SIF 5064 Romlig Statistikk,
Øving 2: Punktprosesser

Oppgave 1: Inhomogen Poisson prosess

Vi skal i denne oppgaven se på en inhomogen Poisson prosess med intensitetsfunksjon

$$\lambda_\theta(x, y) = \mu \exp(-\alpha x - \beta y)$$

hvor $\theta = (\mu, \alpha, \beta)$.

a. Beregn forventningsverdi og varians til antall punkter i denne modellen på et enhetskvadrat.

b. Simuler en inhomogen Poisson prosess med den gitte intensitetsfunksjonen og $\theta = (200, 2.0, 1.0)$, betinget på antall punkter. Prøv med forskjellig antall,f.eks $n = (20, 50, 100)$.

Splus tips : Bruk `par(pty=''s'')` for å plotte figurene på et kvadratisk område.

c. Utvid algoritmen i b, slik at du kan simulere ubetinget på antall punkter,og utfør simuleringen.

d. En alternativ algoritme for ubetinget simulering, er basert på p-tynning. For simulering på enhets kvadratet, inneholder algoritmen følgende skritt:

1. Sett
$$\lambda^* = \max_{(x,y)\in[0,1]^2} \lambda_\theta(x, y)$$
2. Simuler en homogen poisson prosess med intensitet λ^*.
3. Behold et simulert punkt (x_n, y_n) med sannsynlighet
$$p = \frac{\lambda_\theta(x_n, y_n)}{\lambda^*}$$

Implementer denne algoritmen også, og benytt denne til simulering. Hva er forskjellen på denne algoritmen og den ifra c.. Generellt sett hvilke fordeler/ulemper ser du ved denne algoritmen sammenlignet med den i c ?

e. Sett opp ligningene som bestemmer maximum likelihood estimatorene for parametrene $\theta = (\mu, \alpha, \beta)$ i modellen gitt over. Likelihood-funksjonen er gitt ved

$$L(x_1, y_1 \ldots, x_n, y_n; \theta) = \lambda_\theta(x_1, y_1) \ldots \lambda_\theta(x_n, y_n) \exp[-\int_0^1 \int_0^1 \lambda_\theta(x, y) dx dy]$$

når vi er på enhetskvadratet.

f. Velg et punktmønster simulert ubetinget på antall punkter fra modellen med parameterverdier gitt før i oppgaven, og finn maximum likelihood estimatene for parametrene basert på disse dataene. Plott den estimerte intensiteten med funksjonen `image`.

g. Kopier filene intest.s og integral.o fra katalogen `/local/kurs/romstat/` Dette er kode for å begregne ikke-parametrisk estimat av intensiteten basert på kjerneestimering, se Cressie (1993) side 600. Finn ut hvordan funksjonen intest.s virker, og bruk den til å estimere intensiteten for det samme datasettet som du brukte til maximum likelihood estimering. Test gjerne ulike glattingsparametre. Plott intensiteten, og sammenlign de to estimatene. (NB Vær påpasselig med farge skalaene dersom du plotter tethetene vha `image`.)

Oppgave 2: Neyman-Scott prosesser

I denne oppgaven skal vi se på Neyman-Scott prosesser der antall barn er Poisson fordelt. Som fordeling for barnas posisjon relativt til moren kan du velge en symmetrisk normalfordeling (Thomas prosessen) eller uniform fordeling i en sirkel med radius R (Matern cluster prosess). Du kan lese om Neyman-Scott prosesser i Stoyan og Stoyan (1994) kapittel 16.2.

a. Simuler en Neyman-Scott prosess på enhetskvadratet ubetinget på antall punkter. Hva vil du gjøre for å unngå randeffekter? Varier parametrene i modellen og se hvordan det virker inn på punktmønstrene.

b.En Neyman-Scott prosess med Poisson fordelt antall barn er ekvivalent med en Cox prosess med intensitetsfunksjon

$$\Lambda(r) = \omega \sum_{i=1}^{\infty} f(r - s_i)$$

hvor ω er forventet antall barn per mor, f er fordelingen til barnas posisjon relativt til moren og s_i er posisjonen til mor nummer i.

