UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST 101 — Innføring i statistikk og sannsynlighetsregning.
Eksamensdag:	Mandag 30. november 1992.
Tid for eksamen:	09.00 – 15.00.

Oppgavesettet er på 5 sider.

Vedlegg:	Opptrykk av tabell over kumulativ standard-normalfordeling (Tabell A1 fra 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 denne oppgaven skal vi betrakte følgende kontinuerlige fordelinger:

(i) Den uniforme fordeling med tetthet

$$f(x) = \begin{cases} 1/\theta & \text{for } 0 < x < \theta\,; \\ 0 & \text{ellers;} \end{cases}$$

hvor $\theta > 0$.

(ii) Normalfordelingen med tetthet

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}; \qquad -\infty < x < \infty\,;$$

hvor $\sigma > 0$ og $-\infty < \mu < \infty\,.$

(Fortsettes side 2.)

(iii) Den eksponensielle fordeling med tetthet

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & \text{for } \; x > 0\,; \\ 0 & \text{ellers;} \end{cases}$$

hvor $\lambda > 0$.

a) Beregn forventning og varians til den uniforme fordeling.

b) Hvordan defineres medianen til en kontinuerlig fordeling med sannsynlighetstetthet $f(x)$? Bestem medianen til den eksponensielle fordeling.

For en kontinuerlig fordeling med sannsynlighetstetthet $f(x)$ er nedre kvartil q_1 og øvre kvartil q_3 gitt ved

$$\int_{-\infty}^{q_1} f(x)dx = \int_{q_3}^{\infty} f(x)dx = 1/4\,.$$

c) Bestem nedre og øvre kvartil til de tre fordelingene gitt innledningsvis i oppgaven.

Ved stokastisk simulering er det (ved hjelp av datamaskin) trukket 100 uavhengige observasjoner fra hver av fordelingene gitt først i oppgaven (med bestemte valg av verdiene for parametrene θ, μ, σ og λ). Ved hjelp av BLSS-kommandoen "stat" er det videre beregnet en rekke statistiske mål for de genererte datasettene. Noen av disse statistiske målene er gitt nedenfor for hver av de tre datasettene, men ikke (nødvendigvis) i samme rekkefølge som sannsynlighetstetthetene (i)–(iii) først i oppgaven.

Data-sett	Gjennom-snitt (Mean)	Standard-avvik (SD)	Nedre kvartil (25%)	Median (50%)	Øvre kvartil (75%)
A	0.975	0.955	0.200	0.675	1.373
B	1.016	0.597	0.613	1.030	1.429
C	0.987	0.583	0.459	0.981	1.547

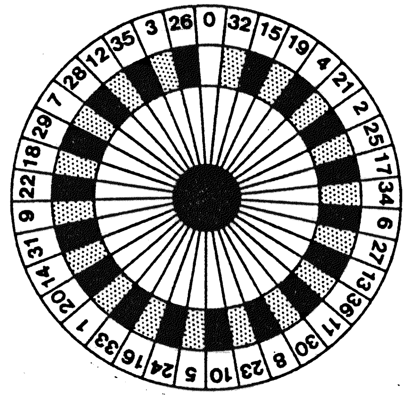
d) Hvilket av datasettene A–C svarer til den eksponensielle fordeling? Svaret skal (selvfølgelig) begrunnes.

e) Hvilket av datasettene svarer til den uniforme fordeling og hvilket svarer til normalfordelingen? Begrunn svarene.

(Fortsettes side 3.)

Oppgave 2.

Rulett er en klassiker blant hasardspillene. Ruletthjulet (vist på figuren nedenfor) er skålformet, horisontalt liggende og kan dreies om midtaksen. En liten kule spretter tilfeldig omkring i skålen under rotasjonen og blir liggende på ett av de 37 nummererte feltene når hjulet stanser. Hver av feltene nummerert $1, 2, \ldots, 36$ er merket med sort eller rød farge, mens feltet nummerert 0 ikke har noen slik farge.



Figur. Ruletthjulet. Prikker angir rød farge.

Spilleren setter sin innsats på visse utvalgte grupper av innsatsfelter, for eksempel de røde feltene, feltene $1, 2, \ldots, 6$ eller ett enkelt felt. Det er ikke tillatt å sette innsats på feltet nummerert 0. Hvis en spiller har satt for eksempel 100 kroner på en gruppe bestående av k innsatsfelter, og kulen stopper på ett av disse, vinner spilleren og hun mottar en gevinst på $100 \cdot (36/k)$ kroner. Hvis kulen stopper på ett av de øvrige feltene, taper spilleren og hun får ingen ting. Uansett beholder kasinoet innsatsen på 100 kroner. Spillerens *nettogevinst* blir altså $100 \cdot \left(\frac{36}{k} - 1\right)$ kroner hvis spilleren vinner, og den blir -100 kroner hvis hun taper. (Negativ nettogevinst svarer til tap.)

