UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST101 — Innføring i statistikk og sannsynlighetsregning.
Eksamensdag:	Torsdag 19. mai 1994.
Tid for eksamen:	09.00 – 15.00.
Oppgavesettet er på 5 sider.	
Vedlegg:	Opptrykk av tabell over kumulative kjikvadratfordelinger. (Redigert fra Tabell A.3 i læreboka til Larsen & Marx.)
Tillatte hjelpemidler:	ST101 formelsamling, kalkulator, Rottmanns "Mathematische Formelsamlung", Jahren og Knutsens "Formelsamling i matematikk" (Tapir forlag).

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1.

I Tabell 1 nedenfor er det gitt 5-årige dødssannsynligheter for norske menn. Av tabellen ser vi for eksempel at sannsynligheten for at en 50 år gammel mann skal dø før han fyller 55 år er 0,036, mens sannsynligheten for at en 65 år gammel mann skal dø før fylte 70 år er 0,138.

Tabell 1. *Dødelighet i 5 års intervaller for norske menn.*

Aldersintervall	[50,55)	[55,60)	[60,65)	[65,70)
Dødssannsynlighet	0,036	0,055	0,091	0,138

a) Hva er sannsynligheten for at en 50 år gammel mann vil oppleve sin 55 års dag? Hva er sannsynligheten for at han vil dø før han fyller 60 år?

b) På en arbeidsplass jobber det fire menn som alle er 50 år gamle. Hva er sannsynligheten for at alle disse vil bli minst 70 år gamle? Hva er sannsynligheten for at nøyaktig to av dem vil leve til de blir 70 år?

(Fortsettes side 2.)

c) På en annen arbeidsplass jobber det tre menn som er henholdsvis 50, 55 og 60 år gamle. Hva er sannsynligheten for at nøyaktig én av disse vil dø i løpet av 10 år?

d) Betrakt de tre mennene fra punkt c). Anta at vi får vite at nøyaktig én av disse døde i løpet av 10 års perioden. Hva er sannsynligheten for at det var den eldste av dem som døde?

Oppgave 2.

På Figurene A – C på neste side er det gitt histogram for tre simulerte datasett, hvert med 250 observasjoner. De (pseudo) stokastiske variable (X) ble generert på følgende måte:

(i) X trukket fra standard normalfordelingen

(ii) X trukket fra Laplace fordelingen med tetthet $\frac{1}{\sqrt{2}}e^{-\sqrt{2}|x|}$

(iii) $X = U_1 + U_2$ hvor U_1 og U_2 er uavhengige og begge trukket fra den uniforme fordelingen over $(-\sqrt{\frac{3}{2}}, \sqrt{\frac{3}{2}})$

Figurene A – C er imidlertid ikke gitt i samme rekkefølge som beskrivelsen av fordelingene *(i)* – *(iii)* .

a) Forklar kort hvorfor $\mathrm{E}X = 0$ for alle de tre fordelingene *(i)* – *(iii)*.

b) Hva er $\mathrm{Var}X$ for hver av fordelingene *(i)* – *(iii)*?

c) Avgjør hvilken av Figurene A – C som svarer til Laplace fordelingen *(ii)*. Svaret skal (selvfølgelig) begrunnes.

d) Avgjør hvilken av Figurene A – C som svarer til fordelingene *(i)* og *(iii)*. Begrunn svarene.

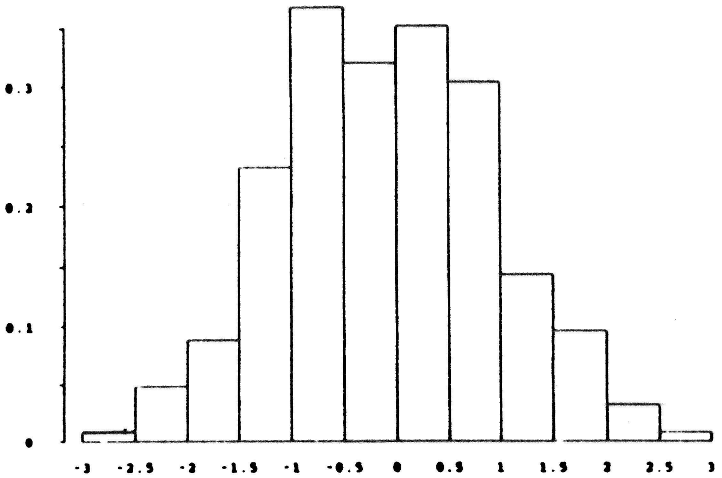
Oppgave 3.