En Cox prosess er en "inhomogen poisson prosess med stokastisk intensitet". Hva er stokastisk i $\Lambda(r)$? Hvordan vil du simulere $\Lambda(r)$ betinget på antall (barn) punkter ? Hvordan vil du gå fram dersom du skal simulere en Neyman-Scott prosess betinget på antall punkter? (Du trenger ikke å utføre selve simuleringen)

c. Beregn L-funksjonen for noen simulerte punktmønstre (se oppgave 3) og sammenlign med teoretisk L-funksjon. NB! Benytt `ppregion` til å initialisere punkt prosess området. For Thomas-prosessen har vi

$$K(t) = \pi t^2 + \rho^{-1}(1 - \exp(-t^2/4\sigma^2))$$

hvor ρ er intensiteten til mor-prosessen, og σ^2 er variansen i fordelingen til barnas posisjon relativt til moren. For Matern cluster-prosessen har vi

$$K(t) = \pi t^2 + \rho^{-1} \times \begin{cases} 2 + \pi^{-1}[(8z^2 - 4)\arccos z - 2\arcsin z & \\ +4z\sqrt{(1-z^2)^3} - 6z\sqrt{1-z^2}] & (t \le 2R) \\ 1 & (t > 2R) \end{cases}$$

hvor $z = t/2R$ og notasjonen ellers er som over.

Oppgave 3: Strauss prosessen

Last inn biblioteket "spatial" ved å bruke kommandoen `library(spatial)` Her finnes blant annet funksjon for beregning av L-funksjon ($L(t) = \sqrt{K(t)/\pi}$) med "envelopes" og rutine for å simulere Strauss prosesser betinget på antall punkter. Relativt til en binomisk prosess er den unormaliserte tettheten til Strausssprosessen her definert som $c^{s(x)}$ der $s(x)$ er antall nabopar med avstand mindre enn en konstant r og $c > 0$.

a. Simuler Strausssprosess med 50 punkter, $r = 0.1$ og $c = 0.01$, $c = 0.5$, $c = 1.0$. Sammenlign punktmønstrene. Pass på at alle punktmønstrene plottes på samme skala. Beregn L-funksjon med "enveloper" basert på 99 simuleringer fra en Binominal prosess (Poisson prosess betinget på antall punkter) for de tre første tilfellene. Hva forteller plottene?(NB! benytt `ppregion()`) Hva skjer dersom $c > 1$ hvis 1) vi betinger på antall punkter 2) dersom vi ikke betinger på antall punkter.

b. Beskriv hvordan Metropolis-Hastings algoritme kan brukes til å simulere fra en Strausssprosess betinget på antall punkter. (Du trenger ikke gjøre det i praksis.)

c. Beskriv korthvordan du ville ha simulert en ubetinget Strauss prosess ved hjelp av en Metropolis-Hastings algoritme ? (Du trenger ikke gjøre det i praksis.)

Oppgave 4: Analyse av punktmønstre

I denne oppgaven skal vi se på tre forskjellige punktmønstre. To ligger i biblioteket "spatial", de heter "cells.dat" og "redwood.dat". Ett datasett ligger på katalogen `/local/kurs/romstat/` og heter "pines2.dat". De kan alle leses inn med kommandoen `ppinit` fra biblioteket "spatial".

a. Plott punktmønstrene. Undersøk om Poisson antakelsen holder for noen av punktmønstrene. Bruk forskjellige tester, f.eks. L-funksjon med tilhørende enveloper, dispersion index test, andre måter?

b.Lag kode for å estimere D funksjonen. Hvor D er fordelingen til avstanden fra et punkt til nærmeste nabopunkt. Plott denne for de tre eksemplene. Nyttige splus funksjoner er: `dist()` (samt funksjonen `dist2full` som beskrevet i hjelp for funksjonen `dist()`) . `apply( , ,min)` som kan brukes til å finne kolonne/radvis minimum.

c. Velg et punktmønster som ikke er Poisson fordelt og foreslå en bedre modell. Forsøk å estimere parametre i den foreslåtte modellen. Sjekk modelltilpasning ved hjelp av L og D funksjonen.

Nyttige kommandoer: `nlmin` ,`integrate` ,`pplik` Du trenger ikke benytte noen av disse.

References:

Cressie; N.; 1993: Statistics for Spatial Data, Wiley & Sons.

Stoyan; D. and Stoyan; H.; 1994: Fractals, Random Shapes and Point Fields; Wiley & Sons. Kap 11,12,13,15,16