UNIVERSITETET I OSLO

Det Matematisk-Naturvitenskapelige Fakultet

EKSAMEN I:	*ST 101 – Innføring i statistikk og sannsynlighetsregning*
TID FOR EKSAMEN:	*Onsdag 13. mai 1992 kl.* 9^{00}–15^{00}
HJELPEMIDLER:	*Formelsamling for ST 101,* *Karl Rottmann's 'Mathematische Formelsammlung',* *lommekalkulator*

OPPGAVESETTET INNEHOLDER TRE OPPGAVER, OG ER PÅ TRE SIDER

Oppgave 1

Et simuleringsprogram lager stokastiske rektangler i (x, y)-planet av typen

$$R = [0, X] \times [0, Y] = \{(x, y)\colon 0 \leq x \leq X \text{ og } 0 \leq y \leq Y\},$$

der (X, Y) er uniformt fordelt i enhetskvadaratet $[0, 1] \times [0, 1]$, altså med tettheten

$$f(x, y) = \begin{cases} 1 & \text{hvis } 0 \leq x \leq 1 \text{ og } 0 \leq y \leq 1, \\ 0 & \text{ellers.} \end{cases}$$

Programmet lager så mange rektangler som brukeren vil. Rektanglene er stokastisk uavhengige av hverandre.

(a) Vis at denne antagelsen om fordelingen til (X, Y) er ekvivalent med at X og Y er uavhengige og uniformt fordelte på $[0, 1]$. (En variabel U sies å være uniformt fordelt på intervallet $[a, b]$ dersom dens sannsynlighetstetthet er lik $1/(b - a)$ på dette intervallet og lik 0 ellers.)

(b) Finn forventning og varians til X samt forventning og varians til Y.

(c) Hva er sannsynligheten for at R inneholder punktet $(0.7, 0.6)$?

(d) Hva er sannsynligheten for at tre R-rektangler på rad skal inneholde punktet $(0.7, 0.6)$?

(e) Hva er sannsynligheten for at punktet $(0.7, 0.6)$ skal havne til høyre for og over rektanglet R?

(f) Finn forventning og varians til arealet av rektanglet R.

(g) La $R_1 \cap R_2$ være snittet av to genererte rektangler. Finn forventet areal av dette.

(h) Noen rektangler blir nesten kvadratiske mens andre blir flate eller høyreiste. Som et mål på rektangelets avvik fra regularitet, eller 'ikke-firkantethet', brukes kvotienten mellom lengste side og korteste side, altså $W = \max\{X/Y, Y/X\}$. Finn sannsynlighetsfordelingen til W.

Oppgave 2

I forbindelse med en mulig omlegging av en bestemt telefonsentral studerer man blant annet lengden av alminnelige telefonsamtaler. En enkel modell for disse lengdene er at de er stokastisk uavhengige og følger sannsynlighetstettheten

$$f(y, \theta) = \frac{y}{\theta^2} e^{-y/\theta}, \quad y > 0,$$

der θ er en ukjent positiv parameter. Tidsenheten er minutter. Anta at data foreligger i form av samtalelengder $Y_1, \ldots, Y_n$.

— Fortsettes side 2 —

(a) Vis at Y_i har kumulativ fordelingsfunksjon

$$F(y) = 1 - (1 + y/\theta)\, e^{-y/\theta}, \quad y > 0.$$

Hva er sannsynligheten for at en telefonsamtale skal vare mellom 1.00 og 2.00 minutter?

(b) Vis at Y_i har forventning 2θ og varians $2\theta^2$. Du kan benytte at $\int_0^\infty x^p e^{-x}\, dx = p!$ for $p = 0, 1, 2, \ldots$.

(c) Vis at momentestimatoren for parameteren θ blir $\widehat{\theta} = \frac{1}{2}\bar{Y}_n$, der $\bar{Y}_n$ er gjennomsnittet av de n samtalelengdene. Finn estimatorens varians.

