf{s})} \lambda_j(\mathbf{s}) R(\mathbf{s}_j) \right] = \\ &= E \left[ \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) R(\mathbf{s}_i) R(\mathbf{s}_j) \right] = \end{aligned} \quad (4.15)$$

upoštevamo, da je  $E[\cdot]$  linearni operator in da uteži  $\lambda_i(\mathbf{s})$  niso slučajne

$$= \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) E [R(\mathbf{s}_i) R(\mathbf{s}_j)] =$$

zapišemo kovariančno funkcijo (enačbi (4.11) in (4.12)) in dobimo končni rezultat

$$= \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}_j).$$

Z drugim členom nimamo dosti dela, saj varianco reziduala preprosto zapišemo kot kovariančno funkcijo za  $h = 0$ :

$$\text{var}[R(\mathbf{s})] = C(0). \quad (4.16)$$



Tretji člen izpišemo podobno, kot prvega:

$$2 \operatorname{cov}[R^*(\mathbf{s}), R(\mathbf{s})] = 2 E[R^*(\mathbf{s}) R(\mathbf{s})] =$$

vključimo enačbo (4.13)

$$= 2 E \left[ \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) R(\mathbf{s}_i) R(\mathbf{s}) \right] =$$

upoštevamo, da je  $E[\cdot]$  linearni operator in da uteži  $\lambda_i(\mathbf{s})$  niso slučajne

$$= 2 \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) E[R(\mathbf{s}_i) R(\mathbf{s})] =$$

zapišemo kovariančno funkcijo (enačbi (4.11) in (4.12)) in dobimo končni rezultat

$$= 2 \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}).$$

Varianca razlike krigiranega in pravega reziduala je odvisna le od neznanih uteži  $\lambda_i(\mathbf{s})$  ter znanih vrednosti kovariančne funkcije  $C(\mathbf{s}_i - \mathbf{s}_j)$  in  $C(\mathbf{s}_i - \mathbf{s})$  in variance  $C(0)$ :

$$\operatorname{var}[R^*(\mathbf{s}) - R(\mathbf{s})] = \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}_j) + C(0) - 2 \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}). \quad (4.18)$$

Z odvajanjem enačbe (4.18) po utežeh  $\lambda_i(\mathbf{s})$  lahko določimo take vrednosti uteži  $\lambda_i(\mathbf{s})$ , da bo varianca najmanjša:

$$\frac{\partial(\operatorname{var}[R^*(\mathbf{s}) - R(\mathbf{s})])}{\partial \lambda_k(\mathbf{s})} = 2 \sum_{j=1}^{n(\mathbf{s})} \lambda_j(\mathbf{s}) C(\mathbf{s}_k - \mathbf{s}_j) - 2 C(\mathbf{s}_k - \mathbf{s}) = 0, \quad (4.19)$$

kjer je  $k = 1, \dots, n(\mathbf{s})$ . Zgornja enačba predstavlja sistem  $n(\mathbf{s})$  linearnih enačb za prav toliko neznank  $\lambda_i(\mathbf{s})$ . Izpišimo nekaj enačb tega sistema v razviti obliki:

$$\begin{aligned} C(0)\lambda_1(\mathbf{s}) &+ C(\mathbf{s}_1 - \mathbf{s}_2)\lambda_2(\mathbf{s}) &+ \dots &+ C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})})\lambda_{n(\mathbf{s})}(\mathbf{s}) &= C(\mathbf{s}_1 - \mathbf{s}), \\ C(\mathbf{s}_2 - \mathbf{s}_1)\lambda_1(\mathbf{s}) &+ C(0)\lambda_2(\mathbf{s}) &+ \dots &+ C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})})\lambda_{n(\mathbf{s})}(\mathbf{s}) &= C(\mathbf{s}_2 - \mathbf{s}), \\ \vdots & & & & \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1)\lambda_1(\mathbf{s}) &+ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_2)\lambda_2(\mathbf{s}) &+ \dots &+ C(0)\lambda_{n(\mathbf{s})}(\mathbf{s}) &= C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}). \end{aligned}$$

Sistem linearnih enačb običajno zapišemo v matrični obliki:

$$\begin{bmatrix} C(0) & C(\mathbf{s}_1 - \mathbf{s}_2) & \dots & C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})}) \\ C(\mathbf{s}_2 - \mathbf{s}_1) & C(0) & \dots & C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})}) \\ \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1) & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_2) & \dots & C(0) \end{bmatrix} \begin{bmatrix} \lambda_1(\mathbf{s}) \\ \lambda_2(\mathbf{s}) \\ \vdots \\ \lambda_{n(\mathbf{s})}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} C(\mathbf{s}_1 - \mathbf{s}) \\ C(\mathbf{s}_2 - \mathbf{s}) \\ \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}) \end{bmatrix}. \quad (4.20)$$

Z rešitvijo sistema enačb (4.20) določimo vrednosti uteži, ki jih uporabimo pri izračunu interpolacije slučajnega polja po enačbi (4.3).

## 4.2 Običajno krigiranje

Pri običajnem krigiranju predpostavimo, da je v okolici točke, za katero interpoliramo vrednost, srednja vrednost  $m(\mathbf{s})$  konstantna, a neznana. Zato predpostavimo:

$$m(\mathbf{s}) = m(\mathbf{s}_i) = \mu(\mathbf{s}). \quad (4.21)$$

Enačbo krigiranja lahko zapišemo takole:

$$Z^*(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) (Z(\mathbf{s}_i) - \mu(\mathbf{s})),$$

po manjši preureditvi pa v naslednji obliki:

$$Z^*(\mathbf{s}) = \mu(\mathbf{s}) \left( 1 - \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \right) + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) Z(\mathbf{s}_i). \quad (4.22)$$

Ker ne želimo, da bi izbira srednje vrednosti vplivala na krigirano vrednost, zahtevamo, da je prvi člen v enačbi (4.22) enak nič. To lahko dosežemo le, če je izpolnjen pogoj:

$$\sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) = 1. \quad (4.23)$$

Prišli smo do problema vezanih ekstremov, kjer mora biti ob določitvi minimuma  $var[Z^*(\mathbf{s}) - Z(\mathbf{s})]$  izpolnjen tudi pogoj (4.23). Problem rešimo z uvedbo Lagrangevega<sup>2</sup> množitelja  $2\mu(\mathbf{s})$ :

$$L(\lambda_i(\mathbf{s}), \mu(\mathbf{s})) = var[Z^*(\mathbf{s}) - Z(\mathbf{s})] + 2\mu(\mathbf{s}) \left( \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) - 1 \right) \quad (4.24)$$

oziroma v razviti obliki:

$$\begin{aligned} L(\lambda_i(\mathbf{s}), \mu(\mathbf{s})) &= \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}_j) + C(0) - \\ &\quad - 2 \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}) + 2\mu(\mathbf{s}) \left( \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) - 1 \right). \end{aligned} \quad (4.25)$$

<sup>2</sup> Joseph-Louis Lagrange, francoski matematik, rojen v Torinu (1736–1813).

Z odvajanjem  $L(\lambda_i(\mathbf{s}), \mu(\mathbf{s}))$  po  $\lambda_i(\mathbf{s})$  in  $\mu(\mathbf{s})$  dobimo sistem  $n(\mathbf{s}) + 1$  linearnih enačb za prav toliko neznank:

$$\begin{aligned} \frac{\partial L(\lambda_i(\mathbf{s}), \mu(\mathbf{s}))}{\partial \lambda_k(\mathbf{s})} &= 0, \\ \frac{\partial L(\lambda_i(\mathbf{s}), \mu(\mathbf{s}))}{\partial \mu(\mathbf{s})} &= 0. \end{aligned} \quad (4.26)$$

Izpišimo nekaj značilnih enačb sistema, pri čemer zaradi preglednejšega zapisa upoštevamo  $\lambda_i(\mathbf{s}) = \lambda_i$  in  $\mu(\mathbf{s}) = \mu$ :

$$\begin{array}{cccccccc} C(0) \lambda_1 & + & C(\mathbf{s}_1 - \mathbf{s}_2) \lambda_2 & + \cdots & + & C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})}) \lambda_{n(\mathbf{s})} & + & 1 \mu & = & C(\mathbf{s}_1 - \mathbf{s}), \\ C(\mathbf{s}_2 - \mathbf{s}_1) \lambda_1 & + & C(0) \lambda_2 & + \cdots & + & C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})}) \lambda_{n(\mathbf{s})} & + & 1 \mu & = & C(\mathbf{s}_2 - \mathbf{s}), \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots & \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1) \lambda_1 & + & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_2) \lambda_2 & + \cdots & + & C(0) \lambda_{n(\mathbf{s})} & + & 1 \mu & = & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}), \\ 1 \mu \lambda_1 & + & 1 \lambda_2 & + \cdots & + & 1 \lambda_{n(\mathbf{s})} & + & 0 \mu & = & 1. \end{array}$$

Sistem linearnih enačb običajno zapišemo v matrični obliki:

$$\begin{bmatrix} C(0) & C(\mathbf{s}_1 - \mathbf{s}_2) & \cdots & C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})}) & 1 \\ C(\mathbf{s}_2 - \mathbf{s}_1) & C(0) & \cdots & C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1) & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_2) & \cdots & C(0) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n(\mathbf{s})} \\ \mu \end{bmatrix} = \begin{bmatrix} C(\mathbf{s}_1 - \mathbf{s}) \\ C(\mathbf{s}_2 - \mathbf{s}) \\ \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}) \\ 1 \end{bmatrix}. \quad (4.27)$$

Ker z dodatno enačbo izpolnimo pogoj (4.23), pri določitvi krigirane vrednosti  $Z^*(\mathbf{s})$  po enačbi (4.22) lahko izpustimo prvi člen in se enačba krigiranja še poenostavi:

$$Z^*(\mathbf{s}) = \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) Z(\mathbf{s}_i). \quad (4.28)$$

### 4.3 Krigiranje s trendom

Pri krigiranju s trendom predpostavimo, da je v okolici točke, za katero interpoliramo vrednost, srednja vrednost  $m(\mathbf{s})$  oziroma  $m(\mathbf{s}_i)$  znana funkcija z neznanimi koeficienti. Zato predpostavimo:

$$m(\mathbf{s}) = \sum_{k=1}^K \alpha_k(\mathbf{s}) f_k(\mathbf{s}) \quad \text{oziroma} \quad m(\mathbf{s}_i) = \sum_{k=1}^K \alpha_k(\mathbf{s}) f_k(\mathbf{s}_i), \quad (4.29)$$

kjer je  $K$  število neznanih koeficientov  $\alpha_k(\mathbf{s})$  oziroma število funkcij  $f_k(\mathbf{s})$ , ki sestavljajo funkcijo trenda. Koeficiente trenda določimo v postopku krigiranja skupaj z utežmi  $\lambda_i(\mathbf{s})$ . Enačbo krigiranja

lahko zapišemo takole:

$$Z^*(\mathbf{s}) = \sum_{k=1}^K \alpha_k(\mathbf{s}) f_k(\mathbf{s}) + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \left( Z(\mathbf{s}_i) - \sum_{k=1}^K \alpha_k(\mathbf{s}) f_k(\mathbf{s}_i) \right),$$

po manjši preureditvi pa v naslednji obliki:

$$Z^*(\mathbf{s}) = \sum_{k=1}^K \alpha_k(\mathbf{s}) \left( f_k(\mathbf{s}) - \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) f_k(\mathbf{s}_i) \right) + \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) Z(\mathbf{s}_i). \quad (4.30)$$

Ker ne želimo, da bi izbira funkcije trenda vplivala na krigirano vrednost, zahtevamo, da je prvi člen v enačbi (4.30) enak nič. To lahko dosežemo le, če so izpolnjeni pogoji:

$$\sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) f_k(\mathbf{s}_i) = f_k(\mathbf{s}), \quad (4.31)$$

za  $k = 1, \dots, K$ . Prišli smo do problema vezanih ekstremov, kjer morajo biti ob določitvi minimuma  $\text{var}[Z^*(\mathbf{s}) - Z(\mathbf{s})]$  izpolnjeni tudi pogoji (4.31). Problem rešimo z uvedbo Lagrangevih množiteljev  $2\alpha_k(\mathbf{s})$ :

$$L(\lambda_i(\mathbf{s}), \alpha_k(\mathbf{s})) = \text{var}[Z^*(\mathbf{s}) - Z(\mathbf{s})] + 2 \sum_{k=1}^K \alpha_k(\mathbf{s}) \left( \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) f_k(\mathbf{s}_i) - f_k(\mathbf{s}) \right) \quad (4.32)$$

oziroma v razviti obliki:

$$\begin{aligned} L(\lambda_i(\mathbf{s}), \alpha_k(\mathbf{s})) &= \sum_{i=1}^{n(\mathbf{s})} \sum_{j=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) \lambda_j(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}_j) + C(0) - \\ &\quad - 2 \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) C(\mathbf{s}_i - \mathbf{s}) + \\ &\quad + 2 \sum_{k=1}^K \alpha_k(\mathbf{s}) \left( \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) f_k(\mathbf{s}_i) - f_k(\mathbf{s}) \right). \end{aligned} \quad (4.33)$$

Z odvajanjem  $L(\lambda_i(\mathbf{s}), \alpha_k(\mathbf{s}))$  po  $\lambda_i(\mathbf{s})$  in  $\alpha_k(\mathbf{s})$  dobimo sistem  $n(\mathbf{s}) + K$  lineranih enačb za prav toliko neznank:

$$\begin{aligned} \frac{\partial L(\lambda_i(\mathbf{s}), \alpha_k(\mathbf{s}))}{\partial \lambda_l(\mathbf{s})} &= 0, \\ \frac{\partial L(\lambda_i(\mathbf{s}), \alpha_k(\mathbf{s}))}{\partial \alpha_m(\mathbf{s})} &= 0. \end{aligned} \quad (4.34)$$

Izberemo lahko različne enačbe trenda. Pogosto uporabimo najpreprostejši linearni trend, ki ga za dve dimenziji zapišemo z naslednjo enačbo:

$$m(\mathbf{s}) = m(x, y) = a_1 + a_2 x + a_3 y. \quad (4.35)$$

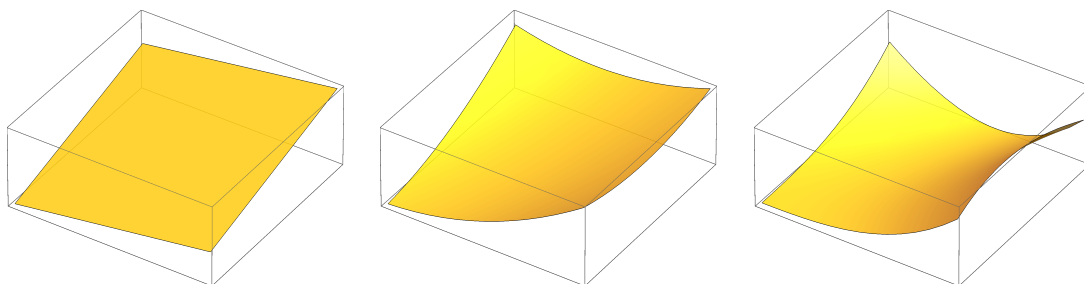
V primeru, da želimo uporabiti kvadratni trend, je enačba trenda enaka:

$$m(\mathbf{s}) = m(x, y) = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 xy. \quad (4.36)$$

Pri linearnem trendu v dveh dimenzijah moramo določiti tri neznane koeficiente  $K = 3$ , pri kvadratnem trendu pa je  $K = 6$ . Na sliki 4.2 prikazujemo tri različne funkcije trenda, ki bi jih lahko uporabili pri krigiranju s trendom: linearni, kvadratni in kubični trend. V primeru, da izberemo konstantno funkcijo trenda:

$$m(\mathbf{s}) = m(x, y) = a_1, \quad (4.37)$$

se primer poenostavi v običajno krigiranje, ki smo ga obravnavali v prejšnjem poglavju.



SLIKA 4.2: Primeri funkcij trenda (linearna, kvadratna in kubična funkcija)

Izpišimo nekaj značilnih enačb sistema za linearno enačbo trenda (4.35), pri čemer zaradi preglednejšega zapisa upoštevamo  $\lambda_i(\mathbf{s}) = \lambda_i$  in  $\alpha_k(\mathbf{s}) = \alpha_k$ :

$$\begin{array}{rcccccccc} C(0) \lambda_1 & + \cdots + & C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})}) \lambda_{n(\mathbf{s})} & + 1 \alpha_1 & + x_1 \alpha_2 & + y_1 \alpha_3 & = & C(\mathbf{s}_1 - \mathbf{s}), \\ C(\mathbf{s}_2 - \mathbf{s}_1) \lambda_1 & + \cdots + & C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})}) \lambda_{n(\mathbf{s})} & + 1 \alpha_1 & + x_2 \alpha_2 & + y_2 \alpha_3 & = & C(\mathbf{s}_2 - \mathbf{s}), \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1) \lambda_1 & + \cdots + & C(0) \lambda_{n(\mathbf{s})} & + 1 \alpha_1 & + x_{n(\mathbf{s})} \alpha_2 & + y_{n(\mathbf{s})} \alpha_3 & = & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}), \\ 1 \lambda_1 & + \cdots + & 1 \lambda_{n(\mathbf{s})} & + 0 \alpha_1 & + 0 \alpha_2 & + 0 \alpha_3 & = & 1, \\ x_1 \lambda_1 & + \cdots + & x_{n(\mathbf{s})} \lambda_{n(\mathbf{s})} & + 0 \alpha_1 & + 0 \alpha_2 & + 0 \alpha_3 & = & x, \\ y_1 \lambda_1 & + \cdots + & y_{n(\mathbf{s})} \lambda_{n(\mathbf{s})} & + 0 \alpha_1 & + 0 \alpha_2 & + 0 \alpha_3 & = & y. \end{array}$$

Sistem linearnih enačb običajno zapišemo v matrični obliki:

$$\left[ \begin{array}{cccc|ccc} C(0) & C(\mathbf{s}_1 - \mathbf{s}_2) & \cdots & C(\mathbf{s}_1 - \mathbf{s}_{n(\mathbf{s})}) & 1 & x_1 & y_1 \\ C(\mathbf{s}_2 - \mathbf{s}_1) & C(0) & \cdots & C(\mathbf{s}_2 - \mathbf{s}_{n(\mathbf{s})}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_1) & C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}_2) & \cdots & C(0) & 1 & x_{n(\mathbf{s})} & y_{n(\mathbf{s})} \\ \hline 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ \hline x_1 & x_2 & \cdots & x_{n(\mathbf{s})} & 0 & 0 & 0 \\ y_1 & y_2 & \cdots & y_{n(\mathbf{s})} & 0 & 0 & 0 \end{array} \right] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n(\mathbf{s})} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} C(\mathbf{s}_1 - \mathbf{s}) \\ C(\mathbf{s}_2 - \mathbf{s}) \\ \vdots \\ C(\mathbf{s}_{n(\mathbf{s})} - \mathbf{s}) \\ 1 \\ x \\ y \end{bmatrix}. \quad (4.38)$$

Ker z dodatnimi  $K$  enačbami zagotovimo, da so izpolnjeni pogoji (4.31), lahko prvi člen v enačbi (4.30) izpustimo in se enačba krigiranja še poenostavi:

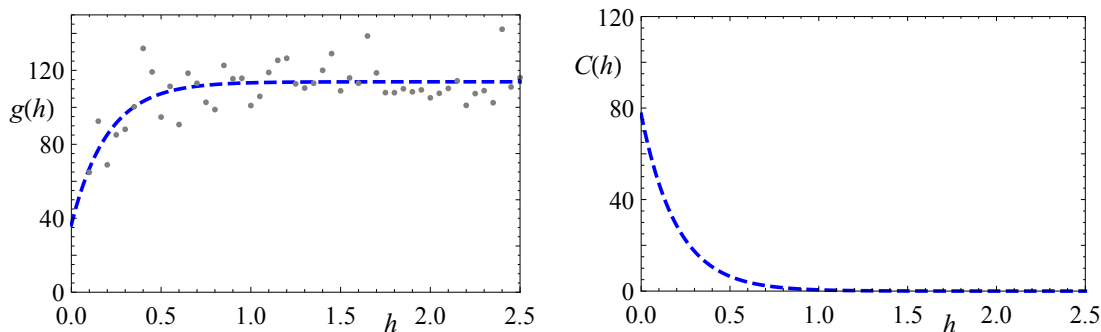
$$Z^*(\mathbf{s}) = \sum_{i=1}^{n(\mathbf{s})} \lambda_i(\mathbf{s}) Z(\mathbf{s}_i). \quad (4.39)$$

S primerjavo enačb (4.20), (4.27) in (4.38) vidimo, da prvih  $n(\mathbf{s})$  enačb v (4.27) in (4.38) predstavlja sistem preprostega krigiranja, prvih  $n(\mathbf{s}) + 1$  enačb v (4.38) pa predstavlja sistem enačb običajnega krigiranja. Različni deli sistema so v enačbi (4.38) ločeni z rdečimi črtami.

## 4.4 Primeri krigiranja

Vrnimo se na že večkrat obravnavani primer geostatističnih podatkov z območja Jure. Ocenimo vrednost koncentracije kroma v točki  $T_k$  ( $x_k = 3.3$  km in  $y_k = 2.8$  km), ki leži približno na sredini obravnavanega območja. Točka s pripadajočimi točkami v okolici je prikazana na sliki 4.1. Na podlagi podatkov o koncentraciji kroma smo ocenili variogram, pripravili empirični variogram in ustrezno kovariančno funkcijo, ki sta prikazana na sliki 4.3. Pri aproksimaciji variograma smo uporabili enačbo (3.7), vrednosti koeficientov pa so prikazane v preglednici 3.5. Pri določitvi empirične kovariančne funkcije, ki jo potrebujemo pri krigiranju, smo uporabili zvezo:

$$C(h) = C(0) - \gamma(h) = 77.3 e^{-4.97h}. \quad (4.40)$$



SLIKA 4.3: Empirični variogram in ustrezna kovariančna funkcija

V krigranje vključimo deset najbližjih točk v okolici točke  $T_k$ . Koordinate teh točk in koncentracije kroma v teh točkah so prikazane v preglednici 4.1. Za izračun uteži  $\lambda_i$  moramo izračunati razdalje med temi desetimi točkami, razdalje med temi desetimi točkami in točko  $T_k$  ter uporabiti enačbo empirične kovariančne funkcije (4.40).

#### 4.4.1 Preprosto krigranje

Pri preprostem krigranju moramo oceno srednje vrednosti določiti vnaprej:

$$m = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i = 35.02. \quad (4.41)$$

PREGLEDNICA 4.1: Koordinate in vrednosti koncentracije kroma v 10 najbližjih točkah točki  $T_k$ 

$x_i$	$y_i$	$Z_i$
3.316	2.987	33.00
3.077	2.926	35.32
3.312	2.991	32.60
3.347	2.505	33.20
3.322	3.001	35.00
3.308	2.963	36.00
3.422	2.851	38.72
3.557	2.640	25.72
3.212	2.716	35.52
3.287	3.061	45.92

Sistem enačb za določitev uteži pri preprostem krigiranju je:

$$\begin{bmatrix} 77.3 & 22.7 & 75.2 & 7.0 & 71.7 & 68.2 & 32.8 & 9.5 & 18.3 & 52.1 \\ 22.7 & 77.3 & 23.0 & 6.4 & 21.6 & 24.2 & 13.4 & 4.8 & 22.4 & 22.4 \\ 75.2 & 23.0 & 77.3 & 6.9 & 72.1 & 67.2 & 31.9 & 9.2 & 18.1 & 53.4 \\ 7.0 & 6.4 & 6.9 & 77.3 & 6.6 & 7.9 & 13.3 & 22.4 & 22.3 & 4.8 \\ 71.7 & 21.6 & 72.1 & 6.6 & 77.3 & 63.2 & 31.6 & 9.1 & 16.9 & 54.7 \\ 68.2 & 24.2 & 67.2 & 7.9 & 63.2 & 77.3 & 34.9 & 10.2 & 20.7 & 47.0 \\ 32.8 & 13.4 & 31.9 & 13.3 & 31.6 & 34.9 & 77.3 & 22.3 & 22.4 & 22.4 \\ 9.5 & 4.8 & 9.2 & 22.4 & 9.1 & 10.2 & 22.3 & 77.3 & 13.4 & 6.4 \\ 18.3 & 22.4 & 18.1 & 22.3 & 16.9 & 20.7 & 22.4 & 13.4 & 77.3 & 13.4 \\ 52.1 & 22.4 & 53.4 & 4.8 & 54.7 & 47.0 & 22.4 & 6.4 & 13.4 & 77.3 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \end{bmatrix} = \begin{bmatrix} 30.4 \\ 21.7 \\ 29.9 \\ 17.5 \\ 28.3 \\ 34.4 \\ 40.1 \\ 17.2 \\ 42.2 \\ 21.1 \end{bmatrix},$$

rešitev tega sistema pa je:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \end{bmatrix} = \begin{bmatrix} -0.0102 \\ 0.0586 \\ 0.0026 \\ 0.0327 \\ -0.0224 \\ 0.2103 \\ 0.3036 \\ 0.0328 \\ 0.3784 \\ -0.0090 \end{bmatrix}.$$

Vrednost koncentracije kroma v točki  $T_k$  pri preprostem krigiranju izračunamo po enačbi (4.3):

$$Z^*(3.3, 2.8) = 35.02 + (-0.0102 (33.00 - 35.02) + 0.0586 (35.32 - 35.02) + \dots) = 36.11.$$

#### 4.4.2 Običajno krigiranje

Na podoben način izračunamo tudi uteži in oceno srednje vrednosti pri običajnem krigiranju. Sistem enačb se poveča za eno enačbo in eno neznancko  $\mu$ :

$$\begin{bmatrix} 77.3 & 22.7 & 75.2 & 7.0 & 71.7 & 68.2 & 32.8 & 9.5 & 18.3 & 52.1 & 1.0 \\ 22.7 & 77.3 & 23.0 & 6.4 & 21.6 & 24.2 & 13.4 & 4.8 & 22.4 & 22.4 & 1.0 \\ 75.2 & 23.0 & 77.3 & 6.9 & 72.1 & 67.2 & 31.9 & 9.2 & 18.1 & 53.4 & 1.0 \\ 7.0 & 6.4 & 6.9 & 77.3 & 6.6 & 7.9 & 13.3 & 22.4 & 22.3 & 4.8 & 1.0 \\ 71.7 & 21.6 & 72.1 & 6.6 & 77.3 & 63.2 & 31.6 & 9.1 & 16.9 & 54.7 & 1.0 \\ 68.2 & 24.2 & 67.2 & 7.9 & 63.2 & 77.3 & 34.9 & 10.2 & 20.7 & 47.0 & 1.0 \\ 32.8 & 13.4 & 31.9 & 13.3 & 31.6 & 34.9 & 77.3 & 22.3 & 22.4 & 22.4 & 1.0 \\ 9.5 & 4.8 & 9.2 & 22.4 & 9.1 & 10.2 & 22.3 & 77.3 & 13.4 & 6.4 & 1.0 \\ 18.3 & 22.4 & 18.1 & 22.3 & 16.9 & 20.7 & 22.4 & 13.4 & 77.3 & 13.4 & 1.0 \\ 52.1 & 22.4 & 53.4 & 4.8 & 54.7 & 47.0 & 22.4 & 6.4 & 13.4 & 77.3 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \\ \mu \end{bmatrix} = \begin{bmatrix} 30.4 \\ 21.7 \\ 29.9 \\ 17.5 \\ 28.3 \\ 34.4 \\ 40.1 \\ 17.2 \\ 42.2 \\ 21.1 \\ 1.0 \end{bmatrix},$$



rešitev tega sistema pa je:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \\ \mu \end{bmatrix} = \begin{bmatrix} -0.0099 \\ 0.0628 \\ 0.0021 \\ 0.0370 \\ -0.0211 \\ 0.2106 \\ 0.3060 \\ 0.0371 \\ 0.3809 \\ -0.0055 \\ -0.5671 \end{bmatrix}.$$

Vrednost koncentracije kroma v točki  $T_k$  pri običajnem krigiranju izračunamo po enačbi (4.28):

$$Z^*(3.3, 2.8) = (-0.0099 \cdot 33.00 + 0.0628 \cdot 35.32 + 0.0021 \cdot 32.60 \dots) = 36.11.$$

Vidimo, da smo prišli do enakega rezultata kot pri preprostem krigiranju, kar pa ni pravilo. Pri različnih točkah oziroma podatkih se rezultati preprostega in običajnega krigiranja lahko razlikujejo.

