Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Håkon Tjelmeland 73 59 35 38/73 94 27 25

EKSAMEN I FAG 75563 ROMLIG STATISTIKK

Lørdag 29. mai 1999
Tid: 09:00–13:00

Hjelpemidler:
Godkjent lommekalkulator
Statistiske tabeller og formler, TAPIR.

Sensuren faller: 19. juni 1999.

Oppgave 1 Gaussiske felt

I denne oppgave skal vi betrakte et Gaussisk felt $\{Z(x); x \in R^2\}$. Som kjent er et Gaussisk felt entydig spesifisert ved forventningsfunksjonen, $\mu(x) = \mathrm{E}[Z(x)]$, variansfunksjonen, $\sigma^2(x) = \mathrm{Var}[Z(x)]$, og korrelasjonsfunksjonen $\rho(x, x') = \mathrm{Corr}[Z(x), Z(x')]$.

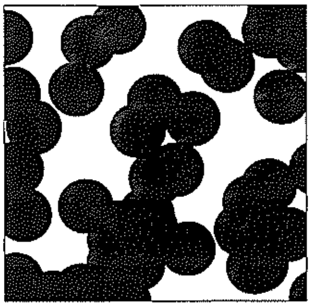
a) En funksjon $\rho(x, x')$ må oppfylle visse krav for at den skal være en lovlig korrelasjonsfunksjon. Spesifisér disse kravene matematisk og forklar bakgrunnen for dem.

   Angi et eksempel på en lovlig korrelasjonsfunksjon og skisser funksjonen.

Anta i resten av oppgaven at $\mu(x)$, $\sigma^2(x)$ og $\rho(x, x')$ er kjente funksjoner. Anta videre at verdien til det Gaussiske feltet er observert i n posisjoner, $x_1, \ldots, x_n$, og at man ønsker å predikere verdien i en ny posisjon x_0. Til dette skal en lineær prediktor benyttes, dvs. en prediktor på formen

$$\widehat{Z}(x_0) = \alpha_0 + \sum_{i=1}^{n} \alpha_i Z(x_i). \tag{1}$$

b) Bestem $\alpha_0, \alpha_1, \ldots, \alpha_n$ slik at prediktoren blir forventningsrett og prediksjonsfeilen, $\widehat{Z}(x_0) - Z(x_0)$, får minst mulig varians.



Figur 1: Eksempel på et bilde med "poisson-disker".

Oppgave 2 Bilde av "Poisson-disker"

I denne oppgaven skal vi anta at vi observerer et bilde av hva man kan kalle "Poisson-disker", et typisk bilde av denne typen er gitt i figur 1. Mer presist skal vi anta følgende modell. La $N = \{x_n\}$ være et stasjonært Poisson punktfelt på R^2 med (konstant) intensitet λ. Til hvert punkt $x \in N$ er der assosiert en sirkelskive $b(x;r)$ med senter i punktet x og med (kjent) radius r, dvs $b(x;r) = \{y \in R^2 : ||y-x|| \leq r\}$. Det observerte bilde er et vindu, W, av denne prosessen der posisjoner som dekkes av (minst) en sirkelskive blir farget sort, mens andre posisjoner er hvite. Det sorte området i bildet kan dermed skrives som

$$B = \left(\bigcup_{x \in N} b(x;r) \right) \cap W.$$

Vi skal anta at parameteren λ er ukjent og vi ønsker å estimere denne fra det observerte bilde.

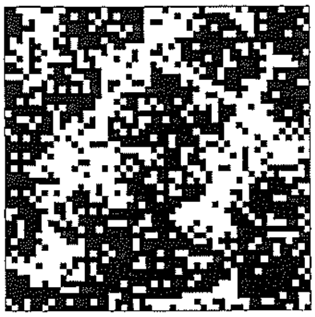
a) For en vilkårlig posisjon x i bildet, vis at

$$\Pr\{x \notin B\} = \exp\{-\lambda \pi r^2\}.$$

Spesifisér spesielt hvilke(n) egenskap(er) ved Possion punktfeltet N du benytter for å komme frem til resultatet.

b) Benytt resultatet i punkt a) til å finne et uttrykk for $\mathrm{E}[|B|]$ som funksjon av λ (samt kjente størrelser). $|B|$ betegner her arealet av mengden B.

Benytt så dette til å foreslå en estimator for λ.



Figur 2: Et diskretisert og støyfull variant av bildet i figur 1, der $n_1 = n_2 = 50$ og $p = 0.8$.

I det siste punktet i denne oppgaven skal vi anta at vi ikke kan observere bildet av "Poisson-diskene" direkte, i stedet observerer man kun et diskretisert bilde med støy. Mer presist skal vi anta at vi observerer et binært bilde på et grid, $Y_{ij} \in \{0, 1\}; i = 1, \ldots, n_1, j = 1, \ldots, n_2$. Dersom senteret av piksel (i, j) er dekket av mengden B, vil $Y_{ij} = 1$ (sort) med kjent sannsynlighet p og $Y_{ij} = 0$ (hvit) med sannsynlighet $1 - p$. Dersom derimot senteret til piksel (i, j) ikke er dekket av B, er $Y_{ij} = 1$ med sannsynlighet $1 - p$ og $Y_{ij} = 0$ med sannsynlighet p. Vi skal dessuten anta at de ulike Y_{ij} er stokastisk uavhengige (for gitt N). Et typisk eksempel på et slikt bilde er vist i figur 2.

c) Modifiser utregningene i punkt a) og b) og kom frem til en estimator for λ basert på $Y_{ij}, i = 1, \ldots, n_1, j = 1, \ldots, n_2$

   Hva skjer dersom $p = 0.5$, og hvorfor ?

Oppgave 3 Markovfelt

Betrakt et Markovfelt definert på et rektangulært grid med n piksler. Anta at pikslene er nummerert fra 1 til n slik at tilstandsvektoren kan skrives $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ der X_i betegner verdien til piksel nummer i. Vi skal videre benytte vanlig notasjon ved at $S = \{1, \ldots, n\}$ er mengden av alle piksler og $\boldsymbol{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Anta at $X_i \in \{0, 1\}$ og at fordelingen til $\boldsymbol{X}$ er gitt ved

$$\Pr\{\boldsymbol{X} = \boldsymbol{x}\} = \text{konst} \cdot \exp\left\{\beta \sum_{i \sim j} I(x_i = x_j)\right\}$$

der summen er over alle par av piksler som er nærmeste naboer, dvs. alle horisontale og vertikale nabopar. Denne modellen kalles (som kjent) gjerne Ising-modell.

a) Beskriv hvordan man (i prinsippet) kan simulere $\boldsymbol{X}$ direkte (uten å bruke MCMC). Forklar også hvorfor denne metoden i praksis ikke er gjennomførbar (for eksempel når $n = 100 \times 100$).

   Spesifisér en Metropolis-Hastings algoritme for å simulere fra den spesifiserte modell (spesifisér spesielt uttrykk for overgangssannsynlighetene i forslagsmatrisa, Q, og finn uttrykk for tilhørende akseptsannsynligheter).

Anta så at det er observert en realisasjon, $\boldsymbol{x}$, fra Ising-modellen gitt over. Fra det observerte bildet er det ønskelig å estimere parameteren β.

b) Forklar hvorfor det i praksis er problematisk å regne ut sannsynlighetsmaksimeringsestimatoren for denne modellen.

   Definér pseudolikelihood-funksjonen for modellen over. Finn så maksimum-pseudolikelihoodestimatoren for modellen (Det er tilstrekkelig at du finner en ligning der estimatoren er løsningen).