UNIVERSITETET I OSLO
Matematisk Institutt

EKSAMEN I:	**STK 1130: Modellering av stokastiske prosesser**
VED:	**Nils Lid Hjort**
TIDSPUNKT:	**Tirsdag 12. juni 2007, 14:30–17:30**
HJELPEMIDLER:	**«Formelsamling til STK 1100 og STK 1110»,** **samt kalkulator**

Oppgavesettet har fire oppgaver og er på seks sider (inkludert tre plass-trengende figurer).

Oppgave 1

NORGES SKØYTEFORBUND har masseprodusert samlekort med i alt N skøyteløpere (norske såvel som internasjonale). Kortene er glansede og informative, med dramaturgisk velvalgt fotografi på den ene siden og opplysninger om personlige rekorder og spesielle bedrifter på den andre, og selges i landets kiosker for kr. 1 pr. kort. Dog er det slik at den som kjøper et kort får et vilkårlig kort, fra kioskens bunke, uten å kunne sjekke på forhånd hvem det er bilde av. Samleren risikerer altså å få et kort hun allerede har fra før, og kan følgelig forvente å måtte kjøpe langt flere enn N kort for å få en komplett samling av de N forskjellige kort.

Vi følger nå en av landets kortsamlere, og lar X_n betegne *antall forskjellige kort* hun har etter n innkjøp (der hun altså kjøper ett kort om gangen). Vi skal anse $\{X_1, X_2, \ldots\}$ som en Markovkjede på utfallsrommet $\{1, \ldots, N\}$, med overgangssannsynligheter

$$\begin{aligned} p_{i+1,i} &= \Pr\{X_{n+1} = i+1 \mid X_n = i\} = 1 - i/N, \\ p_{i,i} &= \Pr\{X_{n+1} = i \mid X_n = i\} = i/N, \end{aligned} \tag{1}$$

for $i = 1, \ldots, N$. Kjeden starter altså med $X_1 = 1$, og samlingen er komplett i det øyeblikk $X_n = N$.

(a) Forklar kort hvilke forhold som bør gjelde for at denne Markovkjede-antagelsen og formlene (1) kan anses å være korrekte. Gjør rede for hvilke tilstander som (eventuelt) er transiente, hvilke som (eventuelt) er rekurrente, og hvilke som (eventuelt) er absorberende. Hva er sannsynligheten for at samlingen aldri blir komplett?

(b) La u_i være forventet tid (målt i antall nye kjøp) til samlingen er komplett, gitt at samleren har rukket å skaffe seg i forskjellige kort. Utled følgende formel, for sammenhengen mellom u_i og u_{i+1}:

$$(1 - i/N)u_i = 1 + (1 - i/N)u_{i+1} \quad \text{for } i = 1, \ldots, N-1.$$

(Ligningene er ikke spesielt vanskelige å løse, men skal ikke løses før etter kl. 17:30.)

(c) Anta at $X_n = i$, og la T_i være antallet nye kortkjøp som må til inntil kjeden har nådd nivået $i+1$. Vis at

$$\Pr\{T_i = m\} = (i/N)^{m-1}(1 - i/N) \quad \text{for } m = 1, 2, 3, \ldots.$$

(d) Finn forventet antall kjøp samleren må gjennomføre før samlingen er komplett. Hint: forklar først hvorfor denne variabelen kan skrives

$$Z = 1 + T_1 + T_2 + \cdots + T_{N-1},$$

med notasjon fra forrige punkt.

MAREN HAUGLI *(39.94 / 1.55.99 / 4.00.34 / 6.54.98, foto: Bjarne Hetland), under allround-verdensmesterskapet i Heerenveen 2007.*

Oppgave 2

«FACEBOOK» ER ET INTERNETT-BASERT sosialt nettverk, som ble etablert i 2004 ved Harvard University. Pr. juni 2007 har antall brukere rukket å stige til ca. 25 millioner, og antallet norske brukere har steget fra ca. 3,000 ved starten av 2007 til ca. 300,000 i juni 2007. Facebook leder til ca. 30 milliarder webside-oppslag pr. måned.

