

# Første sett med obligatoriske oppgaver i STK1100 våren 2008

Dette er det første settet med obligatoriske oppgaver i STK1100 våren 2008. Oppgavesettet består av to oppgaver. Den første oppgaven er en ren teorioppgave. Den andre oppgaven inneholder både teoretiske punkter, diskusjonspunkter og punkter der du må bruke MATLAB. Det er valgfritt om du vil skrive besvarelsen for hånd eller om du vil bruke et tekstbehandlingsprogram. Uansett skal resultater og figurer fra MATLAB-kjøringene i oppgave 2 tas med i besvarelsen på en hensiktsmessig måte. Hvis flere studenter samarbeider om å løse oppgavene, må likevel hver student levere sin *selvstendige* besvarelse. Det må gå fram av besvarelsen hvem du har samarbeidet med. Se ellers "Regelverk for obligatoriske oppgaver" som er gitt på kurssets hjemmeside.

Besvarelsen leveres ved ekspedisjonen til Matematisk institutt, 7. etasje, Niels Henrik Abels hus.  
*Frist for innlevering er fredag 29. februar kl. 14.30.*

## Oppgave 1

Det er rimelig å anta at antall fødsler  $X$  i løpet av  $t$  timer ved en bestemt fødeavdeling er Poisson-fordelt med punktsannsynlighet

$$P(X = x) = \frac{(\lambda t/24)^x}{x!} e^{-\lambda t/24} \quad x = 0, 1, 2, 3, \dots$$

- a) Anta i dette punktet at  $\lambda = 6$ . Hva er forventet antall fødsler pr. døgn? Hva er sannsynligheten for at det blir født minst ett barn i løpet av et døgn? Hva er sannsynligheten for at det blir født flere enn ett barn i løpet av en time?
- b) La  $T$  være tiden i timer fra midnatt (kl. 00.00) til første fødsel skjer. Vis at

$$P(T > t) = e^{-\lambda t/24} \quad t > 0$$

Finn sannsynlighetstettheten til  $T$ . Utled et uttrykk for forventningsverdien til  $T$  som funksjon av  $\lambda$ . Hva er forventet tid fram til første fødsel dersom  $\lambda = 6$ ?

- c) Anta i dette punktet at  $\lambda = 6$ . Finn sannsynligheten for at det ikke blir født noen barn i løpet av de fire første timene etter midnatt. Anta så at det ikke har blitt født noen barn mellom midnatt og klokka 04.00. Finn sannsynligheten for at en da må vente minst fire timer til før det skjer en fødsel.

## Oppgave 2

Under “Obligatoriske oppgaver” på hjemmesiden til STK1100 er det gitt en tabell over ettårige dødssannsynligheter for norske kvinner og menn (basert på dødelighetsstatistikk for 1993). Du skal i denne oppgaven gjøre en del beregninger med utgangspunkt i denne dødelighetstabellen.

Du får tallene i tabellen inn i MATLAB på følgende måte:

- Klikk på linken til tabellen for å få fram tallene. Lagre filen med navnet `dod.txt` i en katalog der MATLAB kan finne den (dvs. i den katalogen du er i når du starter MATLAB, eller i en katalog i søkestien til MATLAB).
- Siden det er noe tekst i filen før selve dataene, kan du ikke bruke `load` for å lese inn dataene. Du må i stedet bruke den mer generelle funksjonen `textread`. Gå inn i MATLAB og les inn dataene med kommandoen

```
X = textread('dod.txt', '', 'headerlines', 7);
```

(Opsjonen `headerlines` angir størrelsen på filens “header”, altså hvor mange linjer som skal hoppes over før dataene leses inn.)

- Det er greit å ha hver kolonne i tabellen i en egen variabel. Det kan du gjøre ved kommandoene

```
alder = X(:,1); menn = X(:,2); kvinner = X(:,3);
```

- Hvis du har problemer med å lese inn dataene, kan du spørre en medstudent, en av vaktene på PC-stua eller en av lærerne i kurset.

a) Dødelighetstabellen gir ettårige dødssannsynlighetene pr. 1000 individer. Divider tallene i variablene `menn` og `kvinner` med 1000 slik at du får sannsynligheter pr. individ. Ta vare på de ettårige dødssannsynlighetene i variablene `qmenn` og `qkvinner`.

b) Lag plott av de ettårige dødssannsynlighetene som funksjon av alder for både kvinner og menn. Diskuter hvordan dødeligheten endrer seg med alder og hvordan den er forskjellig for de to kjønn.