#### 4.4.3 Krigiranje s trendom

Ponovimo izračun še za krigiranje s trendom, pri čemer se odločimo za linearni trend. Sistem enačb se še nekoliko razširi:<sup>3</sup>

$$\begin{bmatrix} 77.3 & 22.7 & 75.2 & 7.0 & 71.7 & 68.2 & 32.8 & 9.5 & 18.3 & 52.1 & 1.0 & 3.3 & 3.0 \\ 22.7 & 77.3 & 23.0 & 6.4 & 21.6 & 24.2 & 13.4 & 4.8 & 22.4 & 22.4 & 1.0 & 3.1 & 2.9 \\ 75.2 & 23.0 & 77.3 & 6.9 & 72.1 & 67.2 & 31.9 & 9.2 & 18.1 & 53.4 & 1.0 & 3.3 & 3.0 \\ 7.0 & 6.4 & 6.9 & 77.3 & 6.6 & 7.9 & 13.3 & 22.4 & 22.3 & 4.8 & 1.0 & 3.3 & 2.5 \\ 71.7 & 21.6 & 72.1 & 6.6 & 77.3 & 63.2 & 31.6 & 9.1 & 16.9 & 54.7 & 1.0 & 3.3 & 3.0 \\ 68.2 & 24.2 & 67.2 & 7.9 & 63.2 & 77.3 & 34.9 & 10.2 & 20.7 & 47.0 & 1.0 & 3.3 & 3.0 \\ 32.8 & 13.4 & 31.9 & 13.3 & 31.6 & 34.9 & 77.3 & 22.3 & 22.4 & 22.4 & 1.0 & 3.4 & 2.9 \\ 9.5 & 4.8 & 9.2 & 22.4 & 9.1 & 10.2 & 22.3 & 77.3 & 13.4 & 6.4 & 1.0 & 3.6 & 2.6 \\ 18.3 & 22.4 & 18.1 & 22.3 & 16.9 & 20.7 & 22.4 & 13.4 & 77.3 & 13.4 & 1.0 & 3.2 & 2.7 \\ 52.1 & 22.4 & 53.4 & 4.8 & 54.7 & 47.0 & 22.4 & 6.4 & 13.4 & 77.3 & 1.0 & 3.3 & 3.1 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 3.3 & 3.1 & 3.3 & 3.3 & 3.3 & 3.3 & 3.4 & 3.6 & 3.2 & 3.3 & 0.0 & 0.0 & 0.0 \\ 3.0 & 2.9 & 3.0 & 2.5 & 3.0 & 3.0 & 2.9 & 2.6 & 2.7 & 3.1 & 0.0 & 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 30.4 \\ 21.7 \\ 29.9 \\ 17.5 \\ 28.3 \\ 34.4 \\ 40.1 \\ 17.2 \\ 42.2 \\ 21.1 \\ 1.0 \\ 3.3 \\ 2.8 \end{bmatrix},$$

<sup>3</sup>Elementi matrike so zaradi omejenega prostora zapisane le na eno decimalno natančno, zaradi česar se izgubijo prave vrednosti koordinat točk. V izračunu smo upoštevali natančnejše vrednosti.

rešitev tega sistema pa je:

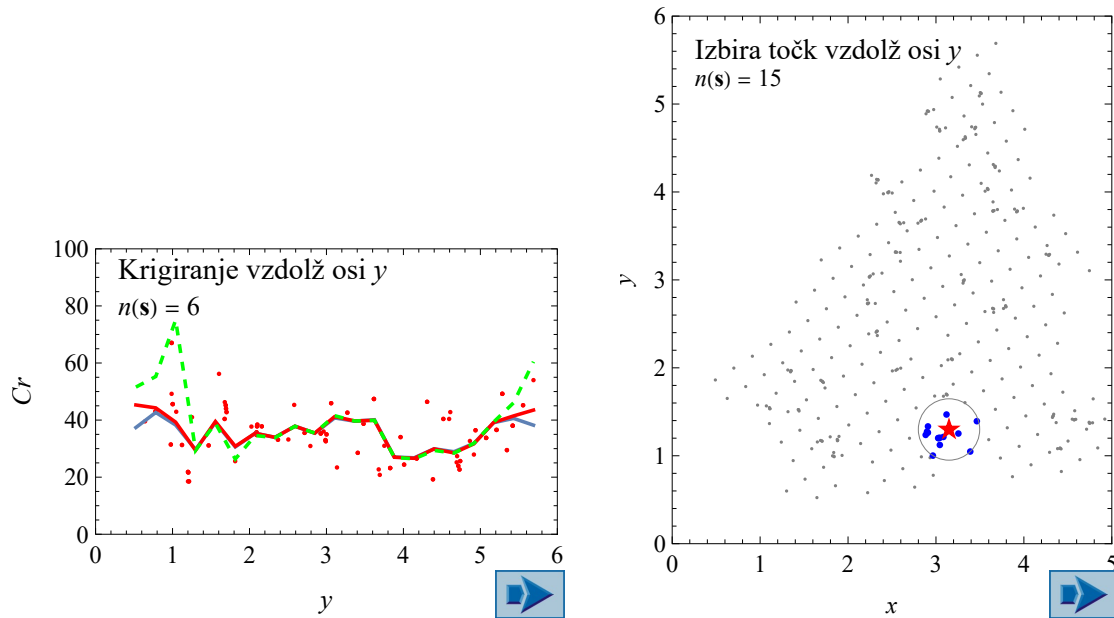
$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} -0.0105 \\ 0.0665 \\ 0.0032 \\ 0.0414 \\ -0.0233 \\ 0.2108 \\ 0.3030 \\ 0.0335 \\ 0.3843 \\ -0.0088 \\ -9.4340 \\ 1.6570 \\ 1.2100 \end{bmatrix}.$$

Vrednost koncentracije kroma v točki  $T_k$  pri krigiranju s trendom izračunamo po enačbi (4.39):

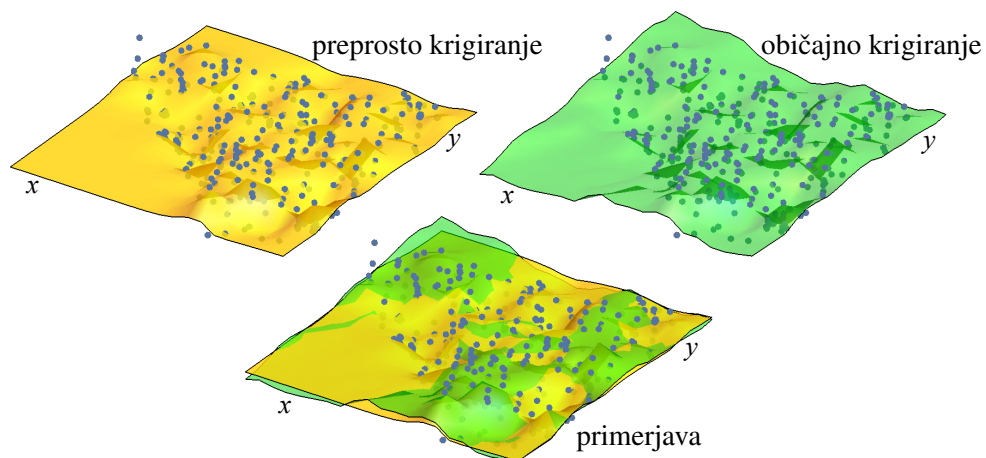
$$Z^*(3.3, 2.8) = (-0.0105 \cdot 33.00 + 0.0665 \cdot 35.32 + 0.0032 \cdot 32.60 \cdot \dots) = 36.09,$$

ki je le malo drugačna od vrednosti, ki smo jo pridobili s preprostim oziroma običajnim krigiranjem.

Pri modeliranju neznanih vrednosti slučajnega polja najpogosteje določamo aproksimacijo vzdolže neke osi ali po nekem območju. Na sliki 4.4 prikazujemo rezultate krigiranja vzdolž osi  $y$  pri  $x = 3.15$  km. Uporabili smo vse tri vrste krigiranja: preprosto (modra), običajno (rdeča) in krigiranje z linearnim trendom (črtkana zelena). Poleg tega smo krigirali z različnim številom okoliških točk, od  $n(s) = 6$  do  $n(s) = 42$ . Opazimo lahko, da so razlike med preprostim in običajnim krigiranjem zelo majhne, pa še to le pri manjšem številu okoliških točk. Krigiranje s trendom se nekoliko bolj razlikuje še posebej za manjše število okoliških točk in na obeh robovih območja, kjer je lahko vpliv trenda večji.

SLIKA 4.4: Krigiranje vzdolž osi  $y$  pri  $x = 3.15$  km

Na sliki 4.5 prikazujemo rezultate krigiranja za celotno območje Jura pri preprostem in običajnem krigiranju za  $n(s) = 10$ . Vidimo, da krigiranje sledi spreminjanju vrednosti slučajnega polja pri obeh načinih krigiranja. Na spodnji sliki vidimo, da se rezultati preprostega in običajnega krigiranja na nekaterih mestih razlikujejo.



SLIKA 4.5: Rezultati krigiranja koncentracije kroma na celotnem območju Jura



## 5 Prostorsko utežena regresija

S pojmom regresija opisujemo postopek za določitev ocen parametrov določenih matematičnih modelov. Če pojav  $Z$  opisujemo z linearnim modelom  $f$ , ki ga v  $k$ -dimenzionalnem prostoru zapišemo z naslednjo enačbo:

$$z^*(x_1, x_2, \dots, x_k) = f(x_1, x_2, \dots, x_k) = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_k x_k, \quad (5.1)$$

potem z regresijsko analizo določimo ocene parametrov  $\hat{a}_i, i = 0, 1, 2 \dots k$ . Taki regresiji pravimo **linearna regresija**. Podatke, ki jih potrebujemo za regresijo, lahko zapišemo v naslednji preglednici in imajo enako strukturo kot geostatistični podatki (glej preglednico 3.1).

PREGLEDNICA 5.1: Struktura podatkov za regresijo

$X_1$	$X_2$	$\dots$	$X_k$	$Z$
$x_{11}$	$x_{12}$	$\dots$	$x_{1k}$	$z_1$
$x_{21}$	$x_{22}$	$\dots$	$x_{2k}$	$z_2$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_{n1}$	$x_{n2}$	$\dots$	$x_{nk}$	$z_n$

### 5.1 Linearna regresija

Opišimo najprej preprost primer linearne regresije za eno samo neodvisno spremenljivko  $x$ . V tem primeru se enačba (5.1) poenostavi:

$$z^*(x) = f(x) = a_0 + a_1 x. \quad (5.2)$$

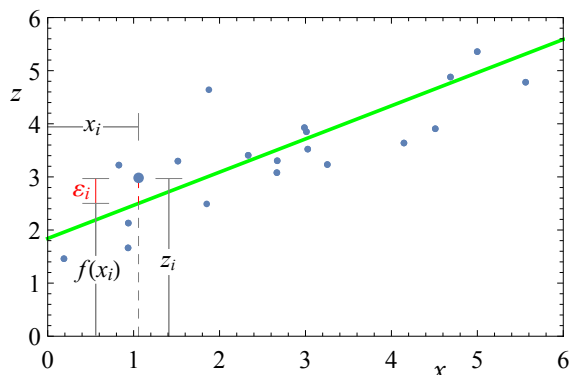
Na sliki 5.1 je prikazan primer linearne regresije za eno neodvisno spremenljivko. Označena sta tudi ena od točk  $(x_i, y_i)$  in **odstopanje** oziroma **rezidual**  $\varepsilon_i$ . Ordinata  $z_i$  je vsota vrednosti funkcije  $f(x_i)$  in reziduala  $\varepsilon_i$ :

$$z_i = f(x_i) + \varepsilon_i \quad \longrightarrow \quad \varepsilon_i = z_i - f(x_i) = z_i - (a_0 + a_1 x_i). \quad (5.3)$$

Za določitev ocen  $\hat{a}_0$  in  $\hat{a}_1$  parametrov  $a_0$  in  $a_1$  postavimo pogoj, da bi bila odstopanja  $\varepsilon_i$  čim manjša. Ob predpostavki, da se odstopanja porazdeljujejo normalno s srednjo vrednostjo nič, je najprimernejša

**metoda najmanjših kvadratov**, pri kateri postavimo pogoj, da je vsota kvadratov odstopanj čim manjša:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (z_i - f(x_i))^2 = \sum_{i=1}^n (z_i - \hat{a}_0 - \hat{a}_1 x_i)^2 \quad \text{naj bo čim manjša.} \quad (5.4)$$



SLIKA 5.1: Linearna regresija za eno neodvisno spremenljivko

V primeru, da se odstopanja ne porazdeljujejo normalno oziroma vključujejo osamelce (angl. outlier), pa je bolje uporabiti eno izmed različic **robustne regresijske metode**:

absolutne vrednosti:

$$S = \sum_{i=1}^n |\varepsilon_i| \quad \text{naj bo čim manjša,} \quad (5.5)$$

Huberjeva<sup>1</sup> metoda:

$$S = \sum_{i=1}^n \begin{cases} 1/2 \varepsilon_i^2 & |\varepsilon_i| \leq h \\ h|\varepsilon_i| - 1/2 h^2 & |\varepsilon_i| > h \end{cases} \quad \text{naj bo čim manjša,} \quad (5.6)$$

bikvadratna enačba:

$$S = \sum_{i=1}^n \begin{cases} h^2/6 (1 - (1 - (\varepsilon_i/h)^2)^3) & |\varepsilon_i| \leq h \\ h^2/6 & |\varepsilon_i| > h \end{cases} \quad \text{naj bo čim manjša,} \quad (5.7)$$

kjer s  $h$  označimo velikost območja, na katerem je funkcija  $S$  nelinearno odvisna od  $\varepsilon_i$ .

V nadaljevanju se bomo omejili le na metodo najmanjših kvadratov. Iz sistema linearnih enačb, ki jih izpeljemo iz pogoja (5.4):

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n z_i \\ \sum_{i=1}^n z_i x_i \end{bmatrix}, \quad (5.8)$$

<sup>1</sup> Peter Jost Huber, švicarski statistik, utemeljitelj robustne statistike (1934–).

določimo vrednosti ocen  $\hat{a}_0$  in  $\hat{a}_1$ :

$$\hat{a}_0 = \bar{Z} - \frac{S_{XZ}}{S_X^2} \bar{X}, \quad \hat{a}_1 = \frac{S_{XZ}}{S_X^2}, \quad (5.9)$$

kjer sta  $\bar{X}$  in  $\bar{Z}$  povprečji vzorcev  $x_i$  in  $z_i$ ,  $S_{XZ}$  je vzorčna kovarianca med  $X$  in  $Z$ ,  $S_X^2$  pa je vzorčna varianca  $X$ .

V primeru, da izvajamo linearno regresijo za  $k$  spremenljivk, iz sistema normalnih enačb

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i2} x_{i1} & \cdots & \sum x_{ik} x_{i1} \\ \sum x_{i2} & \sum x_{i1} x_{i2} & \sum x_{i2}^2 & \cdots & \sum x_{ik} x_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{i1} x_{ik} & \sum x_{i2} x_{ik} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_k \end{bmatrix} = \begin{bmatrix} \sum z_i \\ \sum z_i x_{i1} \\ \sum z_i x_{i2} \\ \vdots \\ \sum z_i x_{ik} \end{bmatrix} \quad (5.10)$$

izračunamo neznane ocene parametrov  $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k$ , ki so enake na celotnem obravnavanem območju.

Linearno regresijo več spremenljivk lahko zapišemo tudi v matričnem zapisu. Definiramo naslednje matrike: matriko podatkov  $\mathbf{X}$ , matriko neznanih koeficientov  $\mathbf{a}$ , matriko vrednosti odvisnih spremenljivk  $\mathbf{Z}$  in matriko odstopanj  $\mathbf{e}$ . Matrike zapišemo z naslednjimi izrazi:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_n \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Linearni model zapišemo z enačbo:

$$\mathbf{Z} = \mathbf{X}\mathbf{a} + \mathbf{e}, \quad (5.11)$$

iz katere lahko izpeljemo enačbo odstopanj:

$$\mathbf{e} = \mathbf{Z} - \mathbf{X}\mathbf{a}.$$

Po metodi najmanjših kvadratov zapišemo najprej vsoto kvadratov odstopanj  $S$ :

$$S = \mathbf{e}^T \mathbf{e} = (\mathbf{Z} - \mathbf{X}\mathbf{a})^T (\mathbf{Z} - \mathbf{X}\mathbf{a}). \quad (5.12)$$

Minimum vsote kvadratov odstopanj dobimo z odvajanjem  $S$  po neznanih koeficientih  $\mathbf{a}$  in upoštevaje, da je ta odvod enak nič:

$$\begin{aligned} \frac{dS}{d\mathbf{a}} = \mathbf{0} &\rightarrow 2\mathbf{X}^T (\mathbf{Z} - \mathbf{X}\hat{\mathbf{a}}) = 2(\mathbf{X}^T \mathbf{Z} - \mathbf{X}^T \mathbf{X}\hat{\mathbf{a}}) = \mathbf{0}, \\ &\rightarrow \mathbf{X}^T \mathbf{X}\hat{\mathbf{a}} = \mathbf{X}^T \mathbf{Z}, \end{aligned} \quad (5.13)$$

kar predstavlja sistem normalnih enačb v matričnem zapisu.

## 5.2 Utežena linearna regresija

Zaradi prostorske odvisnosti, nestacionarnosti in drugih vplivov parametri  $a_0, a_1, \dots$  niso enaki na celotnem obravnavanem območju. To je osnovna ideja utežene linearne regresije, pri kateri za vsako točko, v kateri želimo določiti regresijo, izračunamo drugačne ocene parametrov. Razlika izhaja iz spremenjene enačbe metode najmanjših kvadratov (5.4), ki ji dodamo uteži  $w_i$ :

$$S = \sum_{i=1}^n w_i \varepsilon_i^2 \quad \text{naj bo čim manjša.} \quad (5.14)$$

Ob upoštevanju, da uteži  $w_i$  uredimo v matriko  $\mathbf{W}$ :

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}, \quad (5.15)$$

vsoto kvadratov odstopanj zapišemo takole:

$$S = \mathbf{e}^T \mathbf{W} \mathbf{e} = (\mathbf{Z} - \mathbf{X}\mathbf{a})^T \mathbf{W} (\mathbf{Z} - \mathbf{X}\mathbf{a}). \quad (5.16)$$

Minimum vsote kvadratov odstopanj dobimo z odvajanjem  $S$  po neznanih koeficientih  $\mathbf{a}$  in zahtevo, da je ta odvod enak nič:

$$\begin{aligned} \frac{dS}{d\mathbf{a}} = \mathbf{0} &\rightarrow 2 \mathbf{X}^T \mathbf{W} (\mathbf{Z} - \mathbf{X}\hat{\mathbf{a}}) = 2 (\mathbf{X}^T \mathbf{W} \mathbf{Z} - \mathbf{X}^T \mathbf{W} \mathbf{X}\hat{\mathbf{a}}) = \mathbf{0}, \\ &\rightarrow \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\mathbf{a}} = \mathbf{X}^T \mathbf{W} \mathbf{Z}, \end{aligned} \quad (5.17)$$

kar predstavlja sistem normalnih enačb v matričnem zapisu ob upoštevanju različno uteženih odstopanj.

### 5.2.1 Določitev uteži

Poseben izziv pri uteženi linearni regresiji je določitev uteži  $w_i$ . Intuitivno bomo predpostavili, da naj vrednosti v točkah, ki so bližje točki  $T$  s koordinatami  $(x_{T1}, x_{T2}, \dots, x_{Tk})$ , bolj vplivajo na izračunano vrednost  $z^*$  kot vrednosti v bolj oddaljenih točkah. Uteži  $w_i$  določimo s padajočo funkcijo razdalje  $d_i$  med točkami  $T_i$ , kjer poznamo vrednosti  $Z_i$ , in točko  $T$ , za katero vrednost določimo z modelom (5.1).

Razdaljo  $d_i$  običajno izračunamo kot Evklidovo razdaljo med točkama  $T$  in  $T_i$ :

$$d_i = \sqrt{(x_{T1} - x_{i1})^2 + (x_{T2} - x_{i2})^2 + \cdots + (x_{Tk} - x_{ik})^2} = \sqrt{(\mathbf{X}_T - \mathbf{X}_i)^T (\mathbf{X}_T - \mathbf{X}_i)}. \quad (5.18)$$

Vrednosti uteži lahko določimo z eno od naslednjih enačb:

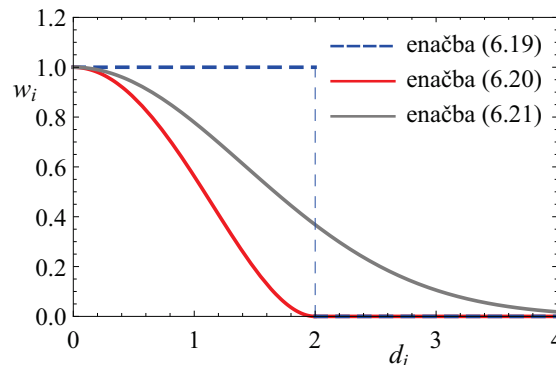
$$w(d) = \begin{cases} 1 & d \leq h \\ 0 & d > h \end{cases}, \quad (5.19)$$



$$w(d) = \begin{cases} \left(1 - \left(\frac{d}{h}\right)^2\right)^2 & d \leq h \\ 0 & d > h \end{cases}, \quad (5.20)$$

$$w(d) = e^{-\left(\frac{d}{h}\right)^2}. \quad (5.21)$$

Enačba (5.19) predstavlja binarno odvisnost, kjer je vrednost uteži enaka 1, če je razdalja manjša od kritične razdalje oziroma praga  $h$ . Pri enačbi (5.20) vrednost uteži  $w$  zvezno pada proti nič, ki jo doseže pri kritični razdalji  $d = h$ . Pri enačbi (5.21) vrednost uteži  $w$  zvezno pada proti nič, a je nikoli ne doseže. Kritična razdalja  $h$  pri tej enačbi predstavlja parameter, ki pove, kako hitro pada vrednost uteži.



SLIKA 5.2: Vrednosti uteži v odvisnosti od razdalje med točkami

### 5.2.2 Izbira funkcije za določitev uteži in kritične razdalje

V prejšnjem razdelku smo pojasnili, kako določimo uteži v odvisnosti od razdalje točke, kjer poznamo vrednost slučajnega polja, in točke, za katero želimo to vrednost oceniti. Prikazali smo tri različne funkcije (5.19)–(5.21), ki so odvisne tudi od parametra  $h$ . Katera funkcija in kolikšen parameter  $h$  je optimalna rešitev, moramo še določiti.

Ena od možnosti je izračun vrednosti navzkrižne validacije (angl. crossvalidation score), pri kateri opravimo prostorsko uteženo regresijo za različne funkcije in različne vrednosti  $h$  v vseh točkah, kjer poznamo vrednost slučajnega polja. Pri vsakokratni določitvi utežene regresije s seznama točk s poznanimi vrednostmi slučajnega polja izključimo prav točko, za katero določamo uteženo regresijo. V tem primeru torej določamo regresijo z  $n - 1$  točk, kjer je  $n$  število točk, v katerih poznamo vrednosti slučajnega polja. Tako določene ocene vrednosti slučajnega polja primerjamo z izmerjenimi oziroma tvorimo naslednjo cenilko:

$$CV(h) = \sum_{i=1}^n (z_i - z_i^*(h))^2. \quad (5.22)$$

Najnižja vrednost  $CV(h)$  ustreza optimalni funkciji in vrednosti  $h$ . Izračun vrednosti  $CV(h)$  za različne enačbe bomo prikazali v naslednjem razdelku.