På Facebook skaffer man seg «friends», blir med i (og forlater) «groups», osv. I denne oppgaven skal vi se på et par enkle modeller for *antallet grupper* forskjellige typer brukere er medlem av. Vi lar $X(t)$ betegne antall forskjellige grupper en bruker er medlem av ved tidspunkt t, målt fra brukerens innmeldingstidspunkt; spesielt er altså $X(0) = 0$. Vi skal se på $X(t)$ som en tidshomogen fødsels- og død-Markovprosess med

$$\begin{aligned} p_{i+1,i}(\Delta t) &= \Pr\{X(t+\Delta t) = i+1 \mid X(t) = i\} = \lambda_i \Delta t + o(\Delta t), \\ p_{i-1,i}(\Delta t) &= \Pr\{X(t+\Delta t) = i-1 \mid X(t) = i\} = \mu_i \Delta t + o(\Delta t), \\ p_{i,i}(\Delta t) &= \Pr\{X(t+\Delta t) = i \mid X(t) = i\} = 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t), \end{aligned} \tag{2}$$

der $\mu_0 = 0$. Ulike strategier, eller ulike typer brukere, svarer til ulike modeller (eller numeriske verdier) for λ_i og μ_i. Tolkningen av λ_i og μ_i avhenger av tidsskalaen, som vi her kan la være i døgn.

(a) Sett opp prosessens infinitesimale generatormatrise $\mathbf{Q}$, og *skriv ned ligninger* som bestemmer prosessens stasjonærfordeling (eller ekvilibriumsfordeling) $\pi_0, \pi_1, \pi_2, \ldots$, om den eksisterer, ved hjelp av av $\mathbf{Q}$-matrisen. – *Løsningen* av disse ligningene er av formen

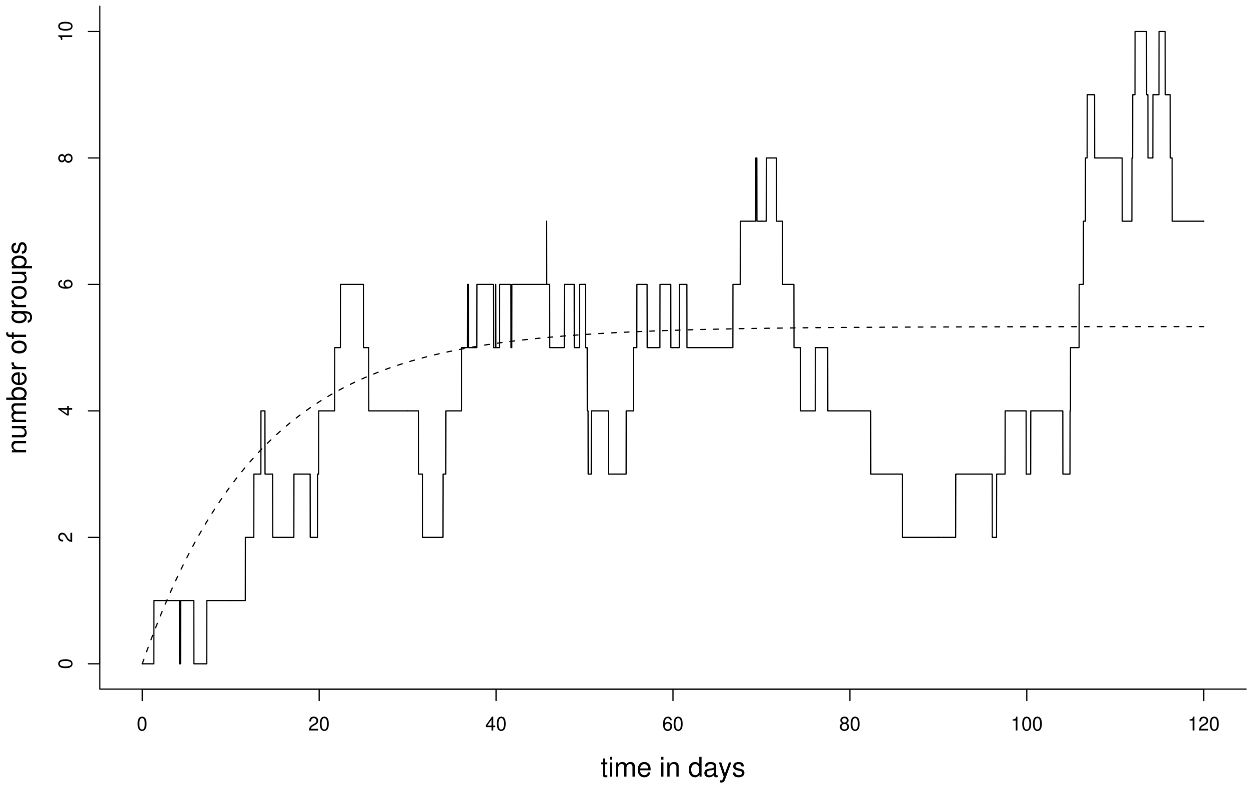
$$\pi_i = \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \pi_0 \quad \text{for } i = 1, 2, 3, \ldots,$$

dersom λ_0 er positiv og summen av π_i-ene er konvergent. Dette er vist i pensum, og skal ikke gjenutledes her, men resultatet kan altså avendes i det følgende.