Her er noen råd som kan være til nytte når du plotter:

- For at du lett skal kunne sammenligne dødeligheten til kvinner og menn er det best å gi dem i samme plott. Bruk kommandoen `hold on` etter første plotting for at det neste plottet ikke skal erstatte eksisterende plott. For at nye plott igjen skal erstatte eksisterende plott, bruk `hold off`.
- Da dødeligheten stiger sterkt med alder, kan det være en fordel å plote den på logaritmisk skala. For å plote med logaritmisk  $y$ -akse, bruk `semilogy`. Funksjonen `semilogy` brukes akkurat som `plot`, men bruker en logaritmisk  $y$ -akse (skriv `help semilogy` og `help plot` for flere detaljer).
- Prøv om du kan få til andre ting som gjør plottene mer informative. For eksempel kan du bruke ulike symboler for kvinner og menn (se hjelpeteksten for kommandoen `plot`) og gi informativ tekst langs aksene.

c) I punkt a beregnet du ettårige dødssannsynligheter for kvinner og menn. De ettårige overlevelsessannsynlighetene er gitt som én minus de ettårige dødssannsynlighetene. Beregn de ettårige overlevelsessannsynlighetene for kvinner og menn og ta vare på dem i variablene `pmenn` og `pkvinner`.

d) Sannsynligheten for at en nyfødt skal bli minst  $k$  år kalles overlevelsessannsynligheten til alder  $k$  år. Denne er gitt som produktet av de ettårige overlevelsessannsynlighetene for alle aldre fra 0 år til og med  $k-1$  år. Forklar hvorfor dette er tilfellet. Beregn overlevelsessannsynlighetene til alder  $k = 1, 2, \dots, 100$  for kvinner og menn og ta vare på dem i variablene `smenn` og `skvinner`. Bruk funksjonen `cumprod` til å gjøre dette (`cumprod` står for “cumulative product”).

e) Det er hensiktsmessig å ha overlevelsessannsynlighetene til alder  $k = 0, 1, 2, \dots, 99$  i stedet for til de  $k$ -verdiene vi beregnet i punkt d. Du oppnår dette ved å lage en ny kolonnevektor hvor det første elementet er 1 og de resterende elementene er de 99 første elementene fra vektorene beregnet i forrige punkt. For eksempel `smenn=[1;smenn(1:99)]`. (Merk at det med dette blir overensstemmelse mellom kolonnen for alder og kolonnene med overlevelsessannsynlighetene.)

f) Lag plott av overlevelsessannsynlighetene fra punkt e som funksjon av alder for både kvinner og menn. Diskuter hvordan overlevelsessannsynlighetene avhenger av alder og hvordan de er forskjellig for de to kjønn. Bestem i denne forbindelse spesielt median levealder, dvs. den alderen det er 50% sannsynlig å overleve til.

g) La den stokastiske variabelen  $X$  betegne levealderen til en nyfødt jente/gutt. Punktsannsynlighetene  $p(k) = P(X = k)$  for  $k = 0, 1, 2, \dots, 99$  gir sannsynlighetene for de ulike verdiene  $X$  kan anta. (Vi ser i dette punktet bort fra muligheten for å bli 100 år eller mer.) Du får beregnet disse punktsannsynlighetene ved å multiplisere (elementvis) kolonnen med ettårige dødssannsynligheter fra punkt a og kolonnen med overlevelsessannsynligheter fra punkt e. Forklar hvorfor dette er tilfellet og foreta beregningene. Ta vare på punktsannsynlighetene for kvinner og menn i nye variabler og lag plott av dem. Kommenter plottene.

h) Beregn forventet levealder for norske kvinner og menn. Hvordan er disse sammenlignet med median levealder? Kommenter.

i) Statistisk Årbok er lagt ut på web ([www.ssb.no/aarbok/](http://www.ssb.no/aarbok/)). I denne kan du blant annet finne opplysninger om forventet gjenstående levetid. Sammenlign de forventede levealderne du fant i punkt h med de som er oppgitt (for 0-åringer) i Statistisk årbok. Diskuter grunnen(e) til eventuelle avvik.