### 5.2.3 Primeri

Vzemimo že znane podatke z območja Jura. Za vrednosti koncentracije kroma opravimo običajno linearno regresijo in uteženo regresijo za vse tri tipe utežnih funkcij (enačbe (5.19)–(5.21)) pri določenih kritičnih razdaljah  $h$ . Model linearne regresije za celotno območje je:

$$Z(x, y) = 37.0827 - 0.842091x + 0.163621y. \quad (5.23)$$

V preglednici 5.2 prikazujemo izračun razdalj med desetimi najbližjimi točkami, kjer poznamo vrednosti, in točko  $T$ , ki je podana s koordinatama  $x_T = 2$  km in  $y_T = 2$  km. V isti preglednici so tudi uteži, izračunane po enačbah (5.19)–(5.21), s kritičnimi razdaljami  $h = 0.3$  km za prvi dve enačbi in  $h = 0.1$  km za zadnjo.

PREGLEDNICA 5.2: Izračun uteži za nekaj točk za točko  $T = (2, 2)$  km

$x$ [km]	$y$ [km]	$Z$ /	$d$ [km]	$w$ (5.19) /	$w$ (5.20) /	$w$ (5.21) /
1.948	1.906	49.60	0.107	1	0.7600	0.3154
2.159	2.041	37.72	0.164	1	0.4906	0.06746
1.813	2.116	33.48	0.220	1	0.2134	0.007887
2.024	2.251	38.00	0.252	1	0.0862	0.001733
2.083	1.695	29.28	0.316	0	0	$4.579 \cdot 10^{-5}$
1.720	1.833	31.68	0.326	0	0	$2.421 \cdot 10^{-5}$
1.709	1.843	33.32	0.331	0	0	$1.786 \cdot 10^{-5}$
1.706	1.848	34.84	0.331	0	0	$1.749 \cdot 10^{-5}$
2.294	1.830	46.80	0.340	0	0	$9.795 \cdot 10^{-6}$
1.690	1.858	38.72	0.341	0	0	$8.927 \cdot 10^{-6}$

Ob upoštevanju vseh 359 točk, kjer so podane vrednosti slučajnega polja, lahko po enačbi (5.17) ocenimo parametra  $a_0$  in  $a_1$  ter izračunamo vrednost polja v točki  $T$ . V preglednici 5.3 prikazujemo rezultate običajne in utežene regresije v točki  $T = (2, 2)$  km.

PREGLEDNICA 5.3: Vrednost  $Z^*$  v točki  $T = (2, 2)$  km z različnimi linearnimi regresijami

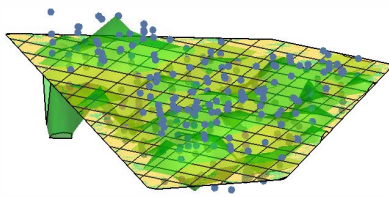
Običajna regresija	Utežena regresija (5.19)	Utežena regresija (5.20)	Utežena regresija (5.21)
35.73	42.49	42.80	42.52

Vidimo, da je rezultat utežene regresije v vseh treh primerih višji od vrednosti, ki smo jo dobili z običajno regresijo. To je rezultat dejstva, da so vrednosti v okolici točke  $T = (2, 2)$  km izmerjene vrednosti

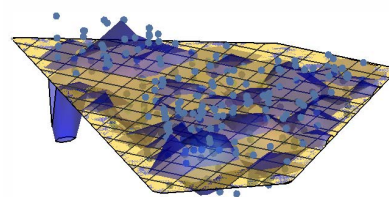
slučajnega polja višje od vrednosti, ki jo dobimo s preprosto linearno regresijo, kar je razvidno tudi iz preglednice 5.2.

Na sliki 5.3 prikazujemo tri regresijske ploskve, ki smo jih določili z uteženo linearno regresijo (5.17) s tremi različnimi enačbami za uteži, z rumeno ploskvijo pa je prikazan rezultat običajne linearne regresije. Na spodnji desni sliki na istem grafu primerjamo vse tri različice utežene linearne regresije in običajno linearno regresijo. Opazimo lahko, da različni načini določitve uteži pripeljejo do različnih ploskev, ki bolj ali manj intenzivno sledijo podatkom, ki odstopajo od linearne ploskve, ki jo dobimo z običajno linearno regresijo.

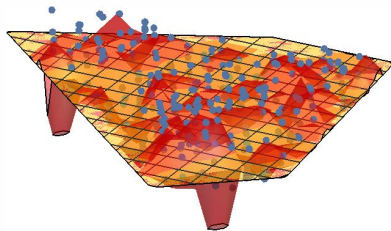
Uteži po enačbi (5.19)  $h = 0.3$



Uteži po enačbi (5.20)  $h = 0.3$



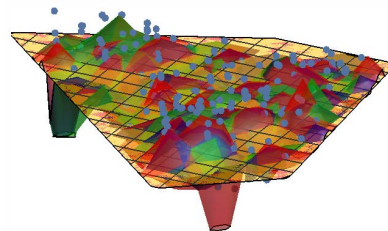
Uteži po enačbi (5.21)  $h = 0.1$



Uteži po enačbi (5.19)  $h = 0.3$

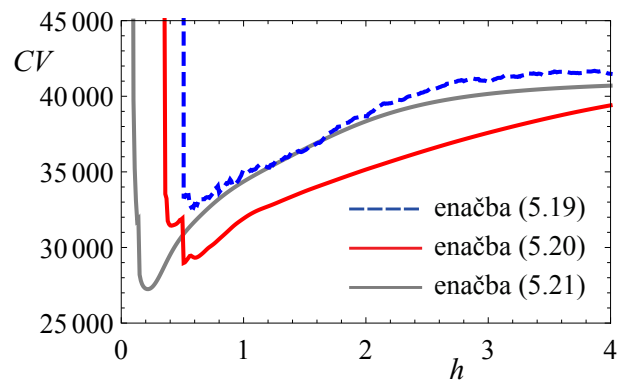
Uteži po enačbi (5.20)  $h = 0.3$

Uteži po enačbi (5.21)  $h = 0.1$



SLIKA 5.3: Prostorsko utežena regresija z različnimi funkcijami  $w(d)$

S ciljem, da bi ocenili, kateri način prostorsko utežene regresije je najboljši, smo opravili postopek navzkrižne validacije. Za vseh 359 točk, v katerih poznamo vrednosti slučajnega polja (koncentracije kroma) smo izračunali prostorsko uteženo regresijo, pri čemer pa nismo upoštevali vrednosti v točki, za katero smo opravili regresijo. Za vsako od 359 točk smo torej uporabili podatke na preostalih 358 točkah. Tako pridobljene ocene vrednosti slučajnega polja smo primerjali z izmerjenimi in po enačbi (5.22) izračunali vrednost navzkrižne validacije  $CV$ . Rezultate za vse tri predlagane enačbe ((5.19)–(5.21)) in različne vrednosti  $h$  prikazujemo na sliki 5.4. Vidimo lahko, da je dala enačba (5.21) najboljše rezultate, saj so tu razlike med izmerjenimi in napovedanimi vrednostmi slučajnega polja najmanjše. S slike tudi opazimo, da je za enačbi (5.19) in (5.20) optimalna razdalja  $h$  približno 0.3 km, medtem ko je za enačbo (5.21) optimalna razdalja  $h$  približno 0.1 km. Ti vrednosti smo uporabili tudi za prikaz prostorske regresije na sliki 5.3.

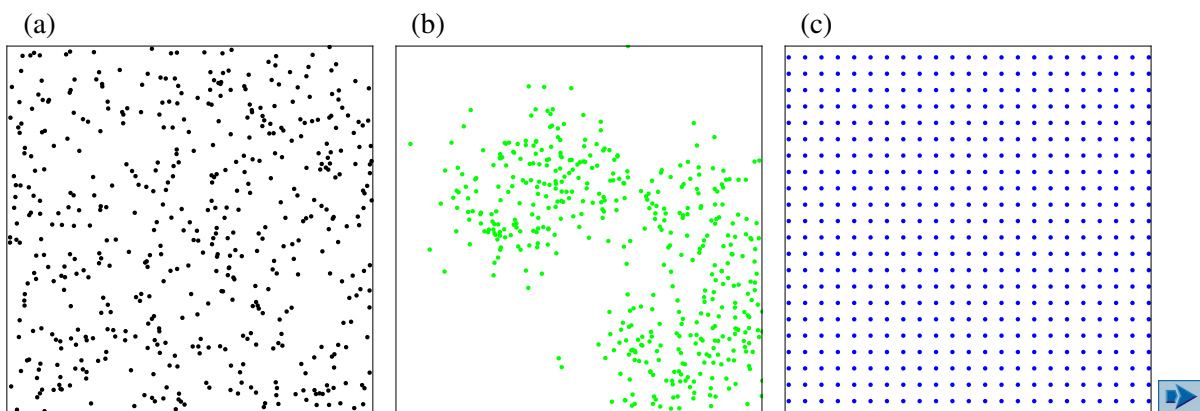


SLIKA 5.4: Določitev optimalne vrednosti  $h$  z uporabo navzkrižne validacije

## 6 Prostorski vzorci

Pri analizi prostorskih podatkov so lege točk, kjer merimo vrednosti slučajne funkcije oziroma slučajnega polja, zelo pomemben podatek. Razpored točk je lahko rezultat načrta poskusa ali pa je slučajen. Zaradi različnih dejavnikov so lahko točke enakomerno razporejene po prostoru ali pa so v gručah oziroma klastrih. Pri razporedu točk po prostoru lahko govorimo o treh značilnih primerih:

- slučajno enakomerno porazdeljene točke v prostoru,
- slučajno porazdeljene točke v gručah (v nadaljevanju v gručah),
- deterministično enakomerno razporejene točke v prostoru.



SLIKA 6.1: Trije tipi razporeditve točk: (a) slučajno enakomerno, (b) v gručah in (c) enakomerno

Osnovna naloga tega področja prostorske statistike je, da znamo ločiti med temi tremi primeri oziroma da lahko s statističnimi preizkusi in izbranim tveganjem  $\alpha$  zavržemo ali ne zavržemo določene domneve o porazdeljenosti točk v slučajnem polju in zapišemo odločitve. Pri preizkušanju domnev bomo za ničelno domnevo vedno izbrali izjavo, da so točke porazdeljene slučajno enakomerno. Oblika alternativne domneve pa je odvisna od tega, ali opravljamo enostranski ali dvostranski preizkus. Poznamo torej tri možnosti ničelne in alternativne domneve:

$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke niso razporejene slučajno enakomerno.

ali

$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke so razporejene v gručah.

ali

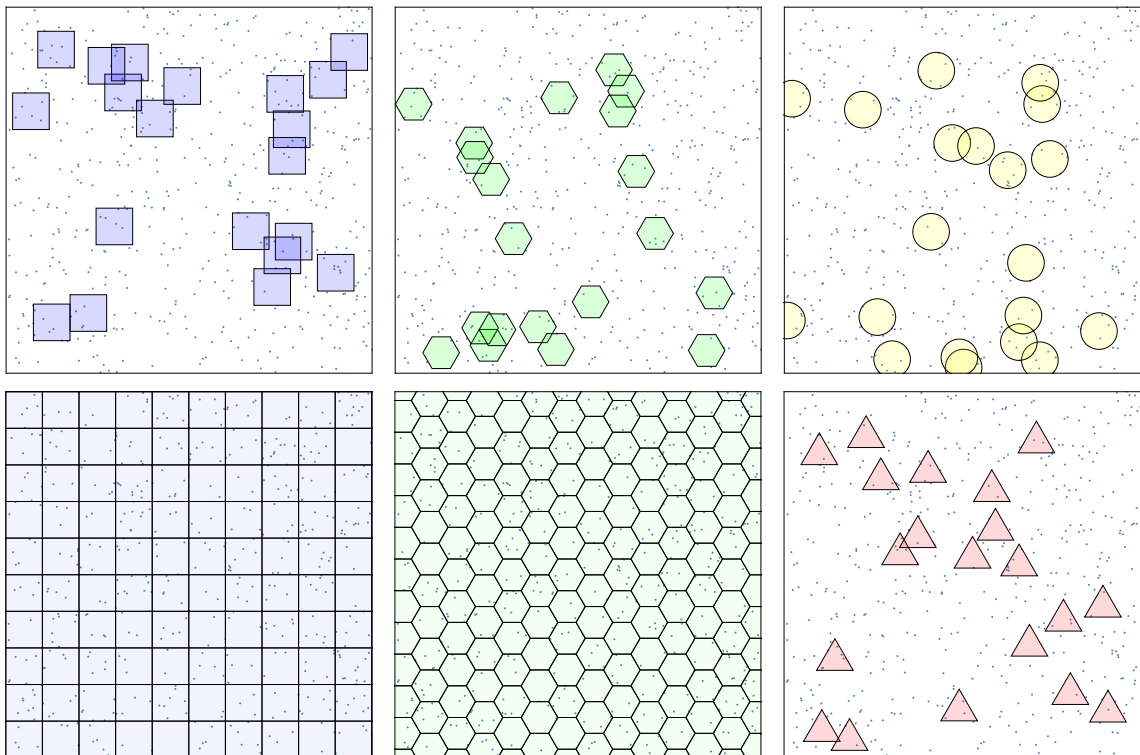
$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke so razporejene deterministično enakomerno.

Analizo prostorskih vzorcev lahko naredimo z **metodo kvadratov** ali pa z **metodo razdalj** oziroma **metodo najbližjega sosedu**.

## 6.1 Metoda kvadratov

Pri metodi kvadratov na območje analize postavimo kvadrate ali druge like ter preštejemo, koliko točk je znotraj posameznih likov. Na sliki 6.2 prikazujemo različne možnosti postavitve likov različnih oblik. Lahko se odločimo za povsem slučajno postavitev likov ali pa like postavimo tako, da z njimi prekrijemo celotno območje. Ne glede na obliko lika, bomo v okviru razlag te metode vse like imenovali „kvadrati“.



SLIKA 6.2: Različne oblike likov, ki jih lahko uporabimo pri metodi kvadratov

Število točk v posameznem kvadratu označimo z  $Y_i, i = 1, \dots, n$ , kjer je  $n$  število kvadratov, ki jih

uporabimo v analizi. Ob predpostavki, da so točke razporejene slučajno enakomerno, je verjetnost, da je točka znotraj kvadrata, enaka razmerju ploščin kvadrata  $A_k$  in celotnega območja  $A$ :

$$p = \frac{A_k}{A}, \quad (6.1)$$

število točk znotraj kvadrata pa je slučajna spremenljivka, porazdeljena po Poissonovi porazdelitvi. Verjetnostna funkcija Poissonove porazdelitve je:

$$p_Y(y) = \frac{\nu^y e^{-\nu}}{y!}, \quad y = 0, 1, 2, \dots, \quad (6.2)$$

kjer parameter  $\nu$  predstavlja pričakovano število točk v posameznem kvadratu in je enako  $\nu = pN$ . Z  $N$  smo označili število vseh točk na obravnavanem območju. Zanimiva lastnost Poissonove porazdelitve je, da sta srednja vrednost in varianca enaki parametru porazdelitve  $\nu$ :

$$E[Y] = \text{var}[Y] = \nu. \quad (6.3)$$

Na osnovi te lastnosti lahko pripravimo eno od možnih statistik za preizkus domneve o prostorski urejenosti:

$$H^* = \frac{S_Y^{*2}}{\bar{Y}} \sim \chi_{n-1}^{*2}, \quad (6.4)$$

kjer smo s  $S_Y^{*2}$  označili varianco vzorca, z  $\bar{Y}$  pa njegovo povprečje. Statistika  $H^*$  se porazdeljuje po normirani porazdelitvi  $\chi^{*2}$ , ki se od običajne porazdelitve  $\chi^2$  loči po tem, da ima srednjo vrednost enako ena, oziroma da je  $H^*$  le  $(n-1)$ -krat manjša od  $H$ :

$$H^* = \frac{H}{n-1} \quad \longrightarrow \quad F_{\chi^{*2}}(h) = F_{\chi^2}((n-1)h). \quad (6.5)$$

V zadnji enačbi smo upoštevali enačbo za določitev porazdelitve funkcije slučajne spremenljivke za primer monotono naraščajoče funkcije, kar zveza v (6.5) za  $n > 1$  seveda je. Enačbo za določitev porazdelitvene funkcije slučajne spremenljivke  $Y$ , ki je monotono naraščajoča funkcija  $X$ , namreč zapišemo takole:

$$Y = g(X) \quad \longrightarrow \quad F_Y(y) = F_X(g^{-1}(y)). \quad (6.6)$$

Poleg statistike  $H^*$ , ki jo določimo z enačbo (6.4), so različni avtorji uporabili tudi druge statistike. Te so na primer:

$$\begin{aligned} ICS &= \frac{S_Y^{*2}}{\bar{Y}} - 1 && \text{(David, Moore, 1954),} \\ \dot{Y} &= \bar{Y} + \frac{S_Y^{*2}}{\bar{Y}} - 1 && \text{(Lloyd, 1967),} \\ IP &= \frac{\dot{Y}}{\bar{Y}} && \text{(Lloyd, 1967),} \\ I_\delta &= \frac{\sum_{i=1}^n Y_i(Y_i - 1)}{\bar{Y}(n\bar{Y} - 1)} && \text{(Morisita, 1959).} \end{aligned} \quad (6.7)$$

Porazdelitve vseh teh statistik niso splošno znane. Eden od načinov za uporabo le-teh je uporaba simulacij. Porazdelitev statistike  $ICS$  je sicer podobna porazdelitvi  $\chi^{*2}$ , le da je zamaknjena za ena proti levi. Za porazdelitev statistike  $\dot{Y}$  lahko za nekoliko večje število kvadratov, na primer, če je število kvadratov večje od 80, uporabimo normalno porazdelitev s srednjo vrednostjo  $\nu$  in standardnim odklonom  $\sqrt{\nu/n + 2/(n-1)}$ .

Pri metodi kvadratov lahko uporabimo eno od statistik v enačbah (6.4) ali (6.7), lahko pa uporabimo preizkus domneve o porazdelitvi. Če vemo, da se število točk v posameznem kvadratu ob predpostavki, da so točke razporejene slučajno enakomerno, porazdeljuje po Poissonovi porazdelitvi, lahko ničelno domnevo preizkusimo s testom Poissonove porazdelitve. To lahko naredimo s preizkusom  $\chi^2$ , s preizkusom Kolmogorova in Smirnova ali kakšno drugo metodo za preizkus domnev o porazdelitvi.

Pri preizkusu  $\chi^2$  preštejemo, koliko kvadratov vsebuje določeno število točk. Glede na pričakovano število točk znotraj kvadrata in število kvadratov se odločimo, na koliko razredov razdelimo podatke in kaj predstavljajo ti razredi. Včasih je primerno, da prvi razred predstavlja kvadrate, ki nimajo nobene točke, drugi kvadrate z eno točko in tako naprej. Zadnji  $K$ -ti razred pa predstavlja vse tiste kvadrate, ki vsebujejo  $K-1$  ali več točk. Če je pričakovano število točk v kvadratu večje, je smiselno, da prvi razred predstavlja kvadrate, ki vključujejo manj kot določeno število točk. V preglednici 6.1 prikazujemo dva značilna primera razporeditev razredov.

PREGLEDNICA 6.1: Dva primera razporeditve razredov glede na število točk v kvadratu

Razred	Število točk	Razred	Število točk
1	0	1	0–3
2	1	2	4
3	2	3	5
4	3	4	6
5	4	5	7
6	5	6	8–9
7	6 ali več	7	10 ali več

Za vsak razred izračunamo teoretična števila kvadratov, ki vsebujejo določeno število točk (teoretične velikosti razredov)  $n_i$ :

$$n_i = P[\text{število točk v kvadratu ustreza razredu } i] n, \quad (6.8)$$

kjer je  $n$  število vseh kvadratov. Empirične velikosti razredov  $\hat{n}_i$  določimo tako, da preštejemo kvadrate, ki ustrezajo razredu  $i$ . Nato lahko izračunamo statistiko  $H$  po naslednji enačbi:

$$H = \sum_{i=1}^K \frac{(n_i - \hat{n}_i)^2}{n_i} \sim \chi_{K-1}^2, \quad (6.9)$$



kjer smo označili, da se statistika  $H$  porazdeljuje po porazdelitvi  $\chi^2$  s  $K - 1$  prostostnimi stopnjami. Pri tem preizkusu ne moremo ločiti med razporeditvijo v gruče in urejeno deterministično razporeditvijo. Ugotovimo lahko le, ali razporeditev točk statistično značilno odstopa od slučajne enakomerne razporeditve. Kritično območje je  $H > \chi_{K-1,1-\alpha}^2$ . Če je vrednost statistike v kritičnem območju, moramo ničelno domnevo zavrniti.

## 6.2 Primeri – metoda kvadratov

Analizirajmo tri različne primere razporeditve točk: slučajno enakomerno (primer A), v gručah (primer B) in deterministično enakomerno (primer C). Postavimo ničelno in alternativno domnevo ter izberimo tveganje  $\alpha$ , pri katerem bi še zavrnili ničelno domnevo:

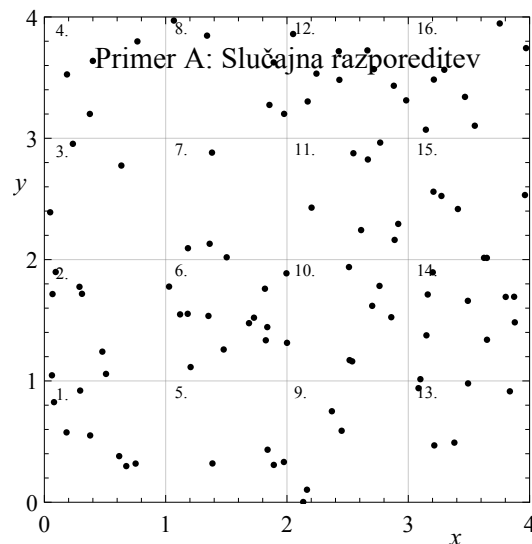
$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke niso razporejene slučajno enakomerno.

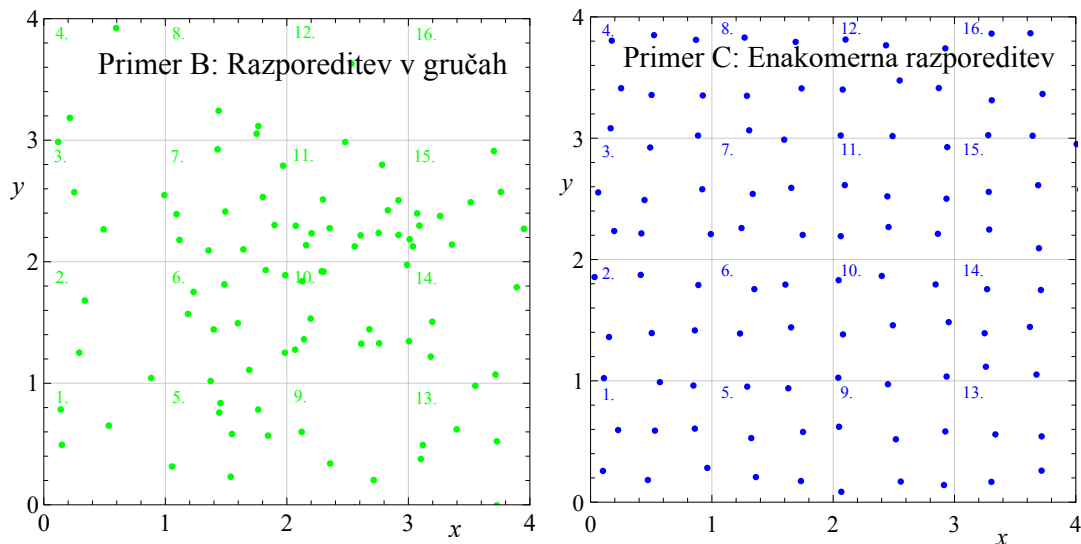
---


$$\alpha = 0.05$$

Razporeditev točk prikazujemo na slikah 6.3 in 6.4. Kvadratno območje razdelimo na 16 kvadratov in preštujemo število točk v posameznih kvadratih. Skupno število točk  $N$  je enako 100 za vse tri primere. Na slikah so označeni tudi kvadrati s številkami od 1 do 16.



SLIKA 6.3: Enakomerna slučajna razporeditev točk – primer A



SLIKA 6.4: Različne razporeditve točk – primera B in C

V preglednici 6.2 je zapisano, koliko točk je v posameznem kvadratu. Iz teh podatkov lahko izračunamo povprečje in varianco ter določimo statistiko  $H^*$  po enačbi (6.4).

PREGLEDNICA 6.2: Število točk v posameznih kvadratih

Kvadrat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Primer A	7	7	3	4	4	12	4	5	4	7	7	9	5	9	6	7
Primer B	3	3	4	2	7	10	9	3	3	10	13	1	5	5	10	0
Primer C	8	7	7	8	6	4	5	5	7	8	7	8	4	6	4	6

V preglednici 6.3 prikazujemo vrednosti statistike z mejami kritičnega območja. V zadnjem stolpcu so vrednosti dejanskega tveganja ob zavrnitvi ničelne domneve. V primeru, da je statistika manjša od spodnje meje  $H^* < H_1$  ali večja od zgornje meje  $H^* > H_2$ , ničelno domnevo zavrnemo. V teh primerih je dejansko tveganje  $\alpha_{dej} < \alpha = 0.05$ .

PREGLEDNICA 6.3: Izračun statistik in kritičnih območij

	$\bar{Y}$	$S_Y^{*2}$	$H^*$	$H_1$	$H_2$	$\alpha_{dej}$
Primer A	6.25	6.42	1.01	0.417	1.83	0.874
Primer B	5.50	14.8	2.69	0.417	1.83	0.000799
Primer C	6.25	0.77	0.35	0.417	1.83	0.021

V naših primerih lahko ugotovimo naslednje: v primeru A statistika ni v kritičnem območju, dejansko tveganje je večje od 5 %, saj znaša kar 87.4 %, ničelne domneve ne moremo zavrniti in s tveganjem  $\alpha$  ne moremo trditi, da točke niso porazdeljene slučajno enakomerno. V primerih B in C je enkrat statistika večja od zgornje meje, drugič pa manjša od spodnje. V obeh primerih je dejansko tveganje manjše od 5 %, ničelno domnevo moramo zavrniti in s tveganjem  $\alpha$ , manjšim od 5 %, lahko trdimo, da točke niso porazdeljene slučajno enakomerno.