Vi betrakter først en "forsiktig" spiller som gjentatte ganger satser 100 kroner på 18 innsatsfelter (for eksempel de røde feltene).

a) Hva er sannsynligheten for at spilleren vinner i første spilleomgang? Hva er sannsynligheten for at spilleren ikke vinner en eneste gang i løpet av de fem første spilleomgangene?

b) La X angi spillerens nettogevinst i én spilleomgang. Vis at $E(X) = -100/37$ og $SD(X) = \sqrt{\mathrm{Var}\,(X)} = 100\sqrt{1368/1369}$.

(Fortsettes side 4.)

c) Selv den "forsiktige" spilleren blir "bitt av spillebasillen", og hun spiller hele 400 spilleomganger i løpet av én kveld. La Y være spillerens samlede nettogevinst i løpet av disse 400 omgangene. Forklar hvorfor Y vil være tilnærmet $N(\mu, \sigma^2)$-fordelt med $\mu = -1081$ og $\sigma = 1999$.

d) Hva er (tilnærmet) sannsynligheten for at spilleren vinner minst 5000 kroner i løpet av kvelden (dvs. $Y \geq 5000$)? Hva er (tilnærmet) sannsynligheten for at hun taper minst 5000 kroner (dvs. $Y \leq -5000$)?

Vi betrakter så en dristig spiller som gjentatte ganger satser 100 kroner på ett innsatsfelt (for eksempel svarende til sin fødselsdato).

e) Hva er (tilnærmet) sannsynligheten for at denne spilleren vinner minst 5000 kroner i løpet av én kveld med 400 spilleomganger? Hva er (tilnærmet) sannsynligheten for at hun taper minst 5000 kroner? Kommenter resultatene i lys av det du fant i punkt d).

Endelig betrakter vi en spiller som har vanskelig for å bestemme seg for om hun skal spille forsiktig, dristig eller "midt på treet". Før denne spilleren går til kasinoet om kvelden kaster hun derfor en terning. Hvis terningen viser 1, 2 eller 3 øyne vil hun denne kvelden hele tiden sette sin innsats på 18 felt, hvis terningen viser 4 eller 5 øyne vil hun hele tiden sette sin innsats på 6 felt, mens hvis terningen viser 6 øyne vil hun hele tiden sette sin innsats på ett felt.

f) Hva er sannsynligheten for at denne spilleren ikke vinner en eneste gang i løpet av de fem første spilleomgangene?

g) Gitt at spilleren ikke har vunnet en eneste gang i løpet av de fem første spilleomgangene, hva er sannsynligheten for at spilleren denne kvelden setter sin innsats på ett felt?

Oppgave 3.

Vi ønsker å studere forekomsten av en *sjelden* begivenhet i en *stor* populasjon i en nærmere angitt tidsperiode. Et eksempel vi skal se på senere i oppgaven er selvmord blant norske kvinner over 15 år i 1990.

Anta at populasjonen består av N individer, og la X være antall tilfeller av den begivenheten vi studerer i løpet av den aktuelle tidsperioden. Vi lar λ betegne sannsynligheten for at et tilfeldig valgt individ i populasjonen skal oppleve begivenheten.

(Fortsettes side 5.)

a) Forklar hvorfor det kan være rimelig å anta at X er Poisson fordelt med parameter λN, dvs.

$$P(X=x) = \frac{(\lambda N)^x}{x!} e^{-\lambda N}; \qquad x = 0, 1, 2, \ldots$$

Hvilke forutsetninger bygger en slik antagelse på?

I resten av oppgaven skal vi anta at X er Poisson fordelt som angitt i punkt a).

b) En estimator for λ er $\hat{\lambda} = X/N$. Er denne estimatoren forventningsrett? Bestem estimatorens varians.

Det er kjent at når λN er rimelig stor, så er

$$\frac{X - \lambda N}{\sqrt{\lambda N}} = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda / N}}$$

tilnærmet standard normal fordelt. Med utgangspunkt i dette resultatet går det an å vise at også

$$\frac{X - \lambda N}{\sqrt{X}} = \frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda} / N}}$$

er tilnærmet standard normal fordelt. (Disse resultatene skal du ikke vise.)

c) Bruk det siste resultatet gitt over til å vise at et tilnærmet 95% konfidensintervall for λ er $(\hat{\lambda} - 1.96\sqrt{\hat{\lambda}/N}\,, \hat{\lambda} + 1.96\sqrt{\hat{\lambda}/N}\,)$.

I 1990 ble det registrert 171 selvmord blant de 1.75 millioner kvinnene over 15 år bosatt i Norge.

d) Gi et estimat for selvmordsrisikoen (dvs. sannsynligheten for å begå selvmord) blant norske kvinner over 15 år. Beregn også et 95% konfidensintervall for denne.

e) I 1950- og 1960-årene ble det i Norge i gjennomsnitt registrert omtrent 5 selvmord hvert år pr. 100000 kvinner over 15 år (svarende til en verdi av λ på $5 \cdot 10^{-5}$). Har det vært en reell økning i selvmordsrisikoen for norske kvinner fra 1950- og 1960-årene og fram til i dag? Diskuter din konklusjon.

SLUTT