(b) Facebook-brukere av «Type A» melder seg inn i og ut av grupper som følger: nye grupper, tilstrekkelig interessante til at man velger å melde seg inn, viser seg som etter en Poisson-prosess med fast rate λ; og for hver gruppe man er medlem av, velger man etterhvert å forlate den med rate μ (dvs. at sjansen for å forlate den i løpet av tidsintervallet $[t, t+\Delta t]$, gitt at man er medlem ved tid t, er $\mu\Delta t + o(\Delta t)$). Forklar kort hvorfor disse antagelsene medfører at modellen (2) over holder, med

$$\lambda_i = \lambda \quad \text{og} \quad \mu_i = i\mu \quad \text{for } i = 0, 1, 2, \ldots.$$

(c) For brukere av Type A, vis at stasjonærfordelingen blir en Poisson. Hva er det forventede antall grupper en Type A-bruker er med i, når prosessen har nådd sitt ekvilibrium?



ANTALL GRUPPEMEDLEMSKAP *over fire måneder for en simulert Facebook-bruker av Type A. Vedkommende har* $\lambda = 0.400$ *og* $\mu = 0.075$.

(d) Vi vil finne $m(t) = \mathrm{E}\,X(t)$ for brukere av Type A. Vis først, ved å arbeide med

$$\mathrm{E}\{X(t+\Delta t) - X(t) \,|\, X(t) = i\},$$

at $m'(t) = \lambda - \mu m(t)$. Vis så at

$$m(t) = \frac{\lambda}{\mu}(1 - e^{-\mu t}).$$

(e) En annen type brukere, som vi kaller Type B, har ett «tak» på antall grupper de orker å delta i, og under visse videre antagelser (som det ikke er tid til å komme inn på før etter eksamen) kan man utlede

$$\lambda_i = (N-i)\lambda \quad \text{og} \quad \mu_i = i\mu \quad \text{for } i = 0, 1, \ldots, N,$$

der N altså er vedkommendes øvre smertegrense for antall grupper. Vis at dette leder til en binomisk stasjonærfordeling.

(f) Brukere av Type A oppfører seg naturligvis ikke helt likt, selv om de altså alle kan være av samme grunntype A, ved at deres parameterverdier for λ og μ varierer fra person til person. Vi ser nå på en populasjon av brukere, som hver for seg oppfører seg som beskrevet i punkt (b) over, men der brøkene λ/μ følger en eksponensiell fordeling med parameter 0.10 (altså med tetthet $0.10\,\exp(-0.10\,\theta)$ for $\theta > 0$). La X være antall grupper en tilfeldig valgt person i denne populasjonen er medlem i, idet vi antar at prosessene har kommet frem til stasjonærfordeling. Finn forventning og varians for X.

Oppgave 3

En bestemt forgreningsprosess arter seg ved at hvert nytt individ får avkom uavhengig av alle andre individer, og på en slik måte at antall nye avkom Y et individ avstedkommer følger den samme fordelingen, gitt ved

$$\Pr\{Y = m\} = q^m p = (1-p)^m p \quad \text{for } m = 0, 1, 2, \ldots, \tag{3}$$

der altså $q = 1 - p$. Vi lar X_n betegne antall avkom i generasjon n, og har med andre ord

$$X_n = Y_{n-1,1} + \cdots + Y_{n-1,X_{n-1}},$$

der $Y_{n-1,j}$-ene er de antall avkom som de ulike individer i generasjon $n-1$ produserer.

(a) Vis at den genererende funksjonen for Y blir

$$g(s) = \mathrm{E}\, s^Y = \frac{p}{1-qs} \quad \text{for } |s| < 1/q.$$

Man kan vise fra dette, via to derivasjoner, at Y har forventning q/p og varians q/p^2, uten at disse formlene skal utledes nå.

(b) Finn

$$\mathrm{E}(X_n \,|\, X_{n-1}) \quad \text{og} \quad \mathrm{Var}(X_n \,|\, X_{n-1}),$$

uttrykt ved parameteren p.

(b) Anta at prosessen starter med $X_0 = 1$ individ. Vis at $\mathrm{E}\, X_n = (q/p)^n$.

(c) Anta også for dette punktet at det er $X_0 = 1$ startindivid. Bruk resultater fra pensum til å finne sannsynligheten π for at populasjonen før eller senere skal dø ut, igjen uttrykt ved parameteren p.

(e) Vi skal nå tenke oss at mange eksperimenter utføres i et laboratorium, der hvert av disse starter med ett enkeltindivid, som danner sin private forgreningsprosess, av typen (3) over. Dessuten skal det antas at parameteren p varierer fra startindivid til startindivid; noen har lav p og høy forventning for Y, mens andre har høy p og lav forventning for Y. Dersom p-verdiene varierer i henhold til en uniform fordeling over $(0, 1)$, hva er da sannsynligheten for at den forgreningsprosess som starter med $X_0 = 1$ og avkomsfordeling som i (3), skal dø ut, før eller siden?