Analizirajmo še porazdelitev točk, v katerih so bile opravljene meritve v območju Jura. Naredili bomo enostranski preizkus, zato je alternativna domneva drugačna, kot v prejšnjem primeru:

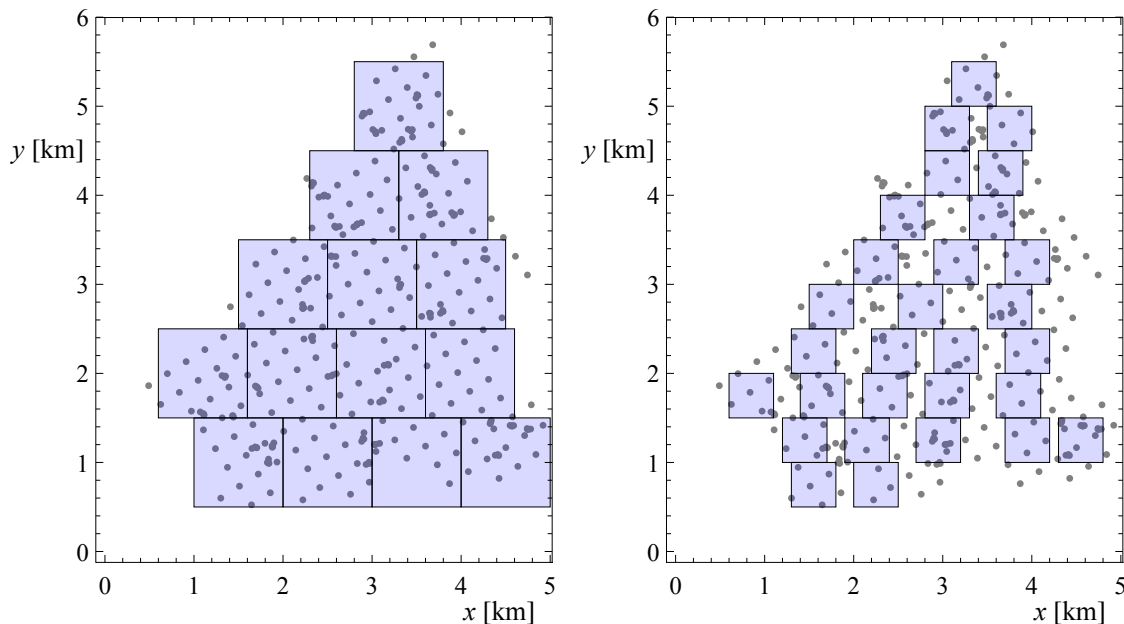
$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke so zbrane v gručah.

---


$$\alpha = 0.05$$

Nalogo ponovimo z dvema različnima postavitvama kvadratov. V prvem primeru uporabimo kvadrate velikosti  $1 \text{ km} \times 1 \text{ km}$  tako, da je bilo pokrito območje točk, v drugem primeru pa uporabimo manjše kvadrate velikosti  $0.5 \text{ km} \times 0.5 \text{ km}$ , razporejeni pa so tako, kot kaže slika 6.5 (desno).



SLIKA 6.5: Dva načina postavitve kvadratov

V preglednici 6.4 prikazujemo izračun statistike, mej kritičnega območja in dejanskega tveganja. Ugotovimo, da postavitve kvadratov odločilno vpliva na končen rezultat in odločitev o tipu razporeditve točk. Vidimo, da je ob prvi postavitvi kvadratov tveganje za zavrnitev ničelne domneve zelo velika. Zato ne moremo trditi, da razporeditev točk ni slučajna enakomerna. Šele ob uporabi druge postavitve manjših

kvadratov metoda prepozna, da so točke združene v gruče. V tem primeru je statistika v kritičnem območju, saj je  $H^* = 2.10 > H_2 = 1.48$ . Dejansko tveganje je seveda manjše od dovoljenega  $\alpha_{dej} = 0.0006 \ll \alpha = 0.05$ . Zato lahko s tveganjem, ki je precej manjše od 5 %, trdimo, da se točke združujejo v gruče.

PREGLEDNICA 6.4: Izračun statistik in kritičnih območij

	$\bar{Y}$	$S_Y^{*2}$	$H^*$	$H_2$	$\alpha_{dej}$
Postavitev A	24.7	25.9	1.05	1.72	0.400
Postavitev B	7.21	15.1	2.10	1.48	0.0006

Pri razporedu A število kvadratov ne zadošča za uporabo preizkusa  $\chi^2$ , zato ga bomo opravili le za razpored B. Odločimo se, da bodo v prvem razredu kvadrati, ki vsebujejo od nič do štiri točke, v drugem kvadrati s pet točkami, v tretjem kvadrati s šest točkami in tako naprej do šestega razreda s kvadrati, ki vsebujejo devet točk. V sedmem, zadnjem razredu so kvadrati, ki imajo več kot devet točk. V preglednici 6.5 prikazujemo razporeditev v razrede, teoretične velikosti razredov  $n_i$  in empirične velikosti razredov  $\hat{n}_i$ .

PREGLEDNICA 6.5: Izračun statistik in kritičnih območij

razred $i$	razporeditev	$n_i$	$\hat{n}_i$
1	0–4	4.493	16
2	5	3.483	4
3	6	4.184	4
4	7	4.308	0
5	8	3.881	2
6	9	3.108	0
7	10 ali več	5.542	3

Postavimo ničelno in alternativno domnevo ter izberimo tveganje  $\alpha$ , pri katerem bi še zavrnilo ničelno domnevo:

$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke niso razporejene slučajno enakomerno.

---


$$\alpha = 0.05$$

Statistiko  $H$  izračunamo po enačbi (6.9) in dobimo:

$$H = \frac{(4.493 - 16)^2}{4.493} + \frac{(3.483 - 4)^2}{3.483} + \dots + \frac{(5.542 - 3)^2}{5.542} = 39.05. \quad (6.10)$$

Meja kritičnega območja je  $\chi_{6,0.95}^2 = 12.59$ . Ker je vrednost statistike bistveno višja od meje kritičnega območja  $H = 39.05 > 12.59$ , dejansko tveganje pa je zanemarljivo  $\alpha_{dej} = 7.01 \cdot 10^{-7}$ , moramo ničelno domnevo zavrniti in s tveganjem, manjšim od 5 %, trdimo, da točke niso razporejene slučajno enakomerno.

### 6.3 Pomanjkljivosti metode kvadratov

Zadnji primer, ko smo zaradi različne razporeditve in velikosti kvadratov prišli do povsem nasprotujočih si rezultatov, kaže na pomanjkljivosti metode kvadratov. Te lahko opredelimo v naslednjih točkah:

- Rezultat je lahko odvisen od velikosti in razporeditve kvadratov. Zato je pri uporabi metode kvadratov primerno, da se preizkus ponovi pri različnem številu, velikosti in razporeditvi kvadratov.
- Ker statistiko izračunamo iz števila točk v kvadratih, dejansko primerjamo le povprečne gostote točk v kvadratih, ne pa razporeditev ene točke glede na druge.
- Pri analizi izgubimo informacijo o razporeditvi točk znotraj kvadrata. Zato gruče, ki so manjše od kvadratov, ta metoda ne zazna.

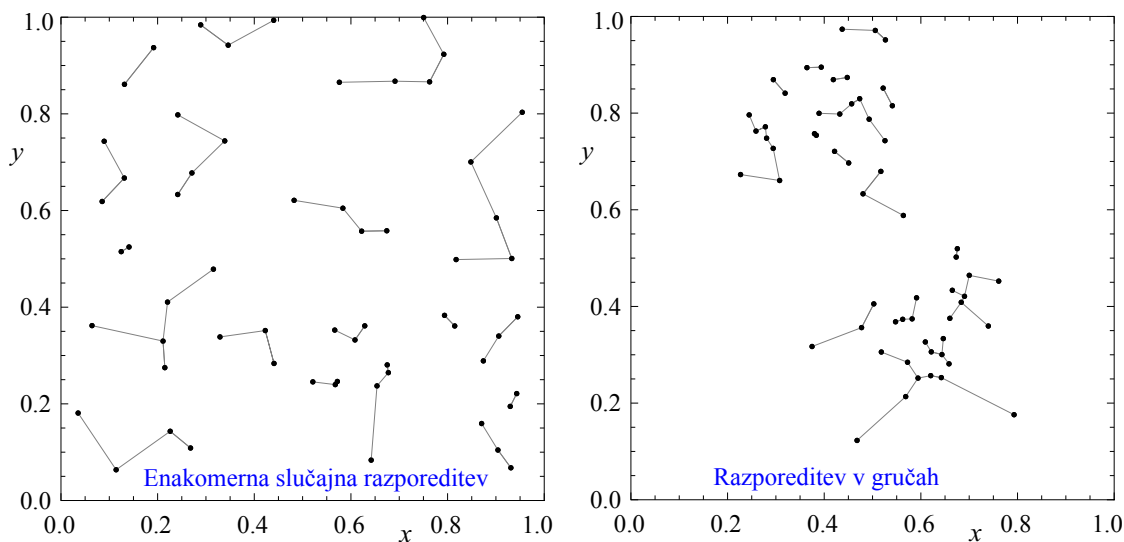
Da bi premostili pomanjkljivosti metode kvadratov, lahko uporabimo alternativno metodo **najbližjega soseda**.

### 6.4 Metoda najbližjega soseda

Pri metodi najbližjega soseda za vsako točko izračunamo razdaljo  $d$  do najbližje sosednje točke:

$$d_i = \min_{j=1}^n d_{ij}, \quad i \neq j. \quad (6.11)$$

Na sliki 6.6 so prikazane točke z enakomerno razporeditvijo in razporeditvijo v gruče. Prikazane so tudi razdalje od posamezne točke do najbližje sosede. Že površna primerjava odkrije, da so razdalje do najbližjih sosedov večinoma krajše v primeru, ko so točke razporejene v gručah.



SLIKA 6.6: Najbližji sosedi za dve različni razporeditvi točk

V primeru, da je razporeditev točk slučajna enakomerna, porazdelitev razdalje do najbližjega soseda lahko zapišemo z naslednjo gostoto verjetnosti:

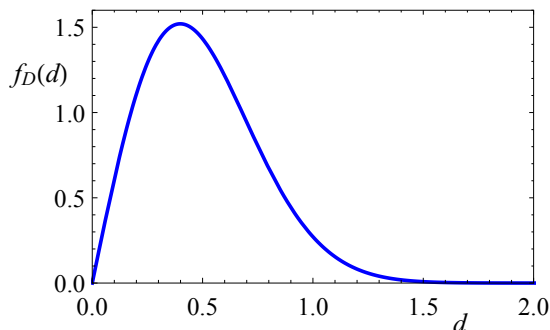
$$f_D(d) = 2\pi \lambda d e^{-\pi \lambda d^2}, \quad \text{za } d > 0, \quad (6.12)$$

kjer

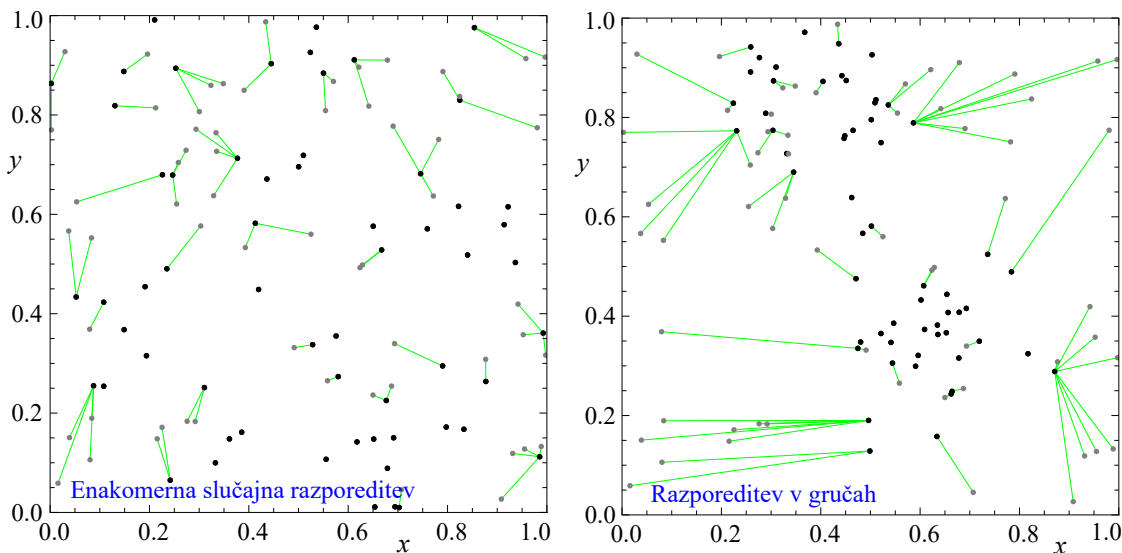
$$\lambda = \frac{n}{A} \quad (6.13)$$

predstavlja gostoto točk ( $n$  je število točk,  $A$  pa ploščina analiziranega območja). Graf te gostote verjetnosti prikazujemo na sliki 6.7. Zapišimo še srednjo vrednost in varianco slučajne spremenljivke  $D$ :

$$\begin{aligned} m_D = E[D] &= \frac{1}{2} \sqrt{\frac{A}{n}} = \frac{1}{2} \sqrt{\frac{1}{\lambda}}, \\ \sigma_D^2 = var[D] &= \frac{A}{n^2} \left( \frac{1}{\pi} - \frac{1}{4} \right) = \frac{1}{n\lambda} \left( \frac{1}{\pi} - \frac{1}{4} \right). \end{aligned} \quad (6.14)$$

SLIKA 6.7: Gostota verjetnosti razdalje do najbližjega soseda za  $\lambda = 1$ 

Pri metodi najbližjega soseda lahko izberemo tudi alternativno pot, da namesto iskanja najbližjega soseda vsake točke določamo razdaljo od navidezne točke do njej najbližje dejanske točke. Pri tem predpostavimo, da je navidezna točka kjerkoli na območju, ki ga analiziramo. To pomeni, da je porazdelitev lege navidezne točke slučajna enakomerna. Na sliki 6.8 prikazujemo najkrajše razdalje od navideznih točk (obarvanih sivo) do dejanskih točk za dve razporeditvi točk. Tudi tu lahko že na prvi pogled opazimo razlike med razdaljami do točk, če so dejanske točke razporejene enakomerno ali v gručah.



SLIKA 6.8: Najkrajša razdalja od navideznih do analiziranih točk za dve razporeditvi dejanskih točk

Porazdelitev razdalje  $X$  od navidezne do najbližje dejanske točke je enaka porazdelitvi razdalje  $D$  do najbližjega soseda za primer slučajne enakomerne razporeditve, saj navidezna točka, od katere merimo razdaljo, predstavlja le še eno točko med slučajno enakomerno razporejenimi točkami. Edina, skoraj zanemarljiva razlika je v izračunu gostote točk, ki je v tem primeru enaka

$$\lambda = \frac{n+1}{A}. \quad (6.15)$$

Možnosti za določitev ustrezne statistike, na osnovi katere bomo lahko izpeljali statistični preizkus, je več. Temeljijo tako na slučajni spremenljivki  $D$ , to je razdalji med najbližjimi sosedami, kot tudi na slučajni spremenljivki  $X$ , torej razdalji med navidezno in njej najbližjo dejansko točko. Zapišimo nekaj različnih statistik z ustreznimi porazdelitvami:

$$U = \frac{2\sqrt{\lambda}}{n} \sum_{i=1}^n d_i \sim N\left(1, \frac{4-\pi}{n\pi}\right) \quad (\text{Clark, Evens, 1954}), \quad (6.16)$$

$$H = 2\pi\lambda \sum_{i=1}^n d_i^2 \sim \chi_{2n}^2 \quad (\text{Skellam, 1952}), \quad (6.17)$$

$$U = \frac{\pi\lambda}{n} \sum_{i=1}^n x_i^2 \sim N\left(1, \frac{1}{n}\right) \quad (\text{Pielou, 1959}), \quad (6.18)$$

$$S = \frac{n \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i\right)^2} \sim \text{porazdelitve ne poznamo, uporabiti moramo simulacije (Eberhardt, 1967)}, \quad (6.19)$$

$$H = \frac{12n}{7n+1} \left( n \ln \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \sum_{i=1}^n \ln(x_i^2) \right) \sim \chi_{n-1}^2 \quad (\text{Pollard, 1971}), \quad (6.20)$$

$$U = \frac{1}{n} \sum_{i=1}^n \frac{x_i^2}{x_i^2 + d_i^2} \sim N\left(\frac{1}{2}, \frac{1}{12n}\right) \quad (\text{Byth, Ripley, 1980}), \quad (6.21)$$

$$F = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n d_i^2} \sim F_{2n,2n} \quad (\text{Hopkins, 1954}). \quad (6.22)$$

Nekateri avtorji pa so pri definiciji statistike uporabili še slučajno spremenljivko  $X_2$ , ki predstavlja razdaljo med navidezno in drugo najbližjo dejansko točko.

$$U = \frac{1}{n} \sum_{i=1}^n \frac{x_i^2}{x_{2i}^2} \sim N\left(\frac{1}{2}, \frac{1}{12n}\right) \quad (\text{Holgate, 1965}), \quad (6.23)$$

$$B = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_{2i}^2} \sim \beta_{n,n} \quad (\text{Holgate, 1965}), \quad (6.24)$$

kjer  $x_i$  predstavlja razdaljo med navidezno točko in njej najbližjo točko,  $x_{2i}$  pa razdaljo med navidezno točko in njej drugo najbližjo točko.



## 6.5 Primeri – metoda najbližjega soseda

Analizirajmo enake primere kot pri metodi kvadratov v podpoglavju [Primeri - metoda kvadratov](#). Obravnavamo tri različne primere razporeditve točk na kvadratnem območju velikosti  $4 \times 4$ : slučajno enakomerno (primer A), v gručah (primer B) in deterministično enakomerno (primer C). Postavimo ničelno in alternativno domnevo ter izberimo tveganje  $\alpha$ , pri katerem bi še zavrnilo ničelno domnevo:

$H_0$ : Točke so razporejene slučajno enakomerno.

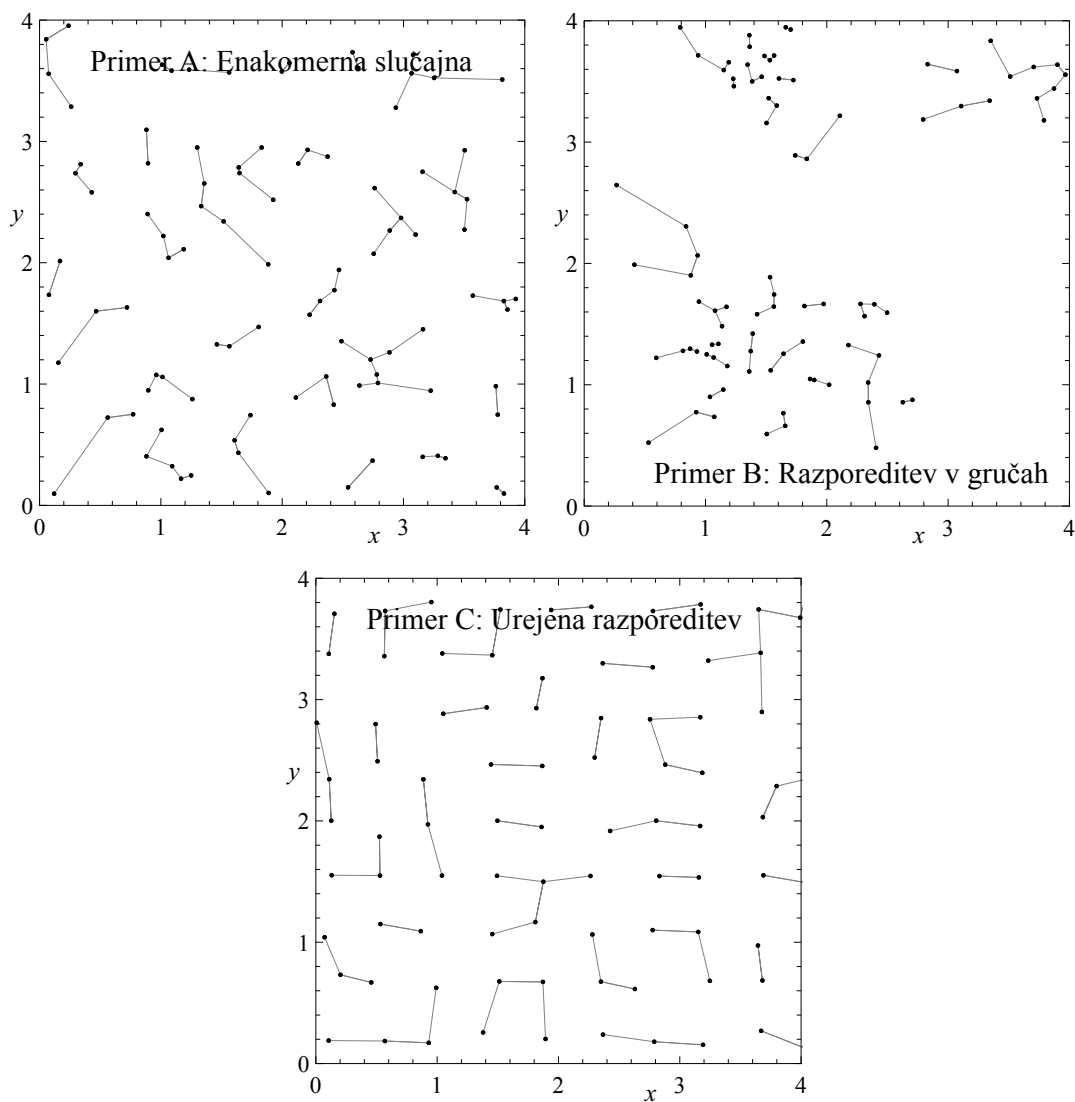
$H_1$ : Točke niso razporejene slučajno enakomerno.

---

$$\alpha = 0.05$$

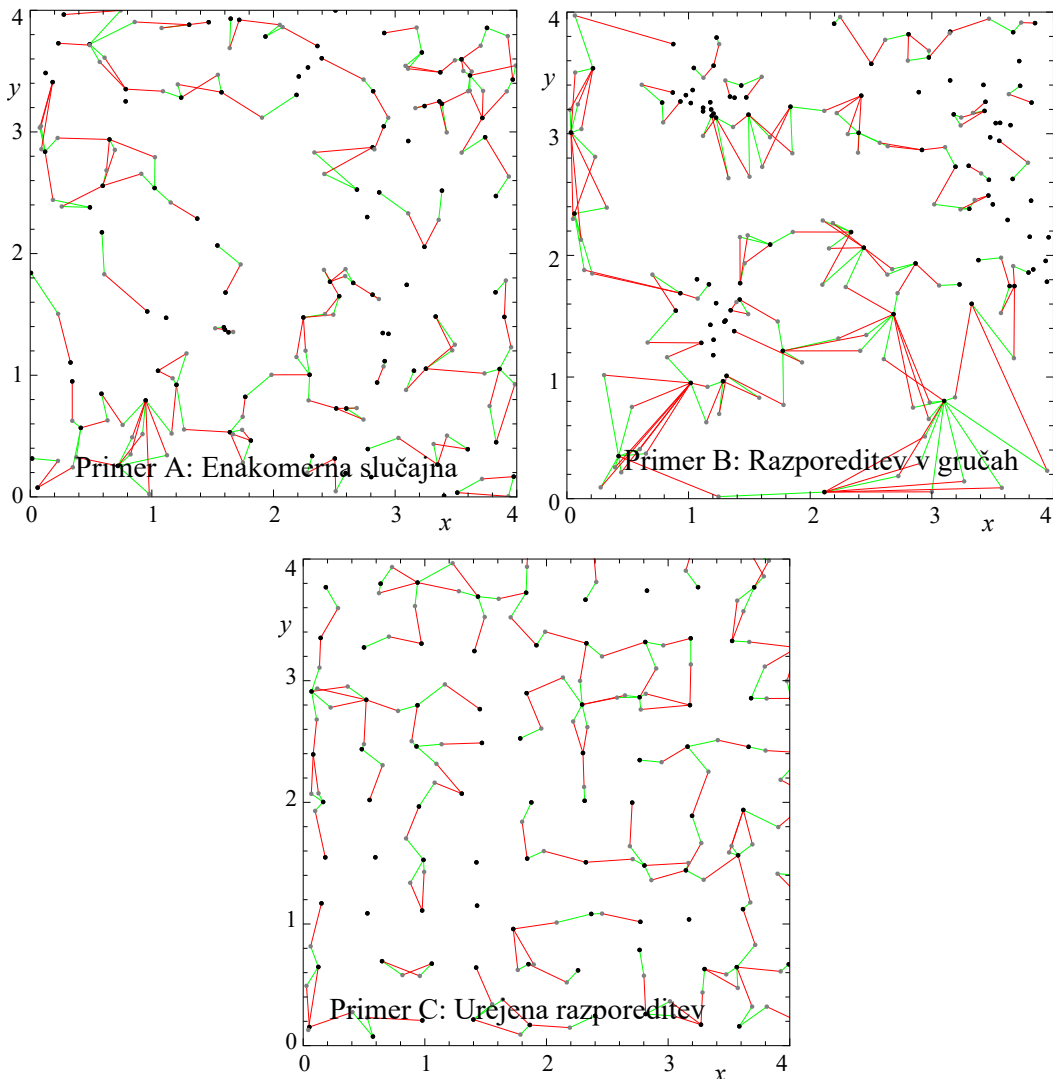
Pripravili smo vzorce s stotimi točkami za različne razporeditve dejanskih točk. Nato smo pripravili še vzorec stotih navideznih točk, ki so razporejene slučajno enakomerno. Iz teh vzorcev smo lahko določili razdalje  $d_i$ ,  $x_i$  in  $x_{2i}$ , iz katerih lahko izračunamo statistike po enačbah (6.16)–(6.24).

Različne razporeditve stotih točk grafično prikazujemo na sliki 6.9, kjer so označene najkrajše razdalje med dejanskimi točkami – vzorec  $d_i$ .



