



Bokmål

Faglig kontakt under eksamen:

Håkon Tjelmeland 73 59 35 38

EKSAMEN I EMNE TMA4300 MODERNE STATISTISKE METODER

Tirsdag 23. mai 2006

Tid: 09:00–13:00

Hjelpemidler: Kalkulator HP30S

Statistiske tabeller og formler, Tapir forlag.

K. Rottman: Matematisk formelsamling.

Ett gult ark (A4 med stempel) med egne formler og notater.

Sensur er ferdig: 13. juni 2006.

Oppgave 1 Bilkollektivet

Per er nylig ferdig med sine studier i Trondheim og ønsker å få tilgang til bil. Han blir tipset av noen kolleger om at det finnes et bilkollektiv i Trondheim. Han bestemmer seg for å undersøke hvor godt dette vil fungere for ham. Per får opplyst at kollektivet har $m = 30$ medlemmer og at det disponerer $k = 5$ biler. I kollektivet kan medlemmene reservere bil for en dag av gangen. Dersom mer enn k medlemmer reserverer bil en dag, avgjøres det med loddtrekning hvem som faktisk får disponere bilene.

Per antar (for $i = 1, \dots, m$) at hver dag reserverer medlem nummer i bil med en sannsynlighet p_i . Videre antar han at medlemmene opptrer uavhengig av hverandre og at hva de gjør en dag er uavhengig av hva de gjør en annen dag. Per får tilgang til kollektivets kartotek over reserverasjoner de siste $n = 50$ dager og lar x_i være antall reserverasjoner medlem nummer i har gjort i denne perioden. Dette gir følgende tabell.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	0	0	7	0	1	13	5	5	7	9	3	0	3	11	1
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
x_i	7	18	2	10	3	0	1	4	9	0	13	6	2	2	10

Per velger videre å betrakte $p_i, i = 1, \dots, m$ som stokastiske variabler og antar apriori at p_i 'ene er uavhengige og beta-fordelt med parametre α og β , dvs

$$f(p_i|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1} \quad , \quad p_i \in [0, 1],$$

og til slutt at parametrene α og β i denne fordelingen er apriori uavhengige og eksponensialfordelt med samme (kjente) parameter λ .

- a)** Tegn DAGen (Directed Acyclic Graph) for modellen Per benytter.

Forklar kort hvordan man fra en DAG generelt kan finne tilhørende uavhengighetsgraf. Tegn så spesielt uavhengighetsgrafene i Pers situasjon.

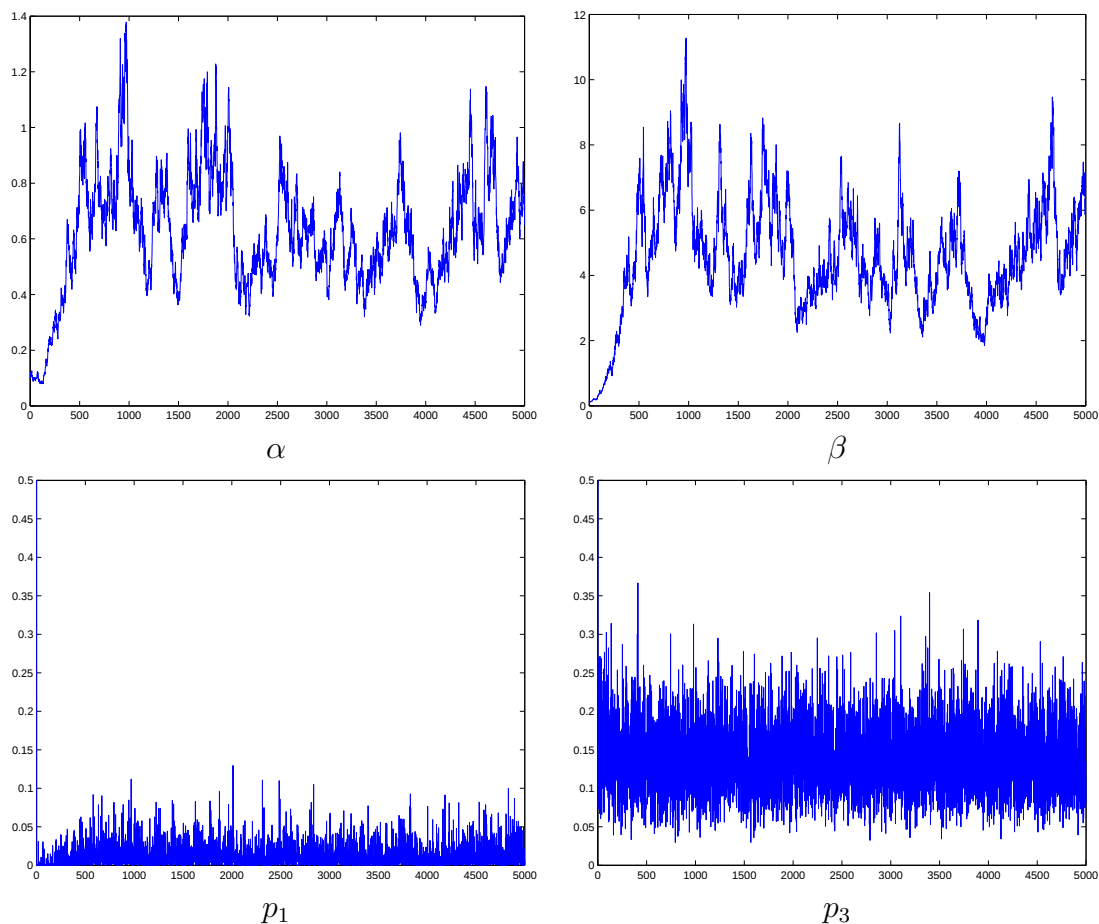
- b)** Utled den fulle betingede fordelinga for p_i , dvs finn $f(p_i|\alpha, \beta, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_m, x_1, \dots, x_m)$, og vis at denne kan uttrykkes som

