Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Jo Eidsvik
Tlf. 90127472

EKSAMEN I FAG TMA4250
ROMLIG STATISTIKK

Mandag 23. mai 2011

Tid: 09:00–13:00

Tillatte hjelpemidler:
Statistiske tabeller og formler, Tapir. Egetprodusert gult ark. Godkjent kalkulator.

Sensur: 13. Juni 2011

Oppgave 1

Vi studerer regresjonsmodellen med romlig korrelerte Gaussiske støyledd. Responsen $y(\boldsymbol{s})$ ved koordinat $\boldsymbol{s}$ modelleres med

$$y(\boldsymbol{s}) = \boldsymbol{x}^t(\boldsymbol{s})\boldsymbol{\beta} + \epsilon(\boldsymbol{s}), \qquad \boldsymbol{s} = (s(1), s(2)) \in (0,1) \times (0,1). \tag{1}$$

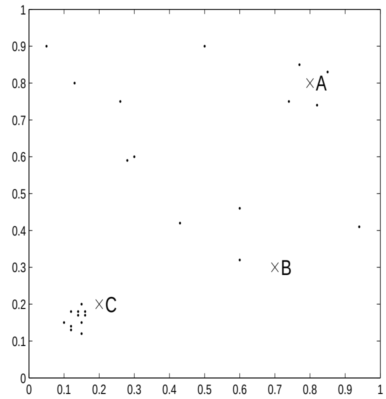
Her er $\boldsymbol{x}(\boldsymbol{s})$ en $p \times 1$ vektor med forklaringsvariable ved koordinat $\boldsymbol{s}$, mens $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ er regresjonsparametre. De Gaussiske støyleddene har forventning 0 og kovarians $\mathrm{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}')) = \Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta})$, der $\boldsymbol{\theta}$ oppsummerer parametre i den romlige kovariansmodellen.

a) Anta at støyprosessen $\epsilon(\boldsymbol{s})$ er stasjonær og isotrop. Hva betyr dette i praksis og hvilke antakelser gir det for den matematiske beskrivelsen av $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta})$?

En mye brukt kovariansfunksjon er den eksponensielle $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \theta_1 \exp(-\theta_2 h)$, der $h = ||\boldsymbol{s} - \boldsymbol{s}'||$ er distansen mellom $\boldsymbol{s}$ og $\boldsymbol{s}'$. En annen er Cauchy $\Sigma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \theta_1/(1+\theta_2 h)^3$. Finn den effektive romlige korrelasjonslengden i disse to modellene, dvs distansen h slik at korrelasjonen er 0.05.

b) Vi ønsker å estimere parametre fra data $y(\boldsymbol{s}_i)$ og kovariater $\boldsymbol{x}(\boldsymbol{s}_i)$, $i = 1, \dots, n$. Vi benytter en eksponensiell kovariansfunksjon for støyleddene, og antar θ_2 kjent. Sett opp likelihoodfunksjonen for $(\boldsymbol{\beta}, \theta_1)$.

Finn maximum likelihood estimatoren for $(\boldsymbol{\beta}, \theta_1)$.



Figur 1: Datapunkter og tre punkter (A, B, C) der vi skal predikere.

c) Vi antar nå at $\boldsymbol{\beta} = 0$. Figur 1 viser en situasjon med $n = 25$ målinger og med tre steder (A, B, C) der vi skal predikere.

For parametre $\boldsymbol{\theta} = (1^2, 7.5)$ blir prediksjonsvariansene i de tre punktene 0.67^2, 0.88^2 og 0.60^2. Hvilke av de tre (A, B eller C) har den minste prediksjonsvariansen (0.60^2)? Hvilken har den høyeste (0.88^2)? Begrunn svaret.

I de neste punktene studerer vi en annen modell for støyprosessen:

$$\epsilon(\boldsymbol{s}) = I[\xi(\boldsymbol{s}) = 0]\epsilon_0(\boldsymbol{s}) + I[\xi(\boldsymbol{s}) = 1]\epsilon_1(\boldsymbol{s}), \qquad \boldsymbol{s} \in (0,1) \times (0,1), \tag{2}$$

der $\xi(\boldsymbol{s})$ er en tilfeldig variabel med utfall 0 eller 1 i hver koordinat $\boldsymbol{s}$. Videre er $\epsilon_0(\boldsymbol{s})$ og $\epsilon_1(\boldsymbol{s})$ *uavhengige* prosesser med forventning 0, varians 1, og kovarians $\Sigma_0(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \exp(-\theta_{0,2}h)$ for $\epsilon_0(\boldsymbol{s})$ og $\Sigma_1(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \exp(-\theta_{1,2}h)$ for $\epsilon_1(\boldsymbol{s})$.

d) Regn ut $\mathrm{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}')|\xi(\boldsymbol{s}), \xi(\boldsymbol{s}'))$ for alle kombinasjoner av $\xi(\boldsymbol{s})$ og $\xi(\boldsymbol{s}')$.