SLIKA 6.9: Najkrajše razdalje med točkami za tri razporeditve dejanskih točk

Na sliki 6.10 prikazujemo podobno razporejene točke kot na sliki 6.9, poleg njih pa še enako veliko skupino slučajno enakomerno razporejenih navideznih točk, ki so označene s svetlejšo sivo barvo. Tu označujemo še razdalje od navideznih točk do najbližje dejanske točke z zelenimi črtami – vzorec  $x_i$  in razdalje od navideznih do druge najbližje dejanske točke z rdečimi črtami – vzorec  $x_{2i}$ .



SLIKA 6.10: Najkrajše razdalje med navideznimi in dejanskimi točkami za tri razporeditve točk

V preglednici 6.6 prikazujemo izračun statistik, mej kritičnega območja in dejanskega tveganja za tri različne tipe razporeditve točk in za osem različnih načinov določitve statistike. Vidimo lahko, da je v primeru A, ko so dejanske točke porazdeljene slučajno enakomerno, statistika vedno v srednjem območju, ki nam ne narekuje zavrnitve ničelne domneve. Zato v tem primeru ničelne domneve ne zavrnemo in zaključimo, da s tveganjem  $\alpha = 5\%$  ne moremo trditi, da točke niso razporejene slučajno ena-

komerno. V tem primeru je za vse različne statistike dejansko tveganje večje od izbranega. V primeru B vidimo, da je statistika v kritičnem območju vedno, razen pri uporabi statistike (6.21). V primerih, ko je statistika v kritičnem območju in je s tem tudi tveganje manjše od izbranega tveganja 5 %, ničelno domnevo zavrnilo in zaključimo, da s tveganjem, manjšim od 5 %, lahko trdimo, da točke niso razporejene slučajno enakomerno. V primeru C, ko so točke razporejene deterministično enakomerno, vsi preizkusi kažejo na to, da moramo ničelno domnevo zavrniti, saj je statistika v kritičnem območju, dejansko tveganje pa nižje od izbranega. Tudi v tem primeru pridemo do enakega zaključka, da s tveganjem, manjšim od 5 %, lahko trdimo, da točke niso razporejene slučajno enakomerno.

PREGLEDNICA 6.6: Izračun statistik, kritičnih območij in dejanskega tveganja za tri razporeditve točk

Enačba	spodnja meja	zgornja meja	Primer A		Primer B		Primer C	
			statistika	tveganje	statistika	tveganje	statistika	tveganje
(6.16)	0.898	1.102	1.022	0.668	0.789	$5.61 \cdot 10^{-5}$	1.827	0.
(6.17)	162.7	241.1	222.3	0.268	160.2	0.0351	535.	0.
(6.18)	0.804	1.196	0.959	0.688	4.093	0.	0.639	0.000303
(6.20)	73.36	128.4	88.98	0.490	238.5	$3.23 \cdot 10^{-13}$	63.77	0.00460
(6.21)	0.443	0.557	0.493	0.818	0.541	0.156	0.404	0.000904
(6.22)	0.431	0.569	0.456	0.215	0.684	$7.78 \cdot 10^{-8}$	0.331	$8.15 \cdot 10^{-7}$
(6.23)	0.443	0.557	0.499	0.976	0.651	$1.79 \cdot 10^{-7}$	0.190	0.
(6.24)	0.757	1.320	0.863	0.301	5.111	0.	0.239	0.

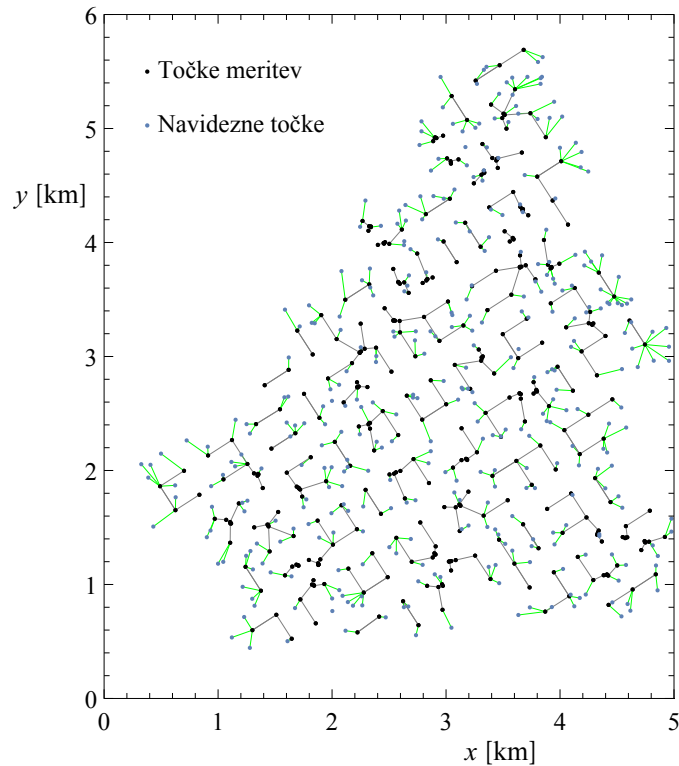
Naredimo še preizkus domneve o razporeditvi točk, v kateri so bile opravljene meritve na območju Jura. Naredimo enak preizkus kot za prejšnji primer, ko smo obravnavali porazdelitev točk v kvadratu  $4 \times 4$ . Število točk je enako 359, prav toliko točk tudi generiramo, in sicer tako, da so razporejene slučajno enakomerno. Gostota točk je  $\lambda = 23.3$  točk/km<sup>2</sup>. Razporeditev točk, kjer so bile opravljene meritve, in enakomerno porazdeljenih navideznih točk ter razdalje do najbližjih točk so prikazane na sliki 6.11. Na tej sliki so s sivo označene razdalje med najbližjimi sosednjimi točkami, kjer so bile opravljene meritve, z zeleno pa razdalje od navideznih do najbližjih točk meritev.

$H_0$ : Točke so razporejene slučajno enakomerno.

$H_1$ : Točke niso razporejene slučajno enakomerno.

---


$$\alpha = 0.05$$



SLIKA 6.11: Razporeditev točk meritev in navideznih točk za primer Jura

V preglednici 6.7 prikazujemo izračune za preizkušanje domneve o razporeditvi merskih točk za primer Jura. Ugotovimo lahko, da metode različnih avtorjev pripeljejo do različnih zaključkov. V primeru uporabe enačb (6.16), (6.17) ter (6.20)–(6.22) je tveganje manjše od izbrane vrednosti  $\alpha = 5\%$ , zato lahko ničelno domnevo zavrnilo in izjavimo, da s tveganjem, manjšim od  $5\%$ , lahko trdimo, da točke niso razporejene slučajno enakomerno. Če pa računamo po enačbi (6.18), ničelne domneve ne moremo zavrnilo. V tem primeru izjavimo, da ob tveganju  $5\%$  ne moremo trditi, da točke niso razporejene slučajno enakomerno. Poudariti moramo, da so statistike, pri katerih uporabimo navidezne točke, odvisne od slučajno generiranega vzorca navideznih točk. Zato ponovitve teh izračunov lahko vodijo do različnih rezultatov in zato tudi do različnih zaključkov.

PREGLEDNICA 6.7: Preizkušanje domneve o razporeditvi točk v primeru Jura

Enačba	spodnja meja	zgornja meja	statistika	tveganje
(6.16)	0.946	1.054	1.209	$3.62 \cdot 10^{-14}$
(6.17)	645.6	794.1	1358.	0.
(6.18)	0.897	1.103	1.084	0.114
(6.20)	307.5	412.3	302.6	0.0305
(6.21)	0.470	0.530	0.550	0.00115
(6.22)	0.864	1.158	0.551	$2.32 \cdot 10^{-15}$

## 7 Metoda Monte Carlo

Integriranje je zelo pomembna in uporabna operacija na mnogih področjih tehnike, tudi v prostorski statistiki. Ker sta v mnogih primerih tako integrant kot integracijsko območje taka, da le za posamezne preproste primere ta integral lahko izračunamo analitično, moramo uporabiti numerične metode. Metoda Monte Carlo je ena od metod numeričnega integriranja, ki temelji na generiranju vzorca slučajnih spremenljivk in vektorjev.

Osnovni postopek metode Monte Carlo je preprost:

1. Definiramo problem, definiramo osnovne lastnosti slučajnih spremenljivk, ki jih uporabimo pri reševanju problema. Določimo število simulacij (število ponovitev).
2. Začetek zanke simulacij (ponavljanje poskusov).
  - 2.1 Generiramo vzorec slučajnih spremenljivk.
  - 2.2 Opravimo analizo, s katero rešimo osnovni problem v okviru ene simulacije.
3. Konec zanke simulacij.
4. Opravimo končno analizo ter numerični in grafični prikaz rezultatov.

Vzorec slučajne spremenljivke tu razumemo kot vrednosti, ki jih slučajna spremenljivka zavzame ob ponavljanju poskusa. Vzemimo, da želimo s to metodo določiti verjetnost, da je rezultat meta poštene kocke manjši ali enak dve. Rezultat seveda poznamo in je enak  $1/3$ :

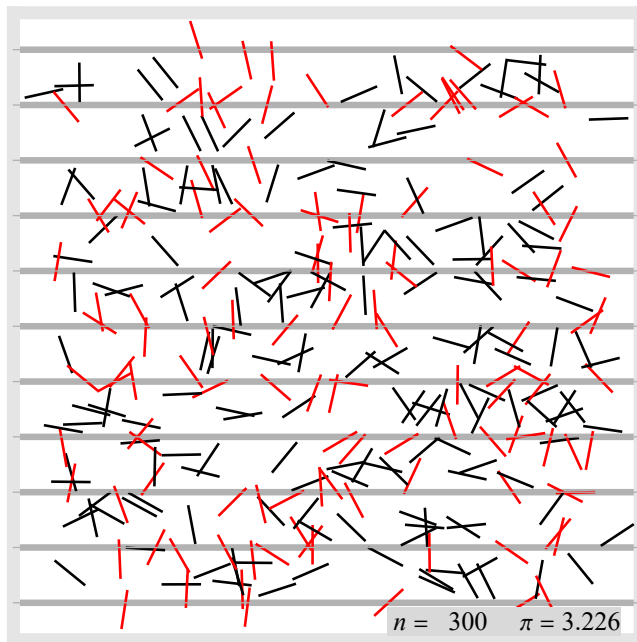
1. Pri definiciji problema zapišemo, kaj želimo določiti in da je slučajna spremenljivka diskretna in enakomerno porazdeljena od 1 do vključno 6. Določimo število simulacij, na primer 10000. Postavimo števec ugodnih izidov na nič.
2. Začetek zanke simulacij (ponavljanje poskusov).
  - 2.1 Generiramo številko od ena do šest po diskretni enakomerni porazdelitvi.
  - 2.2 Ugotovimo, ali je generirana številka manjša ali enaka dve. Če je, povečamo števec ugodnih izidov za ena.
3. Konec zanke simulacij.
4. Oceno iskane verjetnosti dobimo tako, da število ugodnih izidov delimo s številom vseh simulacij.

Ime metode Monte Carlo se je prvič pojavilo med drugo svetovno vojno, ko so znanstveniki von Neuman, Ulam, Fermi in drugi v laboratorijih v Los Alamosu tudi s pomočjo simulacij skonstruirali prvo atomsko bombo. Sicer pa sta ideja o generiranju slučajnih spremenljivk in njihova uporaba pri reševanju problemov bistveno starejši. Prvi dobro dokumentiran primer je opravil Comte de Buffon leta 1777, ko je v problemu, poznanem kot Buffonova igla, s simulacijami poskušal določiti vrednost števila  $\pi$ . Pozneje so generiranje slučajnih števil uporabili tudi Lord Kelvin (1901), Gosset (Student) (1908), Fermi (1930) in drugi (Kalos et al., 1986, Rubinstein, 1981). Z razvojem boljših, hitrejših računalnikov postaja metoda Monte Carlo uporabna za vse širši krog ljudi. Trenutno, v jeseni 2023, lahko v sistemu Web of Science najdemo več kot 143000 člankov, ki vključujejo področje metode Monte Carlo.

Pri primeru, ki ga poznamo kot Buffonova igla, na površino, na kateri so narisane vzporedne črte z medsebojno razdaljo  $d$ , postopoma mečemo igle. Comte de Buffon je pokazal, da je verjetnost, da igla dolžine  $l$  seka eno izmed vzporednih črt, enaka:

$$P[\text{igla seka črto}] = \frac{2l}{\pi d} \quad \rightarrow \quad \pi = \frac{2l}{P[\text{igla seka črto}] d}.$$

Če torej s poskušanjem ocenimo verjetnost, da igla seka črto, lahko po zgornji enačbi ocenimo število  $\pi$ . Animacija takega postopka določitve števila  $\pi$  je prikazana na sliki 7.1.



SLIKA 7.1: Animacija določitve števila  $\pi$



## 7.1 Generiranje vzorca slučajne spremenljivke

Generiranje vzorca slučajne spremenljivke  $X$  temelji na vzorcu enakomerno porazdeljene slučajne spremenljivke  $U$  s porazdelitveno funkcijo  $F_U(u) = u$  (za  $0 \leq u \leq 1$ ). Vzorca povsem slučajne spremenljivke ni lahko generirati. Zato z računalniki običajno generiramo vzorce tako imenovanih psevdoslučajnih števil, ki so pravzaprav zaporedja determinističnih števil z zelo veliko periodo ponovitve. Ta zaporedja imajo enake lastnosti kot zaporedja povsem slučajnih števil. Skoraj vsi računalniški programi, namenjeni računanju (na primer prevajalniki: FORTRAN, C++, PASCAL, BASIC in drugi programi: MATLAB, MATHEMATICA, EXCEL, R, PYTHON), vključujejo tudi „generatorje slučajnih števil“, s katerimi generiramo zaporedja psevdoslučajnih števil, ki ustrezajo vzorcu slučajne spremenljivke  $U$ , ki je porazdeljena enakomerno od 0 do 1. Vzorec poljubno porazdeljene slučajne spremenljivke ali vektorja lahko generiramo z različnimi metodami: z inverzno metodo, metodo sprejema in zavrnitve, polarno metodo, metodo trakov, z mrežno metodo in drugimi. Najpogosteje uporabljamo inverzno metodo.

### 7.1.1 Inverzna metoda generiranja vzorca

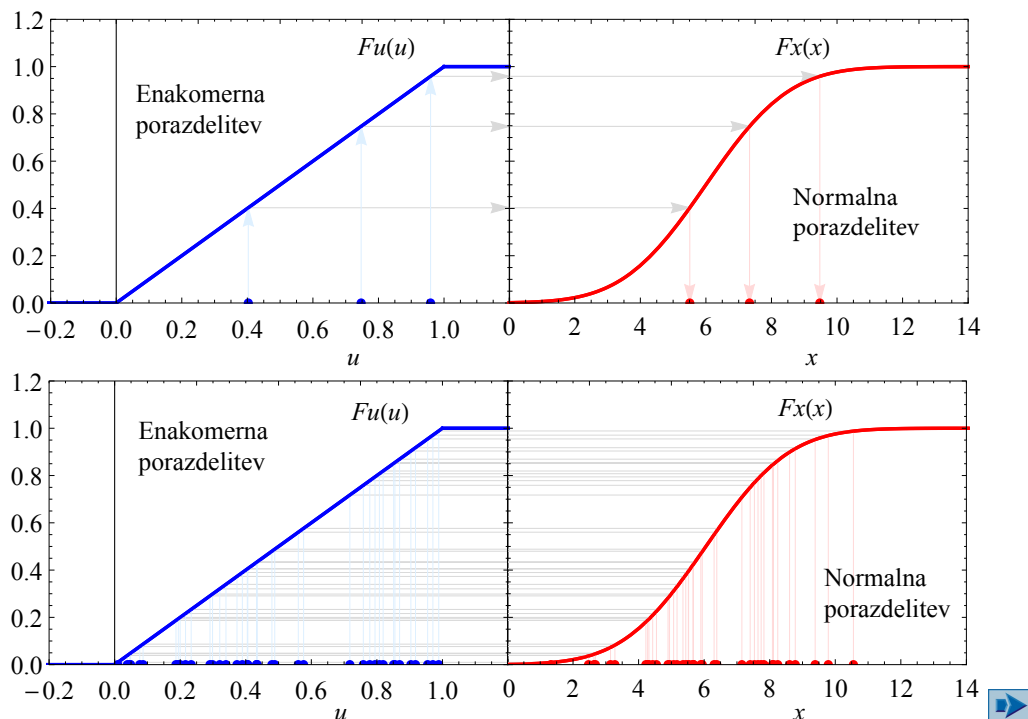
Vzemimo, da imamo enako velika vzorca dveh slučajnih spremenljivk. Prva slučajna spremenljivka,  $X$ , je porazdeljena po porazdelitvi  $F_X(x)$ , druga,  $U$ , pa je enakomerno porazdeljena od 0 do 1:  $F_U(u) = u$  (za  $0 \leq u \leq 1$ ). Elemente v obeh vzorcih razvrstimo po velikosti in predpostavimo, da sta verjetnosti, da sta slučajni spremenljivki  $X$  in  $U$  manjši od  $i$ -tega elementa ustreznega vzorca, enaki, ne glede na porazdelitev. To predpostavko v enačbi za vsak  $i$  zapišemo takole:

$$P[X \leq x_i] = F_X(x_i) = P[U \leq u_i] = F_U(u_i) = u_i.$$

Iz zadnje enačbe lahko izpeljemo izraz za določitev elementa vzorca slučajne spremenljivke:

$$F_X(x_i) = u_i \quad \longrightarrow \quad x_i = F_X^{-1}(u_i). \quad (7.1)$$

Če torej poznamo porazdelitveno funkcijo slučajne spremenljivke oziroma njeno inverzno funkcijo in imamo vzorec  $(u_i, i = 1, \dots, n)$  enakomerno porazdeljene slučajne spremenljivke  $U$ , lahko po zadnji enačbi določimo vzorec slučajne spremenljivke  $X$ , kjer je  $n$  velikost vzorca. Inverzno metodo nazorno prikažemo tudi grafično na sliki 7.2. Z modro je označena enakomerno porazdeljena slučajna spremenljivka  $U$  s pripadajočim vzorcem. Vsakemu elementu vzorca slučajne spremenljivke  $U$  ustreza element vzorca slučajne spremenljivke  $X$ . Na zgornjem delu slike metodo opišemo na primeru treh generiranih vrednosti, na spodnjem delu pa je prikazan nekoliko večji vzorec. Na spodnji sliki lahko opazimo, da so točke v okolici srednje vrednosti  $m_X = 6$  bolj zgoščene kot na obeh repih, kjer je točk manj.



SLIKA 7.2: Inverzna metoda generiranja vzorca

V preglednici 7.1 podajamo nekaj značilnih primerov uporabe inverzne metode za generiranje vzorcev slučajnih spremenljivk, katerih porazdelitvene funkcije poznamo v zaključeni obliki.

PREGLEDNICA 7.1: Uporaba inverzne metode

Porazdelitev	Način generiranja
Dogodek, ki se zgodi z verjetnostjo $p$	Dogodek se zgodi, če je $u_i$ manjši od $p$
Enakomerna zvezna $X$ od $a$ do $b$	$x_i = a + u_i (b - a)$
EkspONENTNA porazdelitev $T$	$t_i = -\frac{1}{\lambda} \ln(1 - u_i)$
Gumbelova porazdelitev maksimuma $Y$	$y_i = u - \frac{1}{\alpha} \ln(-\ln u_i)$
Gumbelova porazdelitev minimuma $Z$	$z_i = u + \frac{1}{\alpha} \ln(-\ln(1 - u_i))$
Fréchetova porazdelitev maksimuma $Y$	$y_i = u (-\ln u_i)^{-\frac{1}{k}}$
Fréchetova porazdelitev minimuma $Z$	$z_i = u (-\ln(1 - u_i))^{-\frac{1}{k}}$
Weibullova porazdelitev maksimuma $Y$	$y_i = \omega - (\omega - u) (-\ln u_i)^{\frac{1}{k}}$
Weibullova porazdelitev minimuma $Z$	$z_i = \varepsilon + (u - \varepsilon) (\ln(1 - u_i))^{\frac{1}{k}}$

### 7.1.2 Generiranje vzorca normalno porazdeljenih slučajnih spremenljivk

Porazdelitvene funkcije normalne porazdelitve v zaključeni obliki ne poznamo, prav tako ne poznamo njene inverzne funkcije. Zato za generiranje vzorca normalno porazdeljene slučajne spremenljivke ne moremo uporabiti analitične inverzne metode. Če računalniški program omogoča numerični izračun inverzne porazdelitvene funkcije normalne porazdelitve, lahko uporabimo inverzno metodo, četudi funkcije  $F_X^{-1}(u_i)$  ne poznamo v eksplisitni obliki.

Na primer: v programu EXCEL je funkcija vgrajena kot `norminv`, v programu MATHEMATICA lahko problem rešimo z ukazom `InverseCDF[NormalDistribution[mx, sx]]`.

Če programska oprema ne omogoča uporabe funkcije  $F_X^{-1}(u_i)$ , lahko vzorec normalno porazdeljene slučajne spremenljivke generiramo z Box-Müllerjevo polarno metodo. S to metodo vzorec standardizirano normalno porazdeljene slučajne spremenljivke izračunamo iz dveh vzorcev  $(u_{1i}, i = 1, \dots, n)$  in  $(u_{2i}, i = 1, \dots, n)$  slučajnih spremenljivk  $U_1$  in  $U_2$ , ki sta neodvisni in porazdeljeni enakomerno od 0 do 1, po naslednjih enačbah:

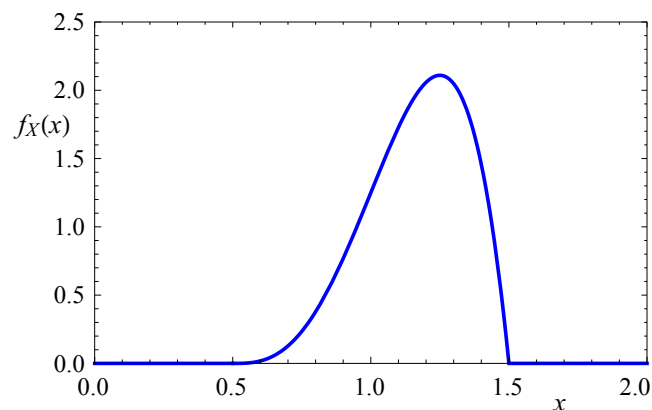
$$\begin{aligned} z_{1i} &= \sqrt{-2 \ln u_{1i}} \sin(2 \pi u_{2i}), \\ z_{2i} &= \sqrt{-2 \ln u_{1i}} \cos(2 \pi u_{2i}), \end{aligned} \quad (7.2)$$

kjer  $z_{1i}$  in  $z_{2i}$  predstavljata vzorec neodvisnih standardizirano normalno porazdeljenih slučajnih spremenljivk  $Z_1$  in  $Z_2$ . Vzorca  $x_{1i}$  in  $x_{2i}$  poljubnih normalno porazdeljenih slučajnih spremenljivk  $X_1$  in  $X_2$  s pričakovanimi vrednostima  $m_{X_1}$  in  $m_{X_2}$  ter standardnima deviacijama  $\sigma_{X_1}$  in  $\sigma_{X_2}$  iz vzorcev  $z_{1i}$  in  $z_{2i}$  določimo s transformacijo:

$$\begin{aligned} x_{1i} &= m_{X_1} + z_{1i} \sigma_{X_1}, \\ x_{2i} &= m_{X_2} + z_{2i} \sigma_{X_2}. \end{aligned} \quad (7.3)$$

### 7.1.3 Metoda sprejema/zavrnitve

Pri inverzni metodi smo uporabili porazdelitveno funkcijo oziroma inverzno porazdelitveno funkcijo, metoda sprejema in zavrnitve pa temelji na **gostoti verjetnosti**. Vzemimo, da želimo generirati vzorec slučajne spremenljivke  $X$  z zalogo vrednosti na končnem intervalu  $[a, b]$ . Taka porazdelitev je na primer porazdelitev  $\text{beta}_{(4,2)}(x - 0.5)$  za  $x \in [0.5, 1.5]$ , ki je prikazana na sliki 7.3.

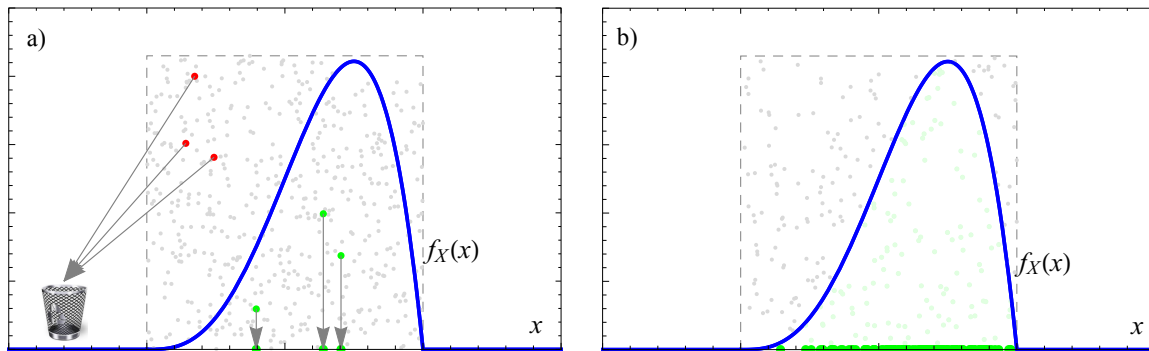


SLIKA 7.3: Gostota verjetnosti, porazdelitev  $\text{beta}(4, 2)$

Ideja metode sprejema in zavrnitve je v generiranju vzorca enakomerno porazdeljenega slučajnega vektorja na območju  $x \in [a, b]$  ter  $y \in [0, \max_{x \in [a,b]} (f_X(x))]$ , kot je prikazano na sliki 7.4.