(d) Se på størrelsen

$$\widetilde{\theta} = k_n \Big\{ \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2 \Big\}^{1/2},$$

der k_n er en konstant, avhengig av sample-størrelsen n. Hvilken verdi vil du gi k_n for at $\widetilde{\theta}$ skal bli en rimelig estimator for θ? (Eksakt forventningsretthet kreves ikke.)

(e) Sentralgrenseteoremet sier at størrelsen $(\bar{Y}_n - E\bar{Y}_n)/\mathrm{stdev}(\bar{Y}_n)$, der nevneren er standardavviket for $\bar{Y}_n$, er tilnærmet standard normalfordelt når n er stor nok. Bruk dette til å konstruere et konfidensintervall for parameteren θ med konfidensgrad tilnærmet 95%.

(f) Finn, om mulig, en estimator for θ med mindre brutto-varians (forventet kvadratavvik) enn brutto-variansen til $\widehat{\theta} = \frac{1}{2}\bar{Y}_n$.

(g) Lengden av 30 forskjellige telefonsamtaler viste seg å være som følger, målt i minutter (så 3.48 minutter betyr ikke 3 minutter og 48 sekunder, men 3 minutter pluss 48/100 av ett minutt):

```
10.37  1.75  1.79  1.80  3.11  0.92  4.71 1.78
 7.45  1.68  1.00  3.67  1.64  0.50  5.08 1.25
 2.65  3.48  6.90  3.62  5.77  2.22 10.91 2.15
 6.28  5.47  0.18  1.04  2.43  8.78
```

Fra en enkel BLiSS-kjøring klipper jeg ut det følgende:

```
. stat y # please
Statistics:  y
Col   N  Mean  SD     Min     25%    50%    75%    Max
1     30 3.679 2.907 0.1800 1.680 2.540 5.470 10.91
```

Estimer parameteren θ ut fra dette.

(h) Bruk til slutt data-materialet til å estimere sannsynligheten for at tre vilkårlige samtaler alle skal vare minst tre minutter.

Oppgave 3

Begrepet fertilitet, eller fruktbarhet, beskriver den evne en mann og en kvinne har til å få barn sammen. I denne oppgaven skal du gjøre visse beregninger i tilknytning til fertilitet. Noen par er *infertile*, som betyr at de ikke kan få barn. Vi skal anta at dette gjelder 10% av alle par. De øvrige par er *fertile* og kan altså få barn. Vi skal videre anta at et fertilt par som ikke bruker prevensjon har 20% sannsynlighet hver måned for å oppnå befruktning (graviditet).

— Fortsettes side 3 —

Vi skal studere par som planlegger å få sitt første barn. De bruker altså ikke prevensjon (eventuelt starter de uten prevensjon fra et bestemt tidspunkt), og vi vil se på hvor lang tid det tar før de (eventuelt) oppnår befruktning.

(a) Hva er sannsynligheten for at et fertilt par oppnår befruktning i deres tredje måned? Gjør rede for de eventuelle tilleggsantagelser du legger til grunn for å finne sannsynligheten.

(b) Hva er sannsynligheten for at et fertilt par ikke oppnår befruktning i løpet av 12 måneder?

(c) Og hva er sannsynigheten for at et vilkårlig valgt par som prøver å få barn ikke oppnår befruktning i løpet av 12 måneder? Dette paret er altså tilfeldig valgt i populasjonen av både fertile og infertile par.

(d) Hvis et par har forsøkt lenge uten å lykkes kan det være mistanke om infertilitet. En annen forklaring er at de er fertile, men bare har vært uheldige, eller ikke prøvd lenge nok. Som en del av en medisinsk utredning er det følgende spørsmål av interesse: Anta at et par har forsøkt å få barn i 12 måneder uten å oppnå graviditet. Hva er sannsynligheten for at paret er infertilt?

(e) Etter hvor mange måneder uten befruktning er sannsynligheten for at paret er infertilt blitt 99%?

SLUTT