UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST101 — Innføring i statistikk og sannsynlighetsregning.
Eksamensdag:	Onsdag 12. mai 1993
Tid for eksamen:	09.00 – 15.00
Oppgavesettet er på 4 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.

Tvillingpar kan være enten eneggede eller toeggede. Sannsynligheten for at det ved en tvillingfødsel blir født eneggede tvillinger er i Nord-Europa omtrent 0.3, og vi vil benytte denne sannsynligheten for enegget tvillingfødsel i denne oppgaven.

De to tvillingene i et enegget tvillingpar har alltid samme kjønn. Anta at sannsynligheten for to gutter ved enegget tvillingfødsel er 0.51. Anta videre at ved en fødsel av toeggede tvillinger er kjønnet til den ene tvillingen uavhengig av kjønnet til den andre, og at sannsynligheten for at en bestemt av tvillingene er gutt er 0.51.

a) Hva er sannsynligheten for to jenter ved enegget tvillingfødsel?

b) Hva er sannsynligheten for at begge tvillingene i et tvillingpar er jenter?
Hva er sannsynligheten for at en er jente og en er gutt?

(Fortsettes side 2.)

c) Begge tvillingene i et tvillingpar er gutter. Hva er sannsynligheten for at de er eneggede?

d) På en skole går det fem tvillingpar hvor begge tvillingene er av samme kjønn. Hva er sannsynligheten for minst ett av disse er toegget?

Oppgave 2.

Den simultane sannsynlighetstettheten for de stokastiske variable X og Y er gitt ved

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x-y} & \text{for } 0 < x < y \\ 0 & \text{ellers.} \end{cases}$$

a) Vis at den marginale sannsynlighetstettheten for X er

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0 \\ 0 & \text{ellers.} \end{cases}$$

b) Finn den marginale sannsynlighetstettheten for Y. Er X og Y uavhengige? Begrunn svaret.

c) Bestem $\mathrm{E}X$ og $\mathrm{E}Y$.

d) Finn sannsynlighetstettheten til $T = (Y/X) - 1$.

Oppgave 3.

I medisinske undersøkelser spiller diagnostiske tester, for eksempel basert på analyse av en blodprøve, en stadig større rolle. Formålet med en diagnostisk test er å avgjøre om en person lider av en gitt sykdom eller ikke.

Nøyaktigheten til en diagnostisk test måles ved dens *sensitivitet* og *spesifisitet*. *Sensitiviteten* er definert som sannsynligheten for at testen er positiv (dvs. indikerer at personen er syk) for en person som faktisk er syk, mens *spesifisiteten* er sannsynligheten for at testen er negativ (dvs. indikerer at personen ikke er syk) for en person som faktisk er frisk.

Vi skal i denne oppgaven betrakte en diagnostisk test som består i å måle konsentrasjonen av et bestemt stoff i en blodprøve. Testen erklæres for å være positiv hvis den målte konsentrasjonen i blodprøven overstiger c μmol/liter, ellers regnes testen som negativ. Konsentrasjonen X er for friske personer $N(\mu, \sigma^2)$-fordelt, mens konsentrasjonen Y for syke personer er $N(\mu + \delta, \sigma^2)$-fordelt.

Anta i punkt a) at $\mu = 40$ μmol/liter, $\delta = 60$ μmol/liter og $\sigma = 20$ μmol/liter.

(Fortsettes side 3.)

a) Hvordan må c velges for at testens sensitivitet skal bli 0.95? Hva blir i såfall spesifisiteten?

I praksis vil ikke μ, δ og σ være kjent, men de må estimeres på grunnlag av analyser av blodprøver. For dette formålet tas det m blodprøver fra friske personer og n blodprøver fra personer som lider av den aktuelle sykdommen. La konsentrasjonene i disse prøvene være henholdsvis $X_1, X_2, \ldots, X_m$ og $Y_1, Y_2, \ldots, Y_n$.

b) Foreslå estimatorer for parameterene μ og δ. Avgjør om estimatorene er forventningsrette og bestem deres standardavvik.

c) Foreslå en forventningsrett estimator for σ^2.

Anta i punkt d) at $m = n = 100$.

d) Foreslå to måter for å estimere c slik at diagnosetesten får en (estimert) sensitivitet på 0.95. Diskuter kort fordeler og ulemper ved de to metodene.

Oppgave 4.

En zoolog ønsker å undersøke forekomsten av en bestemt art dyreplankton i en innsjø. For dette formålet tar hun vannprøver med volum 1 liter og teller antall plankton av den aktuelle arten i prøvene.

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

$$\mathrm{P}(X = x) = \frac{\rho^x}{x!} e^{-\rho} \qquad \text{for } x = 0, 1, 2, \ldots .$$

Forklar hvorfor ρ kan fortolkes som tettheten pr. liter av den aktuelle plankton arten.

I punktene b) og c) vil vi anta at antall plankton i en vannprøve er Poisson fordelt som angitt i punkt a).

Zoologen tar n vannprøver. La $X_1, X_2, \ldots, X_n$ være antall plankton av den aktuelle arten i disse prøvene.

b) Vis at $\widehat{\rho} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er en forventningsrett estimator for ρ. Forklar hvorfor