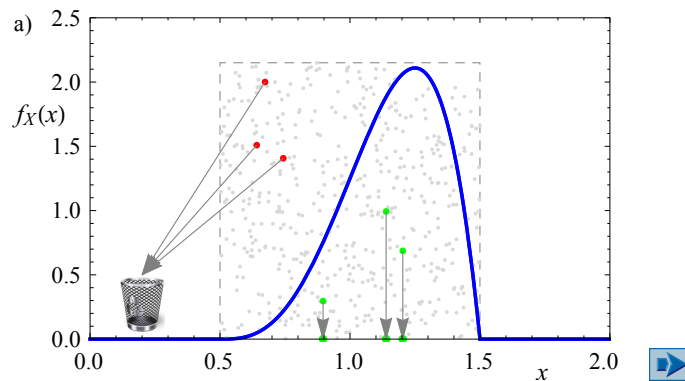
Interval  $[a, b]$  je običajno določen z zalogo vrednosti slučajne spremenljivke  $X$ . Metoda bi sicer delovala tudi, če bi ta interval povečali, a bi s tem povečanjem zmanjšali učinkovitost, saj bi generirali množico točk, za katere zanesljivo vemo, da ne morejo ležati pod grafom gostote verjetnosti in jih zato ne moremo sprejeti. Optimalno je torej najmanjše pravokotno območje, ki očrta graf gostote verjetnosti slučajne spremenljivke  $X$ .

Točke, ki ležijo pod grafom gostote verjetnosti sprejmemo, to pomeni, da vrednost  $x$  dodamo v vzorec. Točke, ki ležijo nad grafom gostote verjetnosti zavrnilimo. Na sliki 7.4a sprejete točke označimo z zeleno barvo, zavrnjene pa z rdečo. Na sliki 7.4b vse sprejete točke označimo z zeleno, na tej sliki se vidi, da je rezultat vzorčenja seznam vrednosti na območju od 0.5 do 1.5, pri čemer so točke bolj gosto razporejene tam, kjer je gostota verjetnosti večja.



SLIKA 7.4: Metoda sprejema/zavrnitve

Animacija generiranja vzorca slučajne spremenljivke po metodi sprejema in zavrnitve je prikazana na sliki 7.5.



SLIKA 7.5: Animacija generiranja po metodi sprejema in zavrnitve

### 7.1.4 Primeri

**Primer 7.1** *Generiranje po inverzni metodi.* Z uporabo inverzne metode generirajmo vzorec slučajne spremenljivke  $T$ , porazdeljene po eksponentni porazdelitvi, ter slučajne spremenljivke  $Y$ , porazdeljene po Gumbelovi porazdelitvi maksimuma. Parameter za eksponentno porazdelitev je:  $\lambda = 0.3$ , za Gumbelovo porazdelitev pa sta:  $u = 10, \alpha = 0.4$ .

Najprej moramo generirati vzorec enakomerno porazdeljene slučajne spremenljivke  $U$ , nato pa z uporabo izrazov v preglednici 7.1 določimo vzorec slučajne spremenljivke  $T$  oziroma  $Y$ . V preglednici 7.2 prikazujemo prvih deset slučajno generiranih vrednosti  $u_i$  in iz njih izračunanih vrednosti  $t_i$ . Na primer: iz prve generirane vrednosti  $u_1 = 0.9384$  določimo prvo vrednost v vzorcu eksponentno porazdeljene

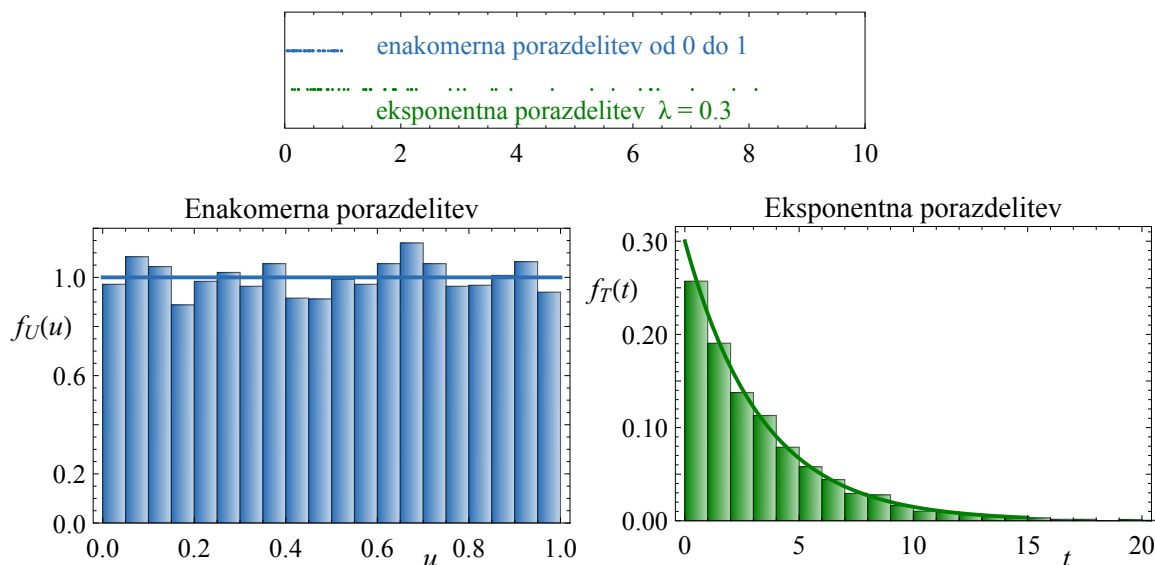
slučajne spremenljivke  $T$ :

$$t_1 = -\frac{1}{\lambda} \ln(1 - u_1) = -\frac{1}{0.3} \ln(1 - 0.9384) = 9.290.$$

PREGLEDNICA 7.2: Razultat inverze metode – eksponentna porazdelitev

Generirane vrednosti enakomerno in eksponentno porazdeljene slučajne spremenljivke										
$u_i$	0.9384	0.3725	0.5234	0.7788	0.0980	0.8043	0.9363	0.4146	0.8976	0.1660
$t_i$	9.2900	1.5536	2.4700	5.0282	0.3438	5.4369	9.1780	1.7851	7.5975	0.6051

Na sliki 7.6 zgoraj grafično prikazujemo vzorca velikosti  $n = 50$  enakomerno porazdeljene slučajne spremenljivke  $U$  in eksponentno porazdeljene slučajne spremenljivke  $T$ . Že pri tako majhnem vzorcu lahko prepoznamo lastnosti eksponentne porazdelitve, pri kateri je gostota verjetnosti večja pri nižjih vrednostih spremenljivke in se postopoma znižuje.



SLIKA 7.6: Inverzna metoda generiranja vzorca – eksponentna porazdelitev

Na sliki 7.6 spodaj sta tudi stolpčna diagrama (histograma), ki prikazujeta razporeditev vrednosti večjega vzorca ( $n = 5000$ ) za obe porazdelitvi. S primerjavo med stolpčnim diagramom in gostoto verjetnosti slučajnih spremenljivk  $U$  in  $T$  vidimo, da je razporeditev vrednosti ustrezna.

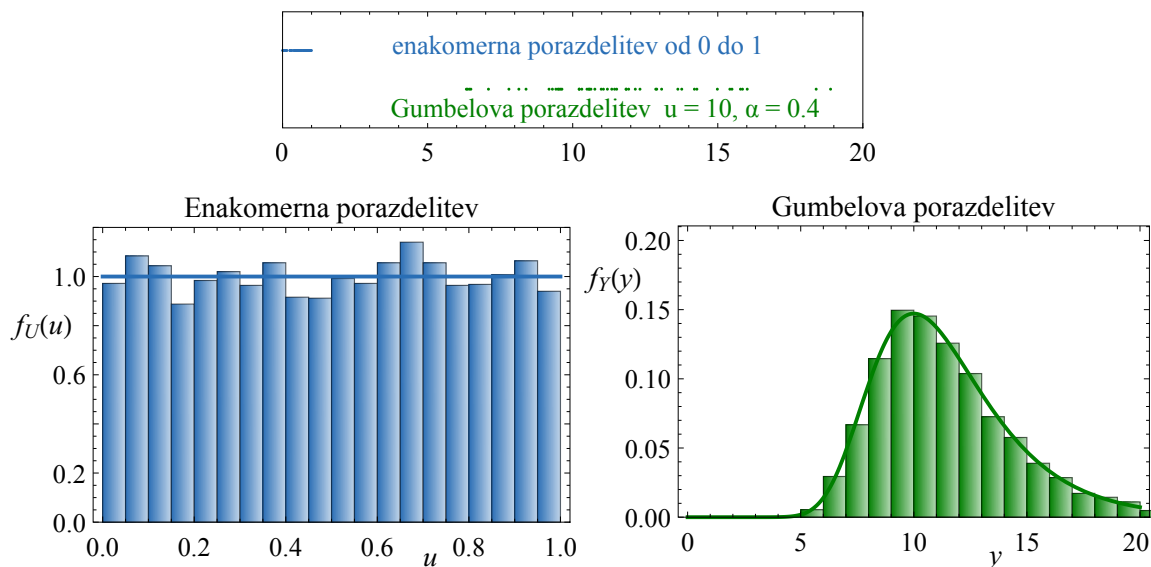
V preglednici 7.3 prikazujemo prvih deset slučajno generiranih vrednosti  $u_i$  in iz njih izračunanih vrednosti  $y_i$ . Na primer: iz prve generirane vrednosti  $u_1 = 0.4009$  določimo prvo vrednost v vzorcu slučajne spremenljivke  $Y$ , porazdeljene po Gumbelovi porazdelitvi:

$$y_1 = u - \frac{1}{\alpha} \ln(-\ln(u_1)) = 10 - \frac{1}{0.4} \ln(-\ln(0.4009)) = 10.225.$$

PREGLEDNICA 7.3: Razultat inverze metode – Gumbelova porazdelitev

Generirane vrednosti enakomerno in Gumbelovo porazdeljene slučajne spremenljivke										
$u_i$	0.4009	0.7661	0.5612	0.1467	0.9921	0.5106	0.8341	0.7233	0.3591	0.5018
$y_i$	10.225	13.307	11.372	8.3700	22.107	10.993	14.268	12.818	9.9403	10.930

Na sliki 7.7 zgoraj grafično s točkami prikazujemo vzorca velikosti  $n = 50$  enakomerno porazdeljene slučajne spremenljivke  $U$  in slučajne spremenljivke  $Y$ , porazdeljene po Gumbelovi porazdelitvi. Že pri tako majhnem vzorcu lahko prepoznamo lastnosti Gumbelove porazdelitve, pri kateri je gostota verjetnosti največja v okolici vrednosti  $u = 10$ , za manjše in večje vrednosti pa je gostota manjša. Opazimo tudi nesimetričnost porazdelitve. Na isti sliki spodaj sta prikazana tudi stolpčna diagrama za večji vzorec ( $n = 5000$ ), s katerima ponazorimo razpored vrednosti vzorcev. Le-ta ustrežata izbrani porazdelitvi.



SLIKA 7.7: Inverzna metoda generiranja vzorca – Gumbelova porazdelitev

**Primer 7.2** *Generiranje po metodi Box-Müller. Z uporabo Box-Müllerjeve metode generirajmo vzorec normalno porazdeljene slučajne spremenljivke s srednjo vrednostjo  $m_X = 5$  in  $\sigma_X = 2$ .*

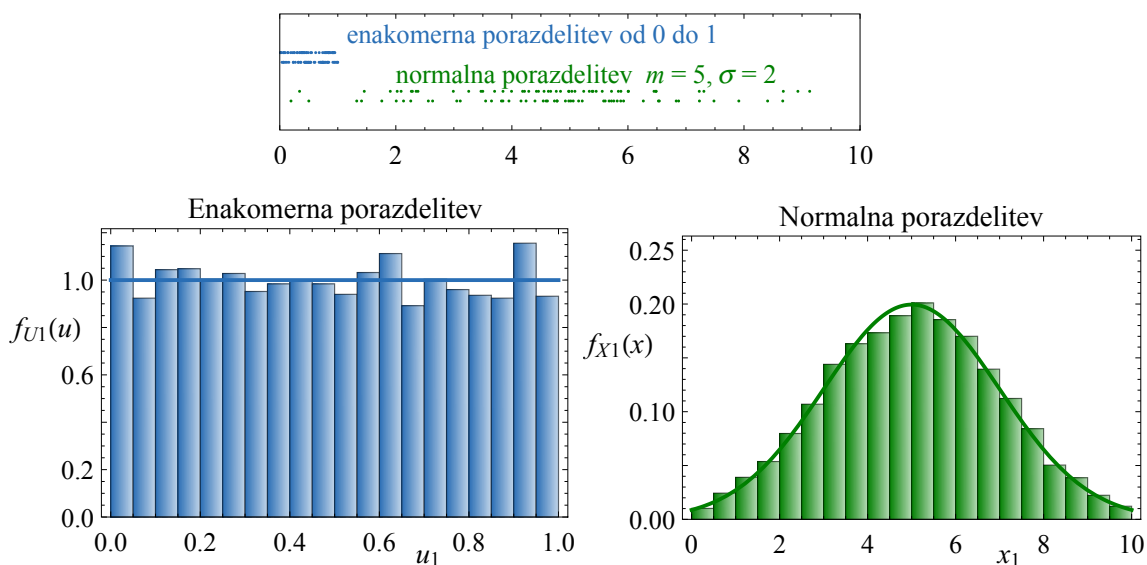
V preglednici 7.4 prikazujemo prvih deset vrednosti vzorca enakomerno porazdeljenih slučajnih spremenljivk  $U_1$  in  $U_2$  ter ustreznih vrednosti normalno porazdeljenih slučajnih spremenljivk  $X_1$  in  $X_2$ . Na primer: iz prvih generiranih vrednosti  $u_{11}$  in  $u_{21}$  določimo prvi vrednosti v vzorcu slučajnih spremenljivk  $X_1$  in  $X_2$ :

$$\begin{aligned} x_{11} &= \sqrt{-2 \ln(u_{11})} \sin(2\pi u_{21}) = \sqrt{-2 \ln(0.3956)} \sin(2\pi \cdot 0.7620) = 2.2840, \\ x_{21} &= \sqrt{-2 \ln(u_{11})} \cos(2\pi u_{21}) = \sqrt{-2 \ln(0.3956)} \cos(2\pi \cdot 0.7620) = 5.2057. \end{aligned} \quad (7.4)$$

PREGLEDNICA 7.4: Generiranje po Box-Müllerjevi metodi

Generirane vrednosti enakomerno in normalno porazdeljenih slučajnih spremenljivk										
$u_{1i}$	0.3956	0.1897	0.7273	0.4939	0.1811	0.5518	0.1928	0.7393	0.4273	0.6994
$u_{2i}$	0.7620	0.0782	0.1713	0.8727	0.5581	0.4212	0.5647	0.2521	0.000031	0.3338
$x_{1i}$	2.2840	6.7210	6.4049	3.2959	3.6793	6.0363	3.5654	6.5545	5.0005	6.4624
$x_{2i}$	5.2047	8.2150	5.7570	6.6549	1.5468	3.0810	1.6667	4.9798	7.6082	4.1503

Na sliki 7.8 s točkami prikazujemo vse štiri vzorce velikosti  $n = 50$  slučajnih spremenljivk  $U_1, U_2, X_1$  in  $X_2$ . Vidimo, da iz dveh vzorcev slučajnih spremenljivk, porazdeljenih enakomerno od 0 do 1, določimo vzorec dveh neodvisnih spremenljivk, porazdeljenih po normalni porazdelitvi s parametri  $m_X = 5$  in  $\sigma_X = 2$ . S stolpčnima diagramoma prikazujemo še razpored vrednosti v večjih vzorcih  $n = 5000$  in ju primerjamo s teoretičnima porazdelitvama.



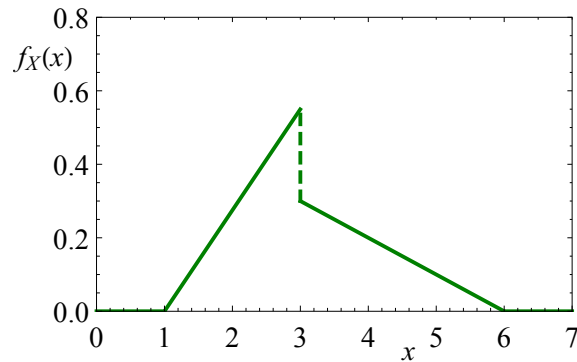
SLIKA 7.8: Box-Müllerjeva metoda – normalna porazdelitev

**Primer 7.3** *Uporaba metode sprejema in zavrnitve. Generirajmo vzorec slučajne spremenljivke  $X$ , ki je porazdeljena po porazdelitvi, definirani z naslednjim izrazom:*

$$f_X(x) = \begin{cases} 0.275(x-1) & \text{za } 1 \leq x \leq 3, \\ 0.1(6-x) & \text{za } 3 < x \leq 6. \end{cases}$$

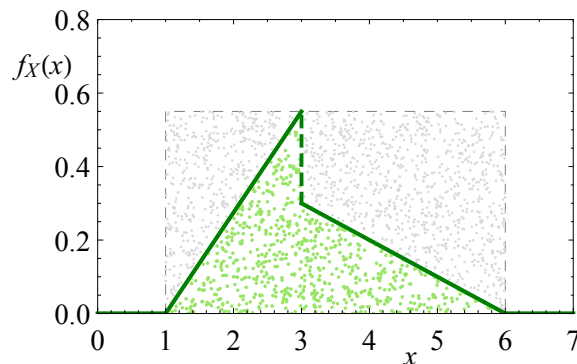
Na sliki 7.9 je prikazana gostota verjetnosti slučajne spremenljivke  $X$ .





SLIKA 7.9: Sestavljena trikotna porazdelitev

Najprej generiramo dva neodvisna vzorca enakomerno porazdeljenih slučajnih spremenljivk,  $U_1$  je enakomerno porazdeljena od 1 do 6,  $U_2$  pa od 0 do 0.55. Vzorec 2000 elementov slučajnih spremenljivk  $U_1$  in  $U_2$  prikazujemo na sliki 7.10. Z zeleno so označene točke, ki jih sprejmemo, s sivo pa tiste, ki jih zavrnemo.



SLIKA 7.10: Sestavljena trikotna porazdelitev

V preglednici 7.5 je zapisanih prvih desetih generiranih vrednosti slučajnih spremenljivk  $U_1$  in  $U_2$ , izračunane vrednosti gostote verjetnosti ter generirana vrednost vzorca slučajne spremenljivke  $X$ , v primeru sprejema in oznaka, če je prišlo do zavrnitve. Na primer: ker je vrednost gostote verjetnosti pri  $u_{11} = 2.803$  večja od vrednosti  $u_{21} = 0.056$ , vrednost sprejmemo:

$$u_{21} = 0.056 < f_X(u_{11}) = f_X(2.803) = 0.496 \quad \rightarrow \quad x_1 = 2.803.$$

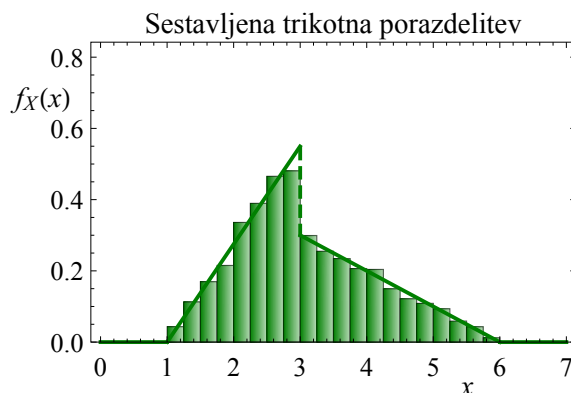
Podobno naredimo tudi za drugo in tretjo vrednost. Drugače je pri četrtem elementu vzorca: ker je gostota verjetnosti pri  $u_{14} = 3.188$  manjša od vrednosti  $u_{24} = 0.529$ , ta element zavrnemo:

$$u_{24} = 0.529 > f_X(u_{14}) = f_X(3.453) = 0.255 \quad \rightarrow \quad \text{ta element zavrnemo.}$$

PREGLEDNICA 7.5: Generiranje po metodi sprejema in zavrnitve

Generiranje po metodi sprejema in zavrnitve										
$u_{1i}$	2.803	1.850	3.188	3.453	5.587	1.461	3.898	2.983	4.253	1.774
$u_{2i}$	0.056	0.174	0.040	0.529	0.277	0.160	0.060	0.405	0.449	0.037
$f_X(u_{1i})$	0.496	0.234	0.281	0.255	0.041	0.127	0.210	0.545	0.175	0.213
$x_i$	2.803	1.850	3.188	-	-	-	3.898	2.983	-	1.774

Na sliki 7.11 s stolpčnim diagramom prikazujemo razporeditev elementov vzorca in za primerjavo tudi gostoto verjetnosti.

SLIKA 7.11: Generirane vrednosti slučajne spremenljivke  $X$ 

## 7.2 Generiranje vzorca slučajnega vektorja

Pri generiranju vzorca slučajnega vektorja bomo obravnavali tri različne primere: generiranje vzorca poljubno porazdeljenih neodvisnih slučajnih spremenljivk, generiranje vzorca poljubno porazdeljenih odvisnih slučajnih spremenljivk in generiranje vzorca normalno porazdeljenih koreliranih slučajnih spremenljivk.

### 7.2.1 Neodvisne slučajne spremenljivke – inverzna metoda

Generiranje vzorca slučajnega vektorja neodvisnih slučajnih spremenljivk je enako generiranju vzorca posameznih slučajnih spremenljivk. Zato lahko uporabimo poljuben način generiranja, kot sta na primer inverzna metoda ali metoda sprejema in zavrnitve. V primeru uporabe inverzne metode lahko generiranje vzorca povzamemo z naslednjimi izrazi:

$$\begin{aligned}
F_{X_1}(x_{1i}) = u_{1i} &\rightarrow x_{1i} = F_{X_1}^{-1}(u_{1i}), \\
F_{X_2}(x_{2i}) = u_{2i} &\rightarrow x_{2i} = F_{X_2}^{-1}(u_{2i}), \\
F_{X_3}(x_{3i}) = u_{3i} &\rightarrow x_{3i} = F_{X_3}^{-1}(u_{3i}), \\
&\dots \\
F_{X_k}(x_{ki}) = u_{ki} &\rightarrow x_{ki} = F_{X_k}^{-1}(u_{ki}),
\end{aligned} \tag{7.5}$$

kjer so  $x_{1i}, x_{2i}, \dots, x_{ki}$  generirane vrednosti neodvisnih slučajnih spremenljivk  $X_1, X_2, \dots, X_k$  in  $u_{1i}, u_{2i}, \dots, u_{ki}$  generirane vrednosti neodvisnih slučajnih spremenljivk  $U_1, U_2, \dots, U_k$ , porazdeljene po enakomerni porazdelitvi. Poudariti velja, da moramo pri generiranju vzorca enakomerno porazdeljenih slučajnih spremenljivk  $U_j$  le-te generirati kot neodvisne spremenljivke.

Za lažje razumevanje generiranja odvisnih slučajnih spremenljivk naredimo še naslednji razmislek, ki bo pripeljal do izrazov (7.5): za neodvisne slučajne spremenljivke velja, da je pogojna porazdelitvena funkcija enaka brezpogojni:

$$F_{X_2|X_1 \leq x_1}(x_2) = P[X_2 \leq x_2 | X_1 \leq x_1] = \frac{P[X_1 \leq x_1 \cap X_2 \leq x_2]}{P[X_1 \leq x_1]} = \frac{F_{X_1, X_2}(x_1, x_2)}{F_{X_1}(x_1)} = F_{X_2}(x_2), \tag{7.6}$$

od koder sledi pogosto uporabljena enačba za določitev porazdelitve slučajnega vektorja neodvisnih spremenljivk, ki pravi, da je porazdelitvena funkcija slučajnega vektorja neodvisnih spremenljivk kar produkt porazdelitvenih funkcij slučajnih spremenljivk:

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2). \tag{7.7}$$

Ob uporabi inverzne metode generiranja slučajnih spremenljivk porazdelitveno funkcijo slučajnega vektorja  $\{X_1, X_2\}$  izenačimo s porazdelitveno funkcijo slučajnega vektorja  $\{U_1, U_2\}$ , ki ga sestavljata neodvisni slučajni spremenljivki, porazdeljeni enakomerno od 0 do 1:

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2) = F_{U_1, U_2}(u_1, u_2) = F_{U_1}(u_1) F_{U_2}(u_2) = u_1 u_2. \tag{7.8}$$

Začnemo z generiranjem vzorca prve slučajne spremenljivke  $X_1$ :

$$F_{X_1}(x_{1i}) = u_{1i} \rightarrow x_{1i} = F_{X_1}^{-1}(u_{1i}). \tag{7.9}$$

Ob upoštevanju (7.9) v (7.8) ostane le izraz, na osnovi katerega lahko generiramo vzorec druge slučajne spremenljivke  $X_2$ :

$$F_{X_2}(x_{2i}) = u_{2i} \rightarrow x_{2i} = F_{X_2}^{-1}(u_{2i}). \tag{7.10}$$

Posplošitev za poljubno število neodvisnih slučajnih spremenljivk je preprosta.