Oppgave 4

VI SIER AT $W = \{W(t)\colon t \geq 0\}$ er en Wienerprosess (eller Brownsk bevegelse, eller virrevandring, som det heter i pressemeldingen om årets Abel-prisvinner Srinivasa Varadhan) dersom $W(0) = 0$ og alle tilvekster $W(t) - W(s)$ er uavhengige og normale, med

$$W(t) - W(s) \sim \mathrm{N}(0, \sigma^2(t-s)).$$

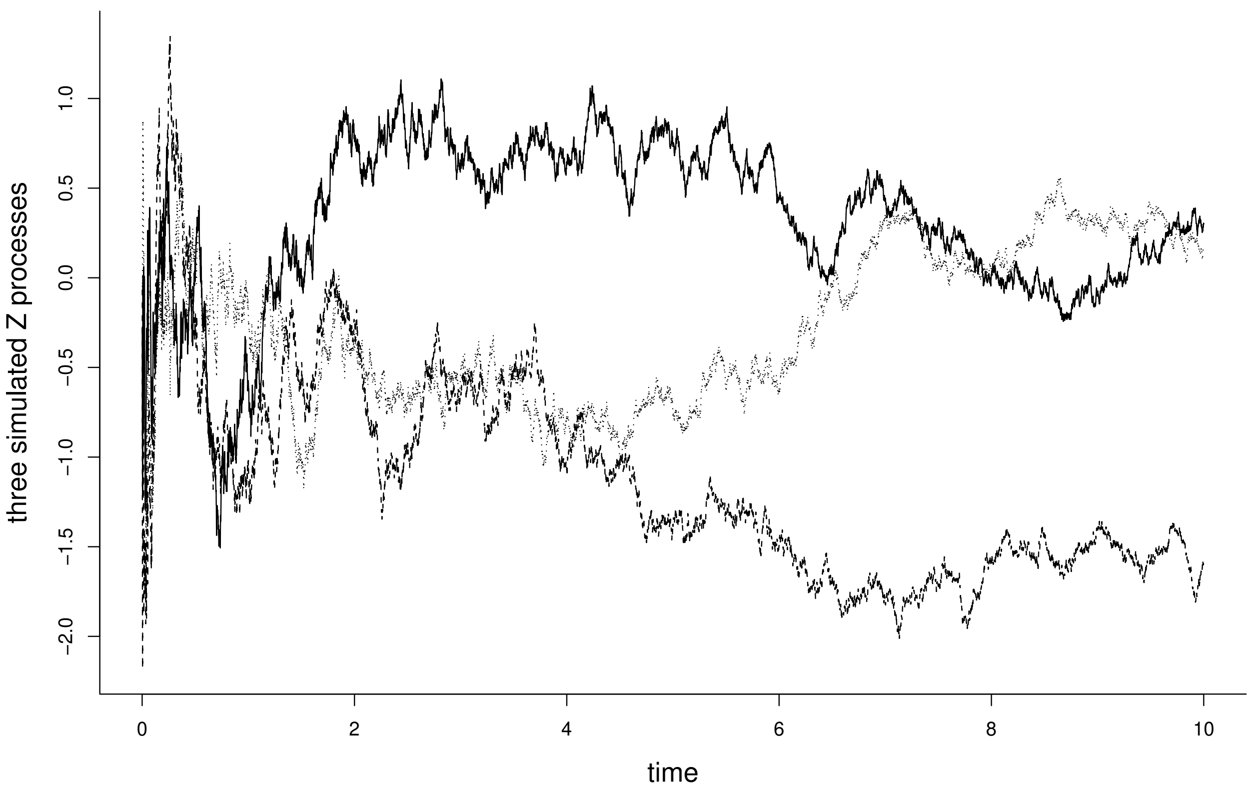
Parameteren σ kalles prosessens spredningsparameter. Spesielt har $W(t)$ varians $\sigma^2 t$.

(a) Finn kovariansen mellom $W(s)$ og $W(t)$, når $s \leq t$.

(b) Se så på den standardiserte prosessen

$$Z(t) = \frac{W(t)}{\sqrt{t}} \quad \text{for } t > 0.$$

Vis at dennes varians er konstant, og finn korrelasjonen mellom $Z(s)$ og $Z(t)$, når $s \leq t$.



TRE SIMULERTE VERSJONER *av den standardiserte Wiener-prosessen Z (med spredningsparameter $\sigma = 1$).*

S-L-U-T-T