UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST101 — Innføring i statistikk og sannsynlighetsregning.
Eksamensdag:	Mandag 3. desember 2001.
Tid for eksamen:	09.00 – 15.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	Tabell over den kumulative standard normalfordelingen.
Tillatte hjelpemidler:	Formelsamlinger for ST100 og ST101, lommeregner, Haugens "Formler og tabeller," Jahren og Knutsens "Formelsamling i matematikk," Rottmanns "Mathematische Formelsammlung."

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

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

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). Videre er det beregnet noen beskrivende statistiske mål for de genererte datasettene. Disse er gitt nedenfor for hver av de tre datasettene, men ikke (nødvendigvis) i samme rekkefølge som sannsynlighetstetthetene (i)–(iii) først i oppgaven.

Datasett	Gjennomsnitt (Mean)	Standardavvik (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.

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

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

I spørsmål a) og c) vil vi anta at $\lambda = 6$.

a) 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

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

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

c) 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.

d) Sannsynligheten for at et barn som fødes er en jente, er 48.6%. La Y være antall jenter som fødes på avdelingen i løpet av et døgn. Gitt at det i løpet av et døgn er født akkurat x barn, er Y binomisk fordelt med parametre $(x, 0.486)$. (Du skal ikke vise dette.)

Vis at (den ubetingede) fordelingen til Y er Poisson med parameter 0.486λ.

Oppgave 3.

Vi minner om at en stokastisk variabel V er kjikvadrat-fordelt med m frihetsgrader hvis den har sannsynlighetstettheten

$$f_V(v)=\begin{cases}\frac{1}{2^{m/2}\Gamma(m/2)}v^{(m/2)-1}e^{-v/2} & \text{hvis} \quad v>0\\ 0 & \text{ellers}\end{cases}$$

For en slik kjikvadrat-fordelt variabel gjelder det at

$$E\left(V^r\right)=2^r\,\frac{\Gamma\left(\frac{m}{2}+r\right)}{\Gamma\left(\frac{m}{2}\right)} \qquad \text{hvis} \quad r>-m/2 \tag{1}$$

(Fortsettes side 4.)

Du skal selv vise (1) i spørsmål f). Men uansett om du klarer å gjøre det eller ikke, kan du bruke dette resultatet i oppgaven.

La nå X være inntekten til en tilfeldig valgt lønnsmottaker i en bestemt befolkningsgruppe. Det er vanlig å anta at X er Pareto-fordelt, det vil si at X har sannsynlighetstettheten

$$f_X(x) = \begin{cases} \theta k^\theta \left(\frac{1}{x}\right)^{\theta+1} & \text{hvis} \quad x > k \\ 0 & \text{ellers} \end{cases} \tag{2}$$

Her er k minsteinntekten i den aktuelle befolkningsgruppen, mens $\theta > 1$ er en parameter som avhenger av lønnsforskjellene i gruppen. Vi vil i hele oppgaven regne med at minsteinntekten k er kjent.

a) Vis at den kumulative sannsynlighetsfordelingen til X blir

$$F_X(x) = \begin{cases} 1 - \left(\frac{k}{x}\right)^\theta & \text{hvis} \quad x > k \\ 0 & \text{ellers} \end{cases}$$

b) Vis at $Y = 2\theta(\log X - \log k)$ er kjikvadrat-fordelt med 2 frihetsgrader.

For å bestemme parameteren θ for den aktuelle befolkningsgruppen, gjøres det $n \geq 3$ observasjoner, $X_1, X_2, \ldots, X_n$, av inntektene i denne gruppen. Du kan gå ut fra at $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte stokastiske variable med sannsynlighetstettheten (2).

c) Forklar hvorfor

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^n \log X_i - n \log k} \tag{3}$$

er en rimelig estimator for θ.

d) Begrunn at $2n\theta/\widehat{\theta}$ er kjikvadrat-fordelt med $2n$ frihetsgrader, og bruk dette og (1) til å vise at $E(\widehat{\theta}) = [n/(n-1)]\theta$. Finn $\text{Var}(\widehat{\theta})$.

e) Estimatoren (3) er ikke forventningsrett. Foreslå en estimator θ^* som er forventningsrett. Vis at θ^* er konsistent, dvs. at for enhver $\epsilon > 0$ vil $\text{P}\left(|\theta^* - \theta| > \epsilon\right) \to 0$ når $n \to \infty$.

f) Vis (1).

SLUTT