## 7.2.2 Odvisne slučajne spremenljivke – inverzna metoda

Za dve odvisni slučajni spremenljivki  $X_1$  in  $X_2$  velja, da porazdelitvene funkcije slučajnega vektorja  $\{X_1, X_2\}$  ne moremo zapisati kot produkt porazdelitvenih funkcij  $F_{X_1}(x_1)$  in  $F_{X_2}(x_2)$ , kot pri neodvisnih slučajnih spremenljivkah. Velja namreč:

$$F_{X_2|X_1 \leq x_1}(x_2) = P[X_2 \leq x_2 | X_1 \leq x_1] = \frac{P[X_1 \leq x_1 \cap X_2 \leq x_2]}{P[X_1 \leq x_1]} = \frac{F_{X_1, X_2}(x_1, x_2)}{F_{X_1}(x_1)}, \quad (7.11)$$

od koder sledi zveza:

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2|X_1 \leq x_1}(x_2) \neq F_{X_1}(x_1) F_{X_2}(x_2). \quad (7.12)$$

Zato moramo vzorec slučajnih spremenljivk generirati postopoma. Najprej generiramo vzorec prve slučajne spremenljivke:

$$x_{1i} = F_{X_1}^{-1}(u_{1i}). \quad (7.13)$$

Nato poiščemo pogojno porazdelitveno funkcijo  $F_{X_2|X_1=x_1}(x_2)$ , na osnovi katere bomo lahko generirali vzorec naslednje slučajne spremenljivke. To porazdelitveno funkcijo najlažje določimo iz pogojne gostote verjetnosti, ki jo določimo po naslednji enačbi:

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)}. \quad (7.14)$$

Porazdelitveno funkcijo določimo z integracijo gostote verjetnosti:

$$F_{X_2|X_1=x_1}(x_2) = \int_{-\infty}^{x_2} f_{X_2|X_1=x_1}(\tilde{x}) d\tilde{x}. \quad (7.15)$$

Vzorec slučajne spremenljivke  $x_2$  generiramo z uporabo inverzne porazdelitvene funkcije (7.15):

$$F_{X_2|X_1=x_{1i}}(x_{2i}) = u_{2i} \rightarrow x_{2i} = F_{X_2|X_1=x_{1i}}^{-1}(u_{2i}). \quad (7.16)$$

Podobno naredimo tudi za vse nadaljnje spremenljivke:

$$f_{X_3|X_1=x_1 \cap X_2=x_2}(x_3) = \frac{f_{X_1, X_2, X_3}(x_1, x_2, x_3)}{f_{X_1, X_2}(x_1, x_2)}, \quad (7.17)$$

opravimo integracijo za določitev pogojne porazdelitvene funkcije:

$$F_{X_3|X_1=x_1 \cap X_2=x_2}(x_3) = \int_{-\infty}^{x_3} f_{X_3|X_1=x_1 \cap X_2=x_2}(\tilde{x}) d\tilde{x} \quad (7.18)$$

in vzorec slučajne spremenljivke  $x_3$  generiramo z uporabo inverzne porazdelitvene funkcije (7.18):

$$F_{X_3|X_1=x_{1i} \cap X_2=x_{2i}}(x_{3i}) = u_{3i} \rightarrow x_{3i} = F_{X_3|X_1=x_{1i} \cap X_2=x_{2i}}^{-1}(u_{3i}). \quad (7.19)$$

Na osnovi enačb (7.13), (7.16) in (7.19) lahko zapišemo izraze za generiranje vzorca poljubno porazdeljenih odvisnih slučajnih spremenljivk:

$$\begin{aligned}
 x_{1i} &= F_{X_1}^{-1}(u_{1i}), \\
 x_{2i} &= F_{X_2|X_1=x_{1i}}^{-1}(u_{2i}), \\
 x_{3i} &= F_{X_3|X_1=x_{1i} \cap X_2=x_{2i}}^{-1}(u_{3i}), \\
 &\dots \\
 x_{ki} &= F_{X_k|X_1=x_{1i} \cap X_2=x_{2i} \cap \dots \cap X_{k-1}=x_{k-1,i}}^{-1}(u_{ki}),
 \end{aligned} \tag{7.20}$$

kjer so  $x_{1i}, x_{2i}, \dots, x_{ki}$  generirane vrednosti neodvisnih slučajnih spremenljivk  $X_1, X_2, \dots, X_k$ , in  $u_{1i}, u_{2i}, \dots, u_{ki}$  generirane vrednosti neodvisnih slučajnih spremenljivk  $U_1, U_2, \dots, U_k$ , porazdeljene po enakomerni porazdelitvi. Če primerjamo enačbe (7.5) in (7.20), opazimo, da za generiranje vzorca odvisne slučajne spremenljivke  $X_2$  potrebujemo vzorčno vrednost slučajne spremenljivke  $x_{1i}$ , za generiranje vzorca  $X_3$  potrebujemo vzorčne vrednosti  $x_{1i}$  in  $x_{2i}$  in tako dalje, medtem ko pri generiranju vzorca neodvisnih slučajnih spremenljivk teh vrednosti ne potrebujemo.

### 7.2.3 Korelirane normalno porazdeljene spremenljivke

Pri generiranju slučajnega vektorja  $\mathbf{X}$  koreliranih normalno porazdeljenih slučajnih spremenljivk bomo izhajali iz vektorja  $\mathbf{Y}$  standardizirano normalnih neodvisnih slučajnih spremenljivk. Vektor  $\mathbf{X}$  je določen z vektorjem srednjih vrednosti  $\mathbf{m}_X$  in kovariančno matriko  $\mathbf{C}_X$ , medtem ko so srednje vrednosti slučajnega vektorja  $\mathbf{Y}$  enake nič in je kovariančna matrika enotska matrika  $\mathbf{C}_Y = \mathbf{I}$ :

$$\mathbf{m}_X = \begin{bmatrix} m_{X_1} \\ m_{X_2} \\ \vdots \\ m_{X_n} \end{bmatrix}, \quad \mathbf{C}_X = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix}, \tag{7.21}$$

$$\mathbf{m}_Y = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{C}_Y = \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{7.22}$$

Med njima velja linearna zveza, ki jo v matrični obliki zapišemo z naslednjim izrazom:

$$\mathbf{X} = \mathbf{m}_X + \mathbf{H}\mathbf{Y} \quad \rightarrow \quad \mathbf{X} - \mathbf{m}_X = \mathbf{H}\mathbf{Y}, \tag{7.23}$$

kjer je  $\mathbf{H}$  trikotna transformacijska matrika, ki jo moramo še določiti. Za določitev matrike  $\mathbf{H}$  izhajamo iz enačbe za kovariančno matriko, v kateri upoštevamo, da je varianca oziroma kovarianca pričakovana vrednost funkcije slučajne spremenljivke:

$$\sigma_{X_1}^2 = E[(X_1 - m_{X_1})^2] \quad \text{ozioroma} \quad \sigma_{X_1 X_2} = E[(X_1 - m_{X_1})(X_2 - m_{X_2})]. \tag{7.24}$$

V matrični obliki kovariančno matriko zapišemo takole:

$$\mathbf{C}_X = E \left[ (\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T \right]. \quad (7.25)$$

Če v enačbo (7.25) vstavimo (7.23), dobimo:

$$\begin{aligned} \mathbf{C}_X &= E \left[ (\mathbf{H}\mathbf{Y})(\mathbf{H}\mathbf{Y})^T \right] = E \left[ \mathbf{H}(\mathbf{Y}\mathbf{Y}^T)\mathbf{H}^T \right] = \\ &\quad \dots \text{upoštevamo, da matrika } \mathbf{H} \text{ ni slučajna, zato jo lahko izpostavimo } \dots \\ &= \mathbf{H} E \left[ \mathbf{Y}\mathbf{Y}^T \right] \mathbf{H}^T = \mathbf{H} \mathbf{C}_Y \mathbf{H}^T = \\ &\quad \dots \text{upoštevamo, da je kovariančna matrika enotska: } \mathbf{C}_Y = \mathbf{I} \dots \\ &= \mathbf{H} \mathbf{H}^T = \mathbf{H} \mathbf{H}^T. \end{aligned} \quad (7.26)$$

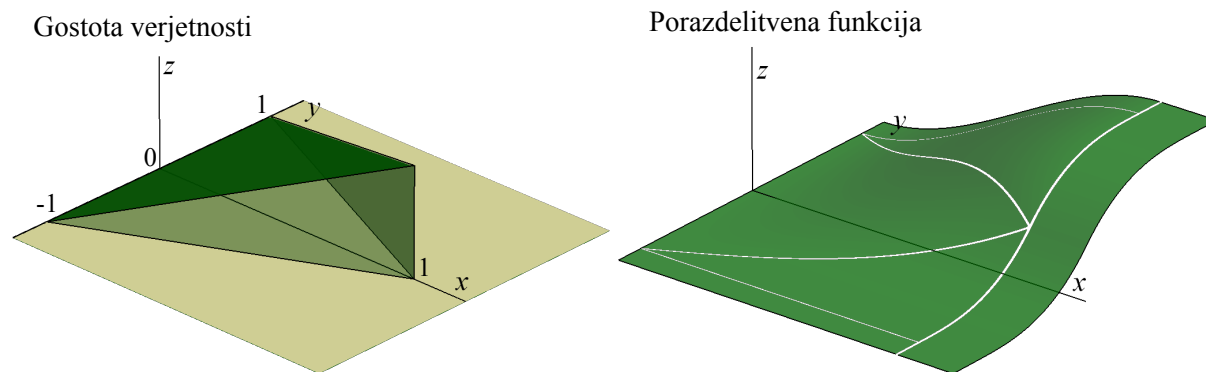
V primeru, da simetrično pozitivno definitno matriko zapišemo kot produkt dveh matrik  $\mathbf{H}$  in  $\mathbf{H}^T$ , lahko matriko  $\mathbf{H}$  določimo z razcepom Choleskega<sup>1</sup>.

## 7.2.4 Primeri

**Primer 7.4** *Generiranje po inverzni metodi. Z uporabo inverzne metode generirajmo vzorec slučajnega vektorja  $\{X, Y\}$ , porazdeljenega po porazdelitvi, definirani z naslednjim izrazom:*

$$f_{XY}(x, y) = 3x, \quad \text{za } 0 \leq x \leq 1, \quad x-1 \leq y \leq 1-x. \quad (7.27)$$

Grafa gostote verjetnosti in porazdelitvene funkcije sta prikazana na sliki 7.12.



SLIKA 7.12: Gostota verjetnosti in porazdelitvena funkcija slučajnega vektorja  $\{X, Y\}$

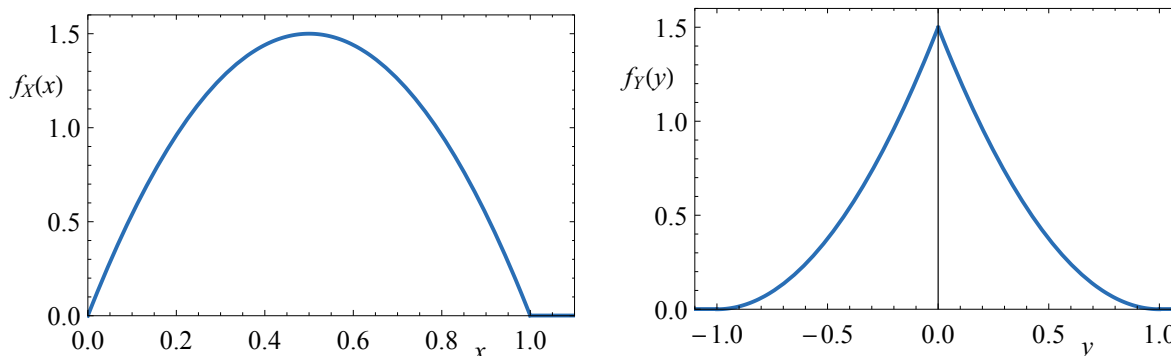
Določimo najprej robni gostoti verjetnosti  $f_X(x)$  in  $f_Y(y)$ :

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_{x-1}^{1-x} 3x dy = 3x y \Big|_{x-1}^{1-x} = 6x(1-x), \quad (7.28)$$

<sup>1</sup> André-Louis Cholesky, francoski oficir in matematik (1875–1918).

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \begin{cases} \int_0^{1+y} 3x dx = \frac{3x^2}{2} \Big|_0^{1+y} = \frac{3(1+y)^2}{2} & \text{za } y \leq 0, \\ \int_0^{1-y} 3x dx = \frac{3x^2}{2} \Big|_0^{1-y} = \frac{3(1-y)^2}{2} & \text{za } y > 0. \end{cases} \quad (7.29)$$

Grafa obeh robnih gostot verjetnosti prikazujemo na sliki 7.13. Za generiranje po inverzni metodi potre-



SLIKA 7.13: Robni gostoti verjetnosti  $f_X(x)$  in  $f_Y(y)$

bujemo inverzne porazdelitvene funkcije, kot je zapisano z enačbami (7.20). Najprej določimo inverzno porazdelitveno funkcijo slučajne spremenljivke  $X$ :

$$F_X(x) = \int_{-\infty}^x f_X(\tilde{x}) d\tilde{x} = \int_0^x 6\tilde{x}(1-\tilde{x}) d\tilde{x} = 3x^2 - 2x^3,$$

... pri določitvi inverzne funkcije kubičnega polinoma smo upoštevali le eno rešitev, tisto, ki ustreza vrednostim  $x \in [0, 1]$  v območju zaloge vrednosti  $X$  ...

(7.30)

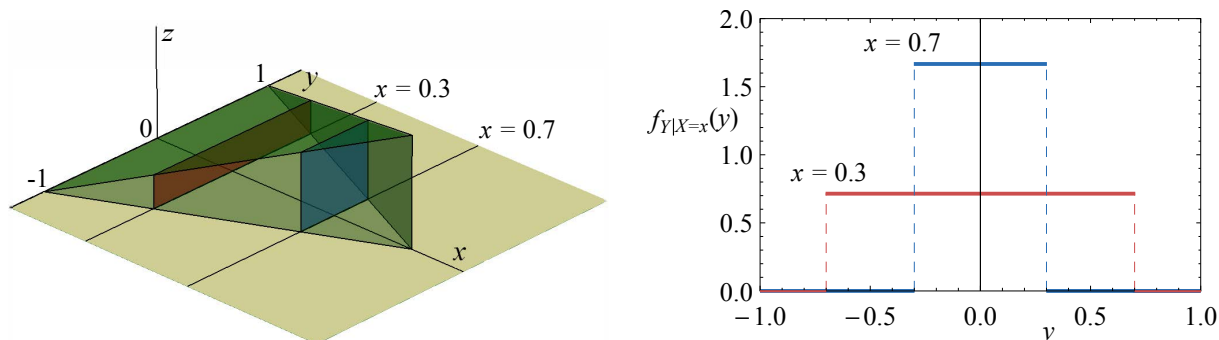
$$F_X^{-1}(u) = \frac{1}{4} \left( \frac{(1 + i\sqrt{3}) \left( 2u + 2\sqrt{(u-1)u-1} \right)^{2/3} + (1 - i\sqrt{3})}{\left( 2u + 2\sqrt{(u-1)u-1} \right)^{1/3}} + 2 \right).$$

Na prvi pogled zadnji izraz predstavlja kompleksno število, a se izkaže, da je za vrednosti  $u \in [0, 1]$  rezultat realno število med 0 in 1. V primeru, da analitične rešitve inverzne funkcije ne moremo najti, lahko uporabimo numerične metode.

Sedaj moramo določiti še pogojno gostoto verjetnosti  $f_{Y|X=x}(y)$  po naslednji enačbi;

$$f_{Y|X=x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = \begin{cases} \frac{3x}{6x(1-x)} = \frac{1}{2(1-x)} & \text{za } x-1 \leq y \leq 1-x, \\ 0 & \text{drugje.} \end{cases} \quad (7.31)$$

Ugotovimo lahko, da je pogojna slučajna spremenljivka  $Y$  pri določeni vrednosti  $x$  enakomerno porazdeljena od  $x-1$  do  $1-x$ . Določitev te pogojne porazdelitve grafično prikazujemo na sliki 7.14.



SLIKA 7.14: Določitev pogojne gostote verjetnosti  $f_{Y|X=x}(y)$  za dve vrednosti  $x$  (0.3 in 0.7)

Če želimo uporabiti inverzno metodo generiranja, moramo določiti še pogojno porazdelitveno funkcijo in njeno inverzno funkcijo:

$$F_{Y|X=x}(y) = \int_{-\infty}^y f_{Y|X=x}(\tilde{y}) d\tilde{y} = \begin{cases} 0 & \text{za } y < x - 1, \\ \frac{y-x+1}{2(1-x)} & \text{za } x - 1 \leq y \leq 1 - x, \\ 1 & \text{za } y > 1 - x, \end{cases} \quad (7.32)$$

... poiščemo inverzno porazdelitveno funkcijo:

$$F_{Y|X=x}^{-1}(u) = 2(1-x)u - (1-x).$$

Pripravimo enačbi inverznih porazdelitvenih funkcij tako, da bomo lahko uporabili izraze (7.20) za generiranje odvisnih slučajnih spremenljivk:

$$x_i = \frac{1}{4} \left( \frac{(1 + i\sqrt{3}) \left( 2u_{1i} + 2\sqrt{(u_{1i} - 1)u_{1i} - 1} \right)^{2/3} + (1 - i\sqrt{3})}{\left( 2u_{1i} + 2\sqrt{(u_{1i} - 1)u_{1i} - 1} \right)^{1/3}} + 2 \right), \quad (7.33)$$

$$y_i = 2(1 - x_i)u_{2i} - (1 - x_i).$$

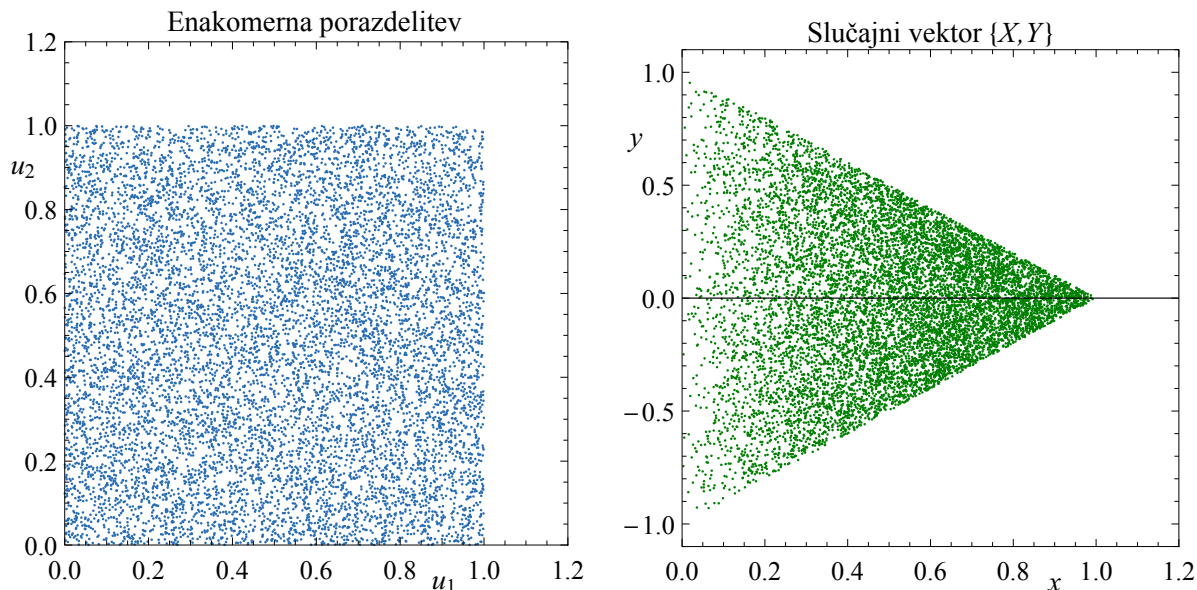
V preglednici 7.6 prikazujemo vrednosti prvih desetih elementov vzorca. Na primer, iz prve vrednosti  $u_{11} = 0.076$  po prvi od enačb (7.33) izračunamo prvo vrednost  $x_1 = 0.169$ . Nato uporabimo drugo izmed enačbo (7.33) in iz  $u_{21} = 0.561$  in  $x_1 = 0.169$  izračunamo še  $y_1 = 0.101$ .

PREGLEDNICA 7.6: Generiranje slučajnega vektorja po inverzni metodi

Generirane vrednosti										
$u_{1i}$	0.076	0.806	0.101	0.405	0.202	0.595	0.709	0.289	0.185	0.871
$u_{2i}$	0.561	0.201	0.732	0.221	0.240	0.894	0.880	0.658	0.914	0.062
$x_i$	0.169	0.718	0.196	0.436	0.289	0.564	0.644	0.355	0.275	0.775
$y_i$	0.101	-0.169	0.373	-0.315	-0.370	0.344	0.271	0.203	0.601	-0.197

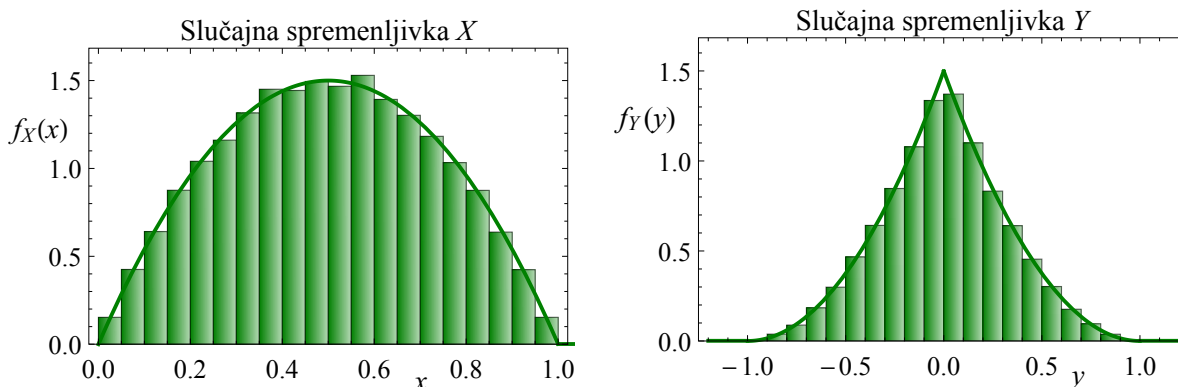


Na sliki 7.15 prikazujemo vzorca 10000 elementov slučajnih vektorjev  $\{U_1, U_2\}$  in  $\{X, Y\}$ . Vidimo razliko med enakomerno porazdeljenim vektorjem  $\{U_1, U_2\}$  in neenakomerno porazdelitvijo vektorja  $\{X, Y\}$ , ki se odraža v povečani gostoti točk pri višjih vrednostih  $X$ .



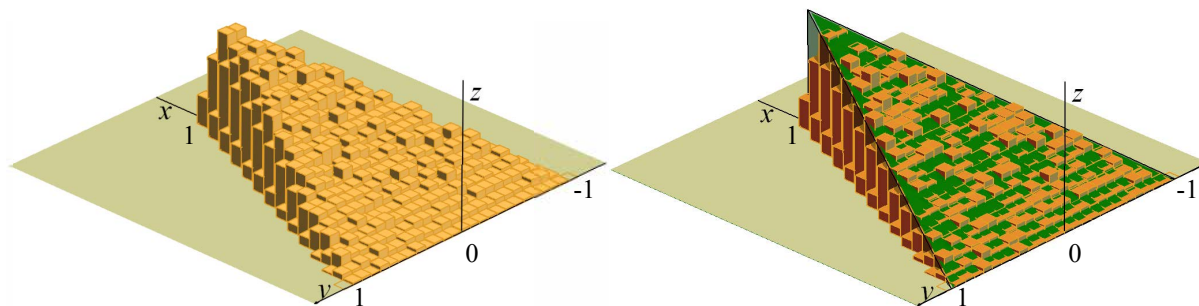
SLIKA 7.15: Vzorec slučajnih vektorjev  $\{U_1, U_2\}$  in  $\{X, Y\}$

Na naslednji sliki 7.16 prikazujemo stolpčne diagrame vzorcev slučajnih spremenljivk  $X$  in  $Y$  ter ju primerjamo s teoretičnimi gostotama porazdelitve. Očitno je, da smo obe slučajni spremenljivki generirali skladno z zahtevano porazdelitvijo.



SLIKA 7.16: Vzorec slučajnih spremenljivk in primerjava z njunima gostotama verjetnosti

Na zadnji sliki 7.17 prikazujemo še stolpčni diagram za vzorec slučajnega vektorja  $\{X, Y\}$  in ga primerjamo s teoretično gostoto verjetnosti. Tudi tukaj lahko ugotovimo, da se generirani vzorec sklada s predpostavljeno porazdelitvijo.



SLIKA 7.17: Vzorec slučajnega vektorja  $\{X, Y\}$  in primerjava z njegovo gostoto verjetnosti

**Primer 7.5** *Generiranje po metodi sprejema in zavrnitve. Z uporabo metode sprejema in zavrnitve generirajmo vzorec slučajnega vektorja  $\{X, Y\}$  iz primera 7.4.*

Pri uporabi te metode moramo generirati tri neodvisne enakomerno porazdeljene slučajne spremenljivke. Glede na zalogo vrednosti slučajnega vektorja  $\{X, Y\}$  je prva slučajna spremenljivka  $U_1$  enakomerno porazdeljena od 0 do 1, druga  $U_2$  je enakomerno porazdeljena od  $-1$  do 1, glede na največjo vrednost gostote verjetnosti pa je tretja slučajna spremenljivka  $U_3$  porazdeljena enakomerno od 0 do 3. Na sliki 7.18 prikazujemo množico enakomerno porazdeljenih točk v kvadru. Zeleno so obarvane tiste točke, ki jih sprejmemo. Rezultat generiranja po metodi sprejema in zavrnitve je prikazan na desni strani iste slike.