Anta at $s(1) = 0.5$, og at $\xi(0.5, s(2)) = 0$ for $s(2) < 0.25$, $\xi(0.5, s(2)) = 1$ for $0.25 \leq s(2) < 0.35$, $\xi(0.5, s(2)) = 0$ for $0.35 \leq s(2) < 0.75$, og $\xi(0.5, s(2)) = 1$ for $s(2) \geq 0.75$. Lag en skisse med 5 mulige realisasjoner av $\epsilon(0.5, s(2))$ prosessen for parametre $\theta_{0,2} = 3$ og $\theta_{1,2} = 30$.

Anta nå at variablene $\xi(\boldsymbol{s})$ er definert fra en Gaussisk prosess $z(\boldsymbol{s})$, $\boldsymbol{s} \in (0,1) \times (0,1)$ slik at:

$$\xi(\boldsymbol{s}) = 0 \;\text{ hvis }\; z(\boldsymbol{s}) < 0, \qquad \xi(\boldsymbol{s}) = 1 \;\text{ hvis }\; z(\boldsymbol{s}) \geq 0. \tag{3}$$

Her settes $E(z(\boldsymbol{s})) = 0$, $Var(z(\boldsymbol{s})) = 1$ og kovarians $\Sigma_z(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}_z) = \exp(-\theta_{z,2}h)$, $h = ||\boldsymbol{s} - \boldsymbol{s}'||$.

e) Bruk bivariat normalfordeling til å vise at $P(\xi(\boldsymbol{s}) = 0 \cap \xi(\boldsymbol{s}') = 0) = \frac{\arctan(-\sqrt{1-\rho^2}/\rho)}{2\pi}$, der $\rho = \exp(-\theta_{z,2}h)$, $h = ||\boldsymbol{s} - \boldsymbol{s}'||$.

Ta her for gitt at $\int_{-\infty}^{0} \phi(x)\Phi(bx)dx = \frac{\arctan(1/b)}{2\pi}$, der $\phi(x)$ er tettheten til standard normal fordeling og $\Phi(x)$ er den kumulative funksjonen til standard normal fordeling.

f) Bruk resultatet fra 1e) til å regne ut $\mathrm{Cov}(\epsilon(\boldsymbol{s}), \epsilon(\boldsymbol{s}'))$.

Oppgave 2

a) Anta en homogen Poissonprosess med intensitet λ på enhetskvadratet $(0, 1) \times (0, 1)$. Hva er sannsynligheten for 0 punkt i området $(0, 0.1) \times (0, 0.1)$?

Vi deler $(0, 1) \times (0, 1)$ inn i 100 like store disjunkte celler med areal 0.1^2. Hva er sannsynligheten for at minst en av cellene har 0 punkter? Finn cirkaverdi for λ slik at det er 0.05 sannsynlighet for at dette skjer.

b) Anta nå en ikke-homogen Poissonprosess med intensitet $\lambda(x, y) = \lambda_0 xy$, $0 < x < 1$, $0 < y < 1$. Bruk den samme inndeling med 100 celler. Hva er sannsynligheten for 0 punkt i $((i-1)/10, i/10) \times ((j-1)/10, j/10)$, $i, j = 1, \ldots, 10$?

Presenter en Monte Carlo algoritme for å estimere sannsynligheten for at minst en av de 100 cellene har 0 punkter.

Oppgave 3

Data $\boldsymbol{x}$ er samlet inn for et regulært todimensjonalt grid av størrelse $m \times n$. Vi modellerer $\boldsymbol{x} = (x_1, \ldots, x_{mn})'$ som et binært Markov felt med celleverdier $x_i \in \{0, 1\}$, $i = 1, \ldots, mn$. Vi bruker en modell med andre-ordens naboskap.

a) Tegn et andre-ordens naboskap for et regult̊ grid. Hva er en clique? Definer cliquene til dette andre ordens naboskapet.

Tegn alle cliquekonfigurasjonene (16 stk). Bruk symmetribetraktninger til å dele inn konfigurasjonene i 4 klasser: 1 'ensfarget', 2 'hjørne', 3 'rad/kollonne', 4 'diagonal'.

Hver av disse fire gruppene av konfigurasjoner angis med et potensial $\phi(\boldsymbol{x}_k)$, der $\boldsymbol{x}_k$ er kombinasjonene av variablene i en clique. $\phi(\boldsymbol{x}_k)$ tar en av verdiene; ϕ_1, ϕ_2, ϕ_3, ϕ_4.

b) Definer pseudolikelihood som produktet av de fulle betingete fordelingene over alle enkeltceller.

Forenkle noen av leddene i pseudolikelihood for modellen over.

Beskriv en metode for å finne maximum pseudolikelihood estimatoren for $(\phi_1, \phi_2, \phi_3, \phi_4)$.