Anta at de stokastiske variablene X og Y er kontinuerlig fordelt med simultan sannsynlighetstetthet

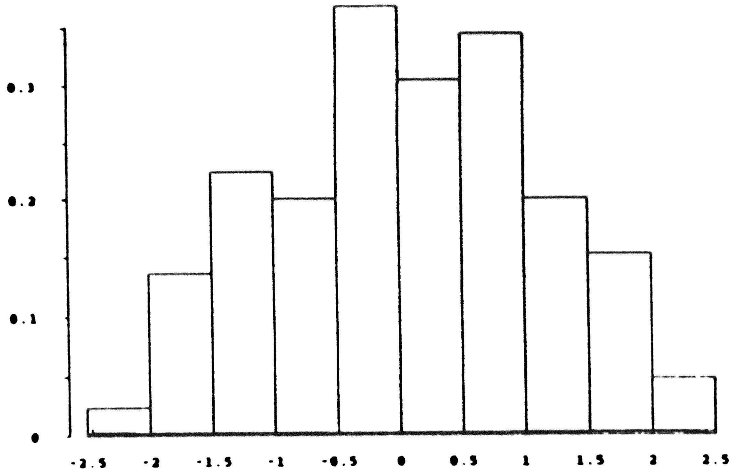
$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{hvis } 0 \le x \le 1 \text{ og } 0 \le y \le 1 \\ \frac{1}{4} & \text{hvis } 1 < x \le 2 \text{ og } 0 \le y \le 2 \\ 0 & \text{ellers.} \end{cases}$$

(Fortsettes side 3.)

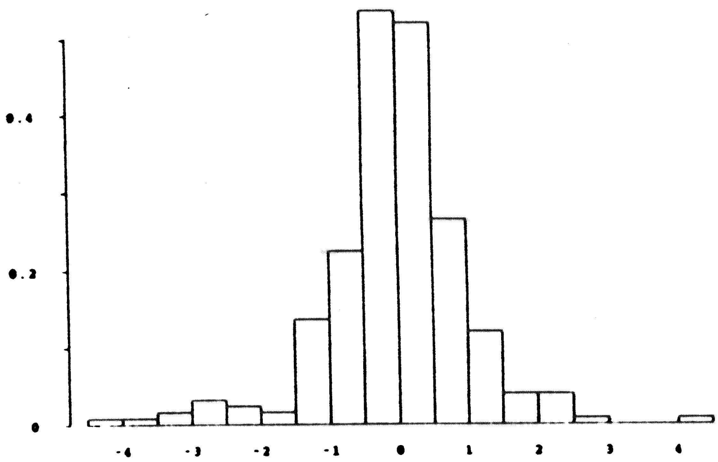
A



B



C



Figur A – C: Histogram for tre simulerte datasett, hvert med 250 observasjoner. (Merk at skalaen langs aksene ikke er lik for de tre datasettene.)

(Fortsettes side 4.)

a) Vis at den marginale sannsynlighetstettheten $f_X(x)$ til X er gitt ved

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{hvis } 0 \leq x \leq 2 \\ 0 & \text{ellers.} \end{cases}$$

b) Bestem den marginale sannsynlighetstettheten $f_Y(y)$ til Y. Er X og Y uavhengige?

Oppgave 4.

Vi minner om at en stokastisk variabel X med sannsynlighetstetthet

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases} \tag{1}$$

sies å være *gammafordelt* med parametre $r > 0$ og $\lambda > 0$. Hvis vi i (1) spesielt har $r = k/2$ og $\lambda = 1/2$, hvor k er et heltall, sier vi at X er *kjikvadratfordelt* med k frihetsgrader. Vi minner også om at hvis $X_1, X_2, \ldots, X_n$ er uavhengige og gammafordelte stokastiske variable, med parametre r_i og λ for X_i, så er $\sum_{i=1}^n X_i$ gammafordelt med parametre $r = \sum_{i=1}^n r_i$ og λ.

Kreftpasienter som gjennomgår en behandling (stråling eller kjemoterapi) kan mange ganger oppnå symptomfrihet uten at sykdommen er helbredet. De kan da senere få tilbakefall slik at symptomene igjen blir manifeste. Således er det i Tabell 2 gitt tid fra behandling til tilbakefall (i uker) for 21 pasienter med akutt leukemi. (Tallene er fra 1960-tallet og er ikke representative for dagens kreftbehandling.)

Tabell 2. *Tid til tilbakefall i uker for pasienter med akutt leukemi*[1]

1	1	2	2	3	4	4	5	5	8	8	8	8	11	11	12
12	15	17	22	23											

[1] Fra Freireich *et al.* (1963). *Blood* **21**, 699-716.

For dataene i Tabell 2 og i mange andre tilfelle kan tid T til tilbakefall beskrives godt ved den eksponensielle fordelingen, dvs. vi kan anta at T har sannsynlighetstetthet

$$f_T(t) = \begin{cases} \mu e^{-\mu t} & \text{for } t > 0 \\ 0 & \text{ellers.} \end{cases} \tag{2}$$

Vi vil først i oppgaven studere den generelle situasjonen med eksponensielt fordelte tilbakefallstider og deretter se nærmere på dataene i Tabell 2.

Anta derfor at tid T til tilbakefall har tettheten (2).

(Fortsettes side 5.)

a) Vis at

$$\lim_{h \to 0} \frac{1}{h} \mathrm{P}(t < T \leq t + h | T > t) = \mu.$$

Dette betyr at μh (tilnærmet) er sannsynligheten for tilbakefall i en liten tidsperiode av lengde h for en pasient som enda ikke har hatt tilbakefall ved starten av perioden. Det er derfor naturlig å kalle μ for *tilbakefalls-intensiteten.*

b) Bestem median tid til tilbakefall og forklar hva denne angir.

c) Vis at $X = 2\mu T$ er kjikvadratfordelt med 2 frihetsgrader.

La nå $T_1, T_2, \ldots, T_n$ være tid til tilbakefall for n kreftpasienter. Vi antar at T_i-ene er uavhengige og identisk fordelte med tettheten (2).

d) Forklar hvorfor $\widehat{\mu} = n / \sum_{i=1}^{n} T_i$ er en rimelig estimator for tilbakefalls-intensiteten.

e) La $x_{\alpha/2}$ og $x_{1-\alpha/2}$ være henholdsvis $\alpha/2$ og $1 - \alpha/2$ fraktilen i kji-kvadratfordelingen med $2n$ frihetsgrader. Dvs. $\mathrm{P}(\chi^2 \leq x_{\alpha/2}) = \alpha/2$ og $\mathrm{P}(\chi^2 \leq x_{1-\alpha/2}) = 1 - \alpha/2$, hvor χ^2 er kjikvadratfordelt med $2n$ frihetsgrader. Vis at

$$\left[\frac{x_{\alpha/2}}{2n} \widehat{\mu}, \frac{x_{1-\alpha/2}}{2n} \widehat{\mu} \right]$$

er et 100(1-α)% konfidensintervall for tilbakefallsintensiteten.

f) Utled et 100(1-α)% konfidensintervall for median tid til tilbakefall.

Vi betrakter så dataene i Tabell 2 over tid fra behandling til tilbakefall for pasienter med akutt leukemi.

g) Estimer tilbakefallsintensiteten for leukemipasientene. Bestem også et 95% konfidensintervall for median tid til tilbakefall.

SLUTT