V preglednici 7.7 prikazujemo določitev vzorca slučajnega vektorja po metodi sprejema in zavrnitve. Prikazujemo prvih deset slučajno izbranih točk. Za prvo točko  $\{0.367, -0.161, 0.736\}$  lahko ugotovimo, da je gostota verjetnosti:

$$u_{31} = 0.736 < f_{XY}(u_{11}, u_{21}) = f_{XY}(0.367, -0.161) = 1.101,$$

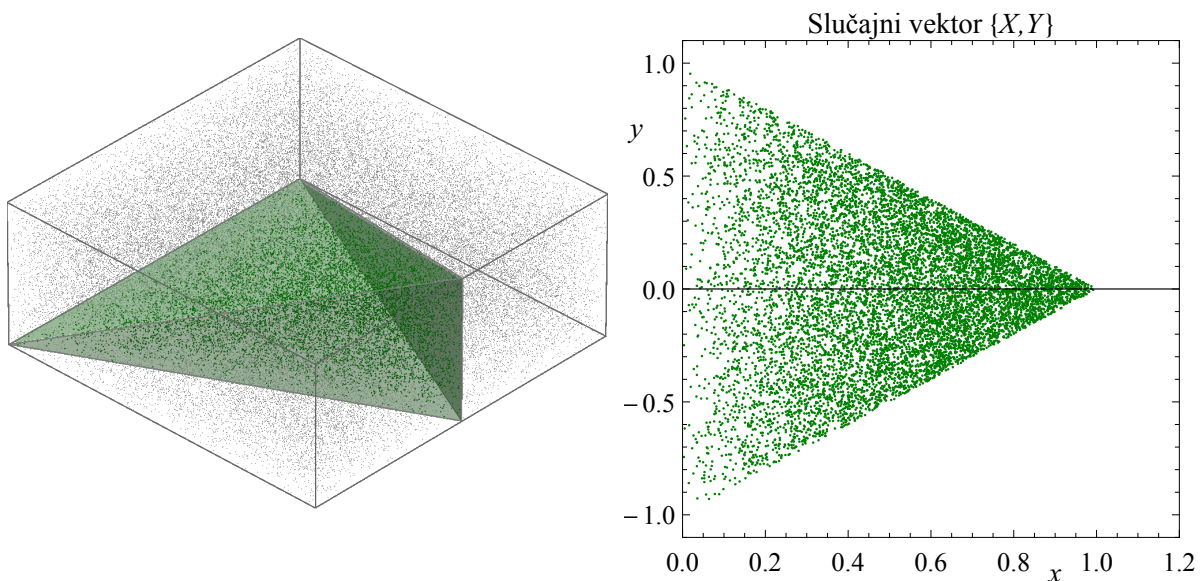
zato točko sprejmemo. V vzorec dodamo par  $\{0.367, -0.161\}$ . Naslednjo točko  $\{0.891, 0.384, 2.640\}$  moramo žal zavrniti, saj je za koordinati  $\{0.891, 0.384\}$  gostota verjetnosti enaka nič (glej enačbo (7.27)):

$$u_{32} = 2.640 > f_{XY}(u_{12}, u_{22}) = f_{XY}(0.891, 0.384) = 0.$$

Vidimo, da lahko od prvih desetih generiranih točk sprejmemo le dve (prvo in šesto), druge pa moramo zavrniti. Delež točk, ki jih bomo sprejeli, običajno lahko ocenimo s primerjavo prostornin: prostornina kvadra, v katerem generiramo točke, je 6, prostornina piramide, ki ustreza območju sprejema, pa je enaka 1. Zato lahko pričakujemo, da bo delež sprejetih točk okoli  $1/6$ . Za vzorec 10000 elementov smo v naši simulaciji generirali 60066 točk, kar potrjuje našo oceno.

PREGLEDNICA 7.7: Generiranje slučajnega vektorja po metodi sprejema in zavrnitve

Generirane vrednosti										
$u_{1i}$	0.367	0.891	0.527	0.678	0.397	0.542	0.603	0.894	0.353	0.370
$u_{2i}$	-0.161	0.384	-0.800	-0.575	0.763	-0.029	0.562	0.729	-0.051	-0.910
$u_{3i}$	0.736	2.640	0.563	1.830	2.204	1.457	2.158	0.784	2.670	1.783
$f_{XY}$	1.101	0	0	0	0	1.627	0	0	1.058	0
$x_i$	0.367	-	-	-	-	0.542	-	-	-	-
$y_i$	-0.161	-	-	-	-	-0.029	-	-	-	-

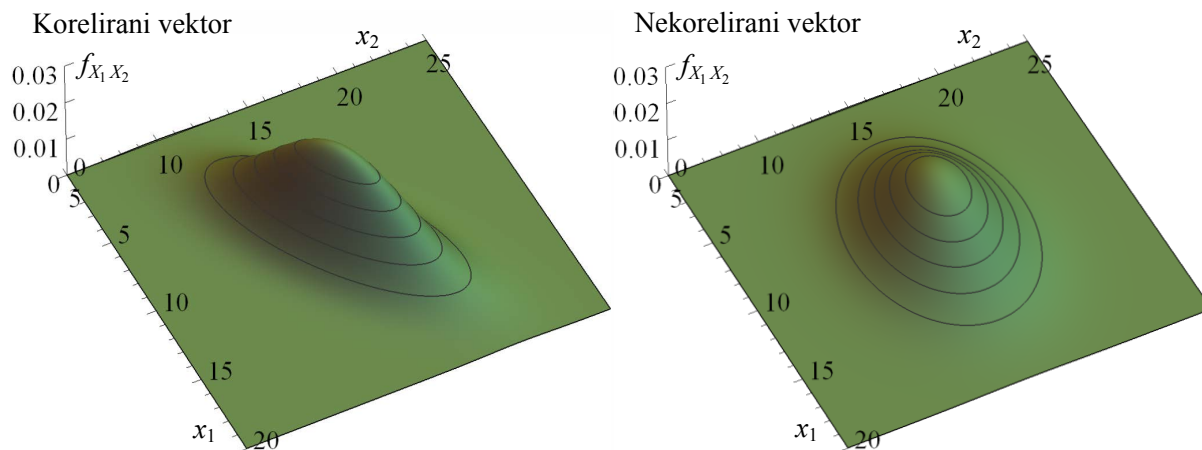


SLIKA 7.18: Generiranje množice enakomerno razporejenih točk po kvadru

**Primer 7.6** *Generiranje koreliranih normalno porazdeljenih slučajnih vektorjev. Z uporabo opisane metode in razcepa Choleskega generirajmo slučajni vektor  $\mathbf{X}$  z naslednjimi lastnostmi:*

$$\mathbf{m}_{\mathbf{X}} = \begin{bmatrix} m_{X_1} \\ m_{X_2} \end{bmatrix} = \begin{bmatrix} 10 \\ 15 \end{bmatrix}, \quad \mathbf{C}_{\mathbf{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_1 X_2} & \sigma_{X_2}^2 \end{bmatrix} = \begin{bmatrix} 12 & 6 \\ 6 & 6 \end{bmatrix}. \quad (7.34)$$

Gostoto verjetnosti slučajnega vektorja  $\mathbf{X}$  in podobnega vektorja z neodvisnimi slučajnimi spremenljivkami ( $\sigma_{X_1 X_2} = 0$ ) prikazujemo na sliki 7.19.



SLIKA 7.19: Gostota verjetnosti vektorja  $\mathbf{X}$  koreliranih in nekoreliranih slučajnih spremenljivk

Pred uporabo enačbe (7.23) moramo določiti matriko  $\mathbf{H}$ , ki jo dobimo z razcepom Choleskega matrike  $\mathbf{C}_X$ . Matrika  $\mathbf{H}$  je:

$$\mathbf{H} = \begin{bmatrix} 2\sqrt{3} & 0 \\ \sqrt{3} & \sqrt{3} \end{bmatrix} = \begin{bmatrix} 3.464 & 0 \\ 1.732 & 1.732 \end{bmatrix}. \quad (7.35)$$

Nato generiramo vzorec neodvisnih standardizirano normalno porazdeljenih slučajnih spremenljivk  $Y_1$  in  $Y_2$ . Prvih deset slučajno generiranih vrednosti slučajnih vektorjev  $\mathbf{Y}$  in  $\mathbf{X}$  prikazujemo v preglednici 7.8. Na primer, za prvi element vzorca izračunamo:

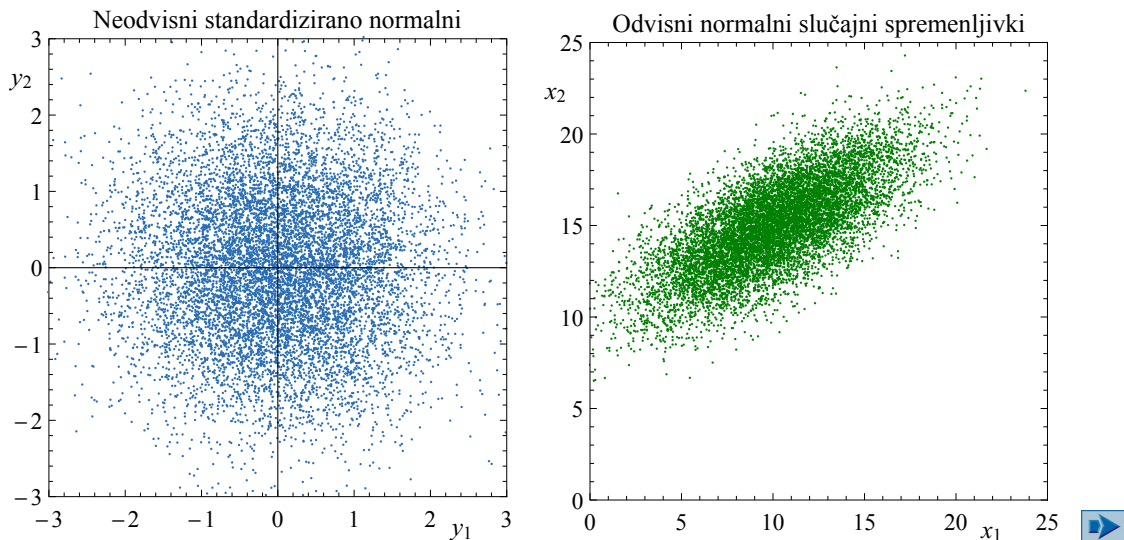
$$\mathbf{X}_1 = \begin{bmatrix} 10 \\ 15 \end{bmatrix} + \begin{bmatrix} 3.464 & 0 \\ 1.732 & 1.732 \end{bmatrix} \begin{bmatrix} 0.548 \\ 1.287 \end{bmatrix} = \begin{bmatrix} 11.90 \\ 18.19 \end{bmatrix}. \quad (7.36)$$

Na podoben način izračunamo tudi vse druge elemente vzorca slučajnega vektorja  $\mathbf{X}$ .

PREGLEDNICA 7.8: Generiranje odvisnih normalno porazdeljenih slučajnih spremenljivk

	Generirane vrednosti									
$y_{1i}$	0.548	-0.416	-1.664	0.392	-1.857	1.984	0.912	1.278	-1.019	-0.738
$y_{2i}$	1.287	0.653	-1.178	1.575	0.763	-1.082	0.751	0.738	1.813	-0.245
$x_{1i}$	11.90	8.56	4.23	11.36	3.57	16.87	13.16	14.43	6.47	7.44
$x_{2i}$	18.19	15.41	10.08	18.41	13.11	16.56	17.88	18.49	16.38	13.30

Na sliki 7.20 prikazujemo vzorca slučajnega vektorja neodvisnih oziroma nekoreliranih standardizirano normalno porazdeljenih spremenljivk  $Y_i$  ter odvisnih oziroma koreliranih normalno porazdeljenih slučajnih spremenljivk  $X_i$ .

SLIKA 7.20: Vzorca slučajnih vektorjev  $\mathbf{Y}$  in  $\mathbf{X}$ 

Če bi želeli generirati neodvisne normalno porazdeljene slučajne spremenljivke, bi uporabili enačbo:

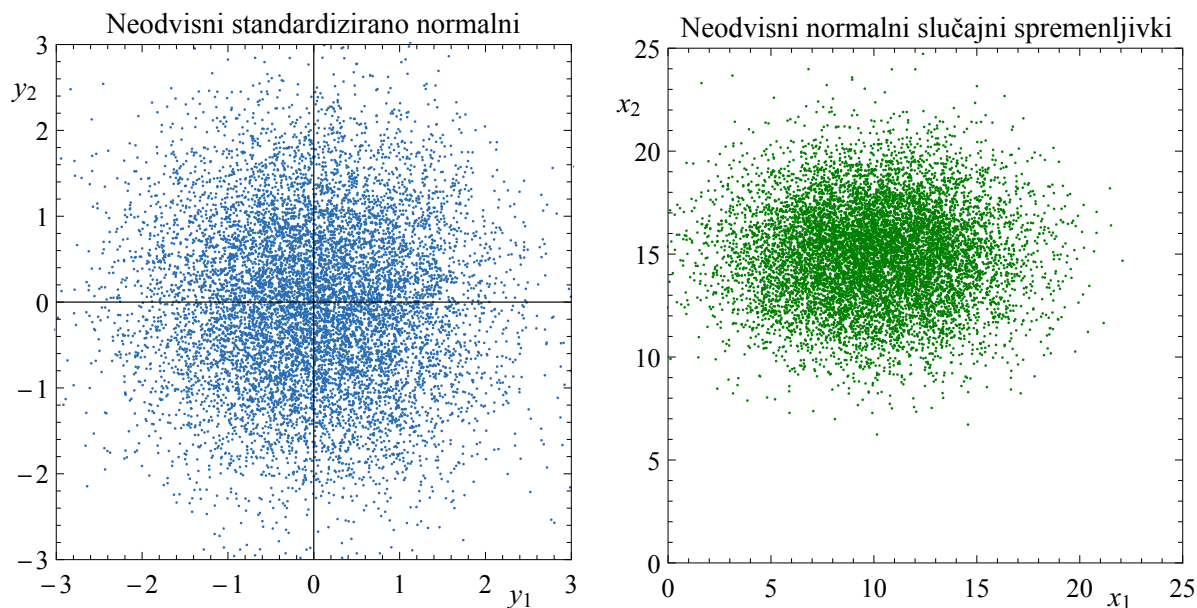
$$x_{ki} = m_{X_k} + \sigma_{X_k} y_{ki}. \quad (7.37)$$

Prvih deset naključno izbranih elementov vzorca prikazujemo v preglednici 7.9. S primerjavo med preglednicama 7.8 in 7.9 vidimo, da se vrednosti vzorca slučajne spremenljivke  $X_1$  ne razlikujejo, do razlik pride pri generiranju druge slučajne spremenljivke  $X_2$ , ki je v zadnjem primeru neodvisna od slučajne spremenljivke  $X_1$ .

PREGLEDNICA 7.9: Generiranje neodvisnih normalno porazdeljenih slučajnih spremenljivk

	Generirane vrednosti									
$y_{1i}$	0.548	-0.416	-1.664	0.392	-1.857	1.984	0.912	1.278	-1.019	-0.738
$y_{2i}$	1.287	0.653	-1.178	1.575	0.763	-1.082	0.751	0.738	1.813	-0.245
$x_{1i}$	11.90	8.56	4.23	11.36	3.57	16.87	13.16	14.43	6.47	7.44
$x_{2i}$	18.15	16.60	12.11	18.86	16.87	12.35	16.84	16.81	19.44	14.40

Na sliki 7.21 prikazujemo vzorca slučajnega vektorja neodvisnih oziroma nekoreliranih standardizirano normalno porazdeljenih spremenljivk  $Y_i$  ter neodvisnih poljubno normalno porazdeljenih slučajnih spremenljivk  $X_i$ . V primerjavi med slikama 7.20 in 7.21 vidimo, da sta pri koreliranih slučajnih spremenljivkah osi oblaka točk poševni, medtem ko sta pri nekoreliranih osi oblaka točk vzporedni s koordinatnima osema.

SLIKA 7.21: Vzorca slučajnih vektorjev  $\mathbf{Y}$  in  $\mathbf{X}$ 

**Primer 7.7** Generiranje koreliranih normalno porazdeljenih slučajnih vektorjev. Z uporabo opisane metode in razcepa Choleskega generirajmo slučajni vektor štirih slučajnih spremenljivk  $\mathbf{X}$  z naslednjimi lastnostmi:

$$\mathbf{m}_{\mathbf{X}} = \begin{bmatrix} m_{X_1} \\ m_{X_2} \\ m_{X_3} \\ m_{X_4} \end{bmatrix} = \begin{bmatrix} 10 \\ 15 \\ 10 \\ 20 \end{bmatrix}, \quad \mathbf{C}_{\mathbf{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \sigma_{X_1 X_3} & \sigma_{X_1 X_4} \\ \sigma_{X_1 X_2} & \sigma_{X_2}^2 & \sigma_{X_2 X_3} & \sigma_{X_2 X_4} \\ \sigma_{X_1 X_3} & \sigma_{X_2 X_3} & \sigma_{X_3}^2 & \sigma_{X_3 X_4} \\ \sigma_{X_1 X_4} & \sigma_{X_2 X_4} & \sigma_{X_3 X_4} & \sigma_{X_4}^2 \end{bmatrix} = \begin{bmatrix} 12 & 6 & 4 & 2 \\ 6 & 6 & 4 & 4 \\ 4 & 4 & 10 & 4 \\ 2 & 4 & 4 & 8 \end{bmatrix}.$$

Najprej moramo določiti matriko  $\mathbf{H}$ , ki jo dobimo z razcepom Choleskega matrike  $\mathbf{C}_{\mathbf{X}}$ :

$$\mathbf{H} = \begin{bmatrix} 2\sqrt{3} & 0 & 0 & 0 \\ \sqrt{3} & \sqrt{3} & 0 & 0 \\ \frac{2}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \sqrt{\frac{22}{3}} & 0 \\ \frac{1}{\sqrt{3}} & \sqrt{3} & 2\sqrt{\frac{2}{33}} & \sqrt{\frac{146}{33}} \end{bmatrix} = \begin{bmatrix} 3.464 & 0 & 0 & 0 \\ 1.732 & 1.732 & 0 & 0 \\ 1.155 & 1.155 & 2.708 & 0 \\ 0.577 & 1.732 & 0.492 & 2.103 \end{bmatrix}. \quad (7.38)$$

V preglednici 7.10 prikazujemo prvih osem naključno izbranih elementov vzorcev slučajnih vektorjev  $\mathbf{Y}$  in  $\mathbf{X}$ . Na primer, tretji element vzorca slučajnega vektorja  $\mathbf{X}$  določimo takole:

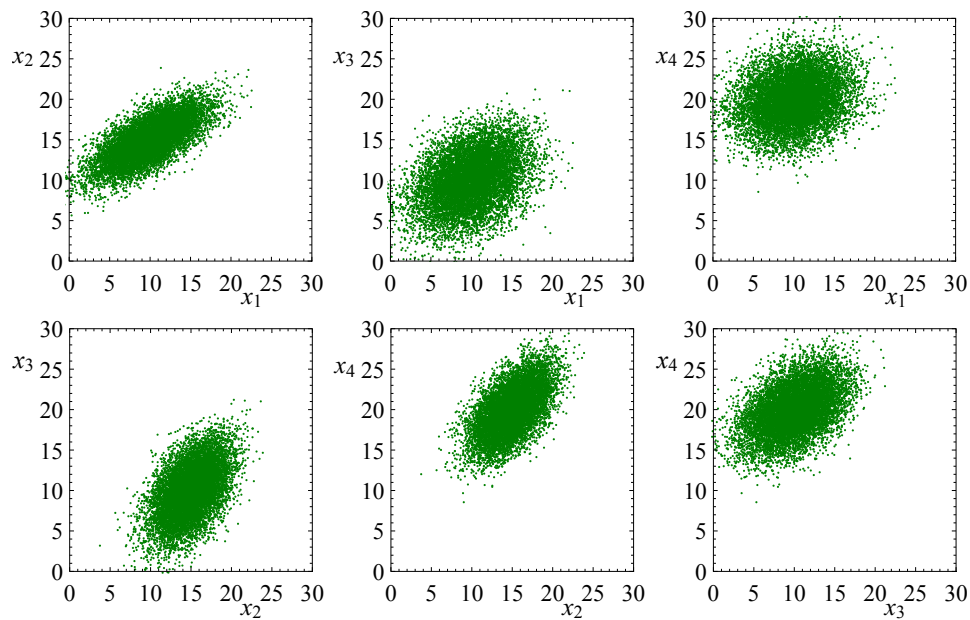
$$\mathbf{X}_3 = \begin{bmatrix} 10 \\ 15 \\ 10 \\ 20 \end{bmatrix} + \begin{bmatrix} 3.464 & 0 & 0 & 0 \\ 1.732 & 1.732 & 0 & 0 \\ 1.155 & 1.155 & 2.708 & 0 \\ 0.577 & 1.732 & 0.492 & 2.103 \end{bmatrix} \begin{bmatrix} 0.723 \\ 0.200 \\ -1.182 \\ -1.784 \end{bmatrix} = \begin{bmatrix} 12.50 \\ 16.60 \\ 7.87 \\ 16.43 \end{bmatrix}. \quad (7.39)$$



PREGLEDNICA 7.10: Generiranje odvisnih normalno porazdeljenih slučajnih spremenljivk

Generirane vrednosti								
$y_{1i}$	-0.644	0.555	0.723	-1.389	0.762	-1.429	1.610	0.627
$y_{2i}$	0.286	0.765	0.200	-0.009	-0.244	-0.101	-1.416	0.739
$y_{3i}$	1.582	-0.668	-1.182	1.026	-0.594	-1.274	-1.671	2.605
$y_{4i}$	0.069	-0.960	-1.784	-1.069	-0.532	-0.938	-0.914	-0.270
$x_{1i}$	7.77	11.92	12.50	5.19	12.64	5.05	15.58	12.17
$x_{2i}$	14.38	17.29	16.60	12.58	15.90	12.35	15.34	17.37
$x_{3i}$	13.87	9.72	7.87	11.16	8.99	4.78	5.70	18.63
$x_{4i}$	21.05	19.30	16.43	17.44	18.61	16.40	15.73	22.36

Na sliki 7.22 prikazujemo vzorec slučajnega vektorja  $\mathbf{X}$ . Ker je vektor štirirazsežen, grafično prikazujemo posamezne dele vektorja in sicer po parih slučajnih spremenljivk. Iz slike lahko vidimo, da je korelacija med  $X_1$  in  $X_2$  močnejša kot korelacija med  $X_1$  in  $X_3$  ali  $X_1$  in  $X_4$ . To se seveda odraža že iz podatkov, saj je korelacijski koeficient  $\rho_{X_1 X_2} = 0.707 > \rho_{X_1 X_3} = 0.365 > \rho_{X_1 X_4} = 0.204$ .

SLIKA 7.22: Vzorec slučajnega vektorja  $\mathbf{X}$





## 8 Literatura

1. R. Andričević, H. Gotovac, I. Ljubenković, Geostatistika, umijeće prostorne analize, Građevinsko-ArHITEKTOSNI Fakultet Sveučilišta u Splitu, Split, 2007.
2. J.-P. Chilès, P. Delfiner, Geostatistics, Modeling Spatial Uncertainty, John Wiley & Sons, New York, 1999.
3. N. Christou, Universal kriging, Class notes, University of California, Los Angeles, 2014.
4. N. E. Cressie, Statistics for Spatial Data, John Wiley & Sons, New York, 1993.
5. T. Hengl, A Practical Guide to Geostatistical Mapping, Office for Official Publications of the European Communities, Luxembourg, 2009.
6. K. Krivoruchko, Spatial Statistical Data Analysis for GIS Users, Esri Press, Redlands, California, 2011.
7. A. Paez, D. M. Scott, Spatial statistics for urban analysis: A review of techniques with examples, *GeoJournal*, let. 61, str. 53–67, 2004.
8. R. Webster, O. Atteia, J.P. Dubois, Coregionalization of trace metals in the soil in the Swiss Jura, *European Journal of Soil Science*, let. 45, str. 205–218, 1994.



# Stvarno kazalo

- aproksimacija variograma, 31
- Box–Müllerjeva metoda, 91
- eksponentna porazdelitev, 90
- enakomerna porazdelitev, 89, 90
- Excel, 91
- Fréchetova porazdelitev, 90
- Gearyjevo razmerje, 36, 40
- generiranje vzorca slučajne spremenljivke, 89
  - Box–Müllerjeva metoda, 91
  - inverzna metoda, 89, 98, 100
  - metoda sprejema/zavrnitve, 92, 93
  - normalna porazdelitev, 91, 101
- generiranje vzorca slučajnega vektorja, 98
- geostatistični podatki, 4, 23
  - Jura, 5, 24
- gostota verjetnosti, 10
- Gumbelova porazdelitev, 90
- indikatorji, 33
- koeficient simetričnosti, 12
- korelacijska funkcija, 14
- korelacijski koeficient, 13
- kovariančna funkcija, 14, 15
  - ocena, 29
- kovariančna funkcija z indikatorji, 34
- kovarianca, 13
- kovariogram, 20
- križna kovariančna funkcija, 20
- krigiranje, 45
  - običajno, 50, 57
  - preprosto, 46, 56
  - s trendom, 51, 58
- krigiranje s trendom, 51
- kurtosis, 12
- Lagrangev množitelj, 50, 52
- linearna regresija, 61
- Mathematica, 91
- metoda kvadratov, 70
- metoda Monte Carlo, 87
- metoda najbližjega soseda, 70, 77
- metoda najmanjših kvadratov, 62, 63
- momenti slučajnih spremenljivk, 12
- Moranov indeks, 36, 37
- neodvisnost, 11
- normalna porazdelitev, 91, 107
- običajno krigiranje, 50
- odstopanje, 61
- osamelec, 62
- podatki v mreži, 6
- Poissonova porazdelitev, 71
- porazdelitev
  - eksponentna, 90
  - enakomerna, 89, 90
  - Fréchetova, 90
  - Gumbelova, 90
  - normalna, 91, 107
  - Weibullova, 90
- porazdelitvena funkcija, 9

- pogojna, 11
  - robna, 11
- preizkušanje domnev, 35
- preizkus  $\chi^2$ , 72
- preprosto krigranje, 46
- pričakovana vrednost, 12
- prostorski vzorci, 8, 70
- prostorsko utežena regresija, 61
- razcep Choleskega, 102, 107
- razsevni graf, 28
- regresija
  - linearna, 61
- rezidual, 61
- robustna metoda, 62
- semivariogram, 15
- slučajna funkcija, 13
- slučajna spremenljivka, 9
- slučajni proces, 13
- slučajni vektor, 10
- slučajno polje, 13
  - gostota verjetnosti, 14, 19
  - korelacijska funkcija, 14
  - kovariančna funkcija, 14
  - porazdelitvena funkcija, 13
  - srednja vrednost, 14, 19
  - varianca, 14, 19
  - variogram, 15
- srednja vrednost slučajnega polja, 14, 19
- stacionarno polje
  - kovariančna funkcija, 15
  - variogram, 15
- stacionarnost, 15
- testiranje hipotez, 35
- varianca, 12
- varianca slučajnega polja, 14, 19
- variogram, 15
  - doseg, 17
  - empirična funkcija, 31
  - ocena, 30
  - prag, 17
  - učinek zrna, 17
- variogram z indikatorji, 34
- vektorsko slučajno polje
  - kovariogram, 20
  - križna kovariančna funkcija, 20
  - križni variogram, 20
  - porazdelitvena funkcija, 19
- verjetnostna funkcija, 10
- vezani ekstrem, 50, 52
- vzorčna kovariančna funkcija, 29
- vzorčni variogram, 30
- Weibullova porazdelitev, 90
- zaloga vrednosti, 9