$$f(p_i|\alpha, \beta, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_m, x_1, \dots, x_m) \propto p_i^{\alpha+x_i-1} (1 - p_i)^{\beta+m-x_i-1} \quad , \quad p_i \in [0, 1].$$

Hva heter denne typen sannsynlighetsfordeling? Hva blir normaliseringskonstanten i denne (betingede) fordelinga?

- c)** Skriv pseudokode for å simulere fra den fulle betingede fordelinga oppgitt i punkt **b)** ved hjelp av “rejection sampling”. Benytt en uniformfordeling på intervallet $[0, 1]$ som forslagsfordeling og finn spesielt akseptanssannsynligheten uttrykt ved parametrene som inngår i problemet.
- d)** Skriv pseudokode for å simulere fra aposteriorifordelingen $f(\alpha, \beta, p_1, \dots, p_m|x_1, \dots, x_m)$ ved hjelp av en Metropolis–Hastings algoritme. Spesifiser spesielt hvilke forslagsfordelinger du velger å benytte og utled uttrykk for tilhørende akseptanssannsynligheter som funksjon av parametrene som inngår i problemet.

Figur 1 viser traceplott av α , β , p_1 og p_3 for de første 5000 iterasjoner av en Metropolis–Hastings algoritme for $f(\alpha, \beta, p_1, \dots, p_m|x_1, \dots, x_m)$ (algoritmen benyttet her er naturlig nok ikke nødvendigvis eksakt den samme som den du valgte i punkt **d)**).



Figur 1: Traceplott av α , β , p_1 og p_3 for en Metropolis-Hastings algoritme for aposteriorfordelingen $f(\alpha, \beta, p_1, \dots, p_m | x_1, \dots, x_m)$. Horisontal akse viser antall iterasjoner, vertikal akse angir parameterverdi.

- e) Cirka hvor lang vil du si burn-in-perioden er for kjøringen vist i figur 1? Svaret skal begrunnes!

Innfør en notasjon for å representere simuleringsresultatene fra algoritmen i punkt **d**) og angi hvordan man kan benytte disse (og eventuelt andre simuleringer, som du i så fall må spesifisere hvordan du vil utføre) til å estimere

1. sannsynligheten for at medlem nummer $i = 5$ på en vilkårlig dag reserverer bil,
2. forventet antall medlemmer som på en vilkårlig dag reserverer bil, og
3. forventet antall medlemmer som på en vilkårlig dag reserverer bil men som likevel ikke får bil til disposisjon.

Oppgave 2 Firmafesten

Et stort firma med 478 ansatte har 50-års jubileum. Det skal arrangeres en jubileumsfest for de ansatte med eventuelle ledsagere. Mulige utfall for hver ansatt er altså: ikke delta/delta alene/delta med ledsager. Ledelsen i firmaet tilbyr subsidiering av jubileumsfesten med

kr 300 for hver ansatt som deltar,
kr 100 for hver ledsager som deltar.

Ledelsen krever imidlertid et anslag på størrelsen av subsidiet.

Firmastatistikeren får oppgaven med å utarbeide anslaget og foretar en rundspørring blant 10 tilfeldig utvalgte ansatte. Statistikeren spør om antallet som hver enkelt av disse vil bringe til festen og får svarene

0 1 1 2 2 0 1 1 0 2.

La N_i benevne den tilfeldige variable som representerer antallet som ansatt nummer i vil bringe til festen og la den tilfeldige variabelen S representere størrelsen på totalsubsidiet fra firmaet.

a) Spesifiser S uttrykt ved $N_i, i = 1, \dots, 478$.

Estimer sannsynlighetsfordelingen til N_i og estimer forventet verdi for S .

Firmastatistikeren ønsker også å bestemme usikkerheten i anslaget på subsidiet S , og beslutter å bruke ikke-parametrisk bootstrapping for å finne denne.

b) Skriv pseudokode for hvordan variansen til estimatet av forventet verdi for S kan estimeres på datamaskin.

c) Regn ut eksakt verdi for bootstrapvariansen til estimatet av forventet verdi for S (*Hint: Betrakt sannsynlighetsfordelingen for antall 0'ere, 1'ere og 2'ere i et bootstrapvalg*).