SIF 5064 Romlig Statistikk, Øving 1: Gaussiske felt og Kriging

Innledning Denne øvingen handler om Gaussiske felt, kriging, kovariansfunksjoner (variogrammer) m.m. Øvingen er laget slik at det den benytter spatial biblioteket i Splus Kapittel 15 i Venables & Ripley: Modern Applied Statistics with Splus, og notatet "S-plus ved IMF, Del 3: Funksjoner, bibliotek og internett", samt web linker fra hjemme siden kan være nyttige. "spatial" laster du inn inn i S-plus ved å skrive `library(spatial)`.

Oppgave 1: 1D Gaussisk felt

(a) Lag plott som viser hvordan eksponensiell og 2.ordens eksponensiell kovariansfunksjon ser ut for forskjellige valg av "sill", og "range". Gjør parameter valg som gjør det mulig å sammenligne de forskjellige korrelasjons funksjonene. Gi en tolkning av parameterene.

Hint: Benytt funksjonene `gaucov` og `expcov` fra biblioteket "spatial" i splus.

Hva er den teoretiske sammenhengen mellom kovariansfunksjon, og variogram ? Når vil du bruke den ene parametriseringen fremfor den andre.

(b) Trekk et normalisert 1D gaussisk felt diskretisert på et grid av størrelse 50, ved å bruke multigaussisk fordeling. Bruk forskjellige covarians funksjoner. Diskuter innvirkningen av parameter valgene på simuleringen.

Hint: Under "Examples" på hjelpeteksten for `rnorm` finner du en funksjon som du kan benytte.

(c) Velg én av realisasjonene med 1. ordens eksponensiell kovarians fra oppgave (b). Bruk verdiene i node 10,25 og 30 som observasjoner. Finn utrykk for betinget fordeling for nodene som ikke er observert gitt de tre observasjonene. Plot betinget forventning og betinget varians (eller standard avvik) i dette eksempelet. Plot også realisasjonen du valgte. Diskuter resultatene.

(d) Lag 25 realisasjoner fra det Gaussiske feltet betinget på observasjonene i (c). Plott de betingede realisasjonene. Sammenlign med resultatene i (b) og (c) og kommenter.

(e) Beregn gjennomsnitt og empirisk varians for de 25 realisasjoner punktvis dvs., for hver av grid nodene. Plot og kommenter resultatet.

(f) Definer en terskel verdi på 1.0. En ønsker å vite hvor stort arealet over terskel verdien og under den utvalgte realisasjonen er. For å predikere dette arealet, betinget på de tre observasjonen kan en tenke seg to prediktorer.

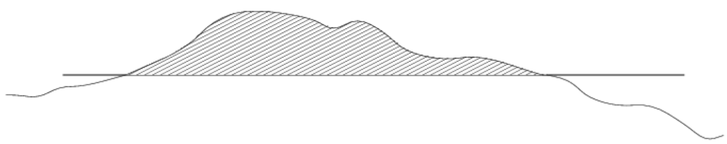
1. arealet mellom betinget forventning fra (c) og terskel verdien.
2. gjennomsnittet av arealet over terskelverdien, og under gjentatte simuleringer fra den betingede fordelingen i (d).

Benytt begge prediktorene for å gi en prediksjon. Hvilken prediktor gir størst verdi ? Kan du gi en teoretisk begrunnelse for dette ? Hvordan vil du presentere usikkerheten i prediktoren for strategi 2 ? Beregn det sanne arealet over terskel verdien og under din utvalgte realisasjon fra (b). Sammenlign dette estimatet med prediksjonen og usikkerheten i strategi 2.

Tips: Splus syntaxen `sum(y[y>0])` kan forenkle programmet.

(g) Repeter oppgave (c)-(f) med 2.ordens eksponensiell kovarians funksjon. Kan noen av forskjellene du finner tilskrives forskjell i kovarians funksjoner.

Merk: Dette eksempelet er en skjematisk presentasjon av "oil in placeproblemet, vi estimerer berg volumet under en geologisk horisont og over olje vann kontakten.



Figur 1: Figuren fremstiller situasjonen over.

Oppgave 2: Trendflater, kriging og plotting i S-plus

Funksjoner og data som brukes i denne oppgaven finnes i biblioteket `spatial`. Data settet `topo` inneholder 52 målinger av høyden over havet (i fot) innenfor et 310×310 fot kvadrat, omregnet til 50 fots enheter.

(a) Vi vil først se på trendflater, polynomiske regresjonsflater av formen

$$f((x,y)) = \sum_{r+s\le p} a_{rs} x^r y^s$$

hvor parameteren p er polynomets orden. Tilpass trendflater med orden p= 2,3,4,6. Bruk splus funksjonen `surf.ls`. Plot de forskjellige flatene med plot funksjonene: ;`contour`, `image` and `persp` inkluder data punkter i disse plottene. Funksjonen `trmat` kan være nyttig. Diskuter fordeler og ulemper med de forskjellige plottings rutinene. Kommenter resultatene. Hvor mange parametere må estimeres for en 6. ordens flate? Hvilke svakheter eller ulemper medfører dette?

(b) Ved universell kriging tilpasses en flate av formen

$$Z(x) = f(x)^T \beta + \epsilon(x)$$

der x er 2D romlig referanse, $f(x)$ er gitte funksjoner, β er ukjente koeffisienter og $\epsilon(x)$ er et Gaussisk felt med forventning 0 og romlig kovariansfunksjon:

$$\mathrm{cov}\{Z(x), Z(y)\} = \mathrm{cov}\{\epsilon(x), \epsilon(y)\}$$

Prøv universell kriging, benytt funksjonene `surf.gls` med varians 1, 1.ordens eksponensiell kovarains med parameter $d = 0.7$. La $f(x)$ være polynomer som i (a), og benytt forskjellige verdier for p,dvs polynomer av forskjellig orden. Benytt spesielt $p = 0$. Plot og kommenter resultatene. Funksjonen `prmat` kan her være nyttig. Prøv også forskjellige verdier av d.

Endre til 2. ordens eksponensiell kovarians funksjon og gjør beregningene paa nytt. Plot og kommenter.

(c) Plot den estimerte trendflaten fra kriging modellene i (b), benytt `trmat`.

Hva er forskjellen mellom dette estimatet og det vi fikk i a, dvs. hva er forskjellen mellom funksjonene `surf.gls` og `surf.ls`.

Hva er forskjellen på trendflatene og krigingflatene?

(d) Plot prediksjonsstandardavvik for krigings flaten fra (b), Kommenter resultatene. Benytt splus funksjonen `semat`.

Oppgave 3: Simulering og estimering

Simulering av Gaussiske felt i dimensjon 2 eller høyere blir fort ganske regnetungt. I denne oppgaven skal vi bruke et ferdig program for simulering av 2D Gaussiske felt. Programmet simulerer et Gaussisk felt med forventning null, varians 1 og eksponensiell kovariansfunksjon $c(r) = \exp(-3r/d)$ der r er avstanden mellom punktene. Input til programmet er gridstørrelse og parameteren d i kovariansfunksjonen. **NB:** Merk 3-tallet! I S-plus er eksponensiell kovariansfunksjon definert som $c(r) = \exp(-r/d)$. Programmet startes ved å skrive `/local/kurs/romstat/simGauss $RANDOM`. Det kryptiske `$RANDOM` gir frø (seed) til randomgeneratoren. Hvis det utelates, brukes en standard verdi. Programmet skriver ut resultatet til en fil med en matrise av verdiene til det Gaussiske feltet i tilhørende gridnoder.

(a) Simuler et Gaussisk felt med gridstørrelse: 50×50 og parameter $d = 20$. Prøv et par andre parametervalg også. Plot realisasjonene i splus. Benytt `sim <- read.table( ''filnavn'' )` for å lese inn datafilen. Ta vare på realisasjonene, de skal brukes videre i oppgaven.

(b) "Observer" det Gaussiske feltet i 36 punkter tilfeldig fordelt utover i gridet. Bruk funksjonen `correlogram` til å estimere den empiriske korrelasjons funksjonen. Plott estimatet og sammenlign med den sanne korrelasjons funksjonen.

(c) Estimer korrelasjons funksjonen ved å bruke 9, 25, 49, 64 og 100 punkter tilfeldig fordelt utover i gridet. Omtrent hvor mange observasjoner trenger du for å et tilfredsstillende resultat?

(d) La observasjonspunktene være systematisk plassert på et grid. Hvordan fungerer dette i forhold til å la observasjonenspunktene være tilfeldig fordelt? (Bruk tilsvarende antall punkter som du gjorde i forrige delspørsmål.)

(e) Hvilke problemer kan du se i forbindelse med estimering av kovariansfunksjon dersom funksjonsformen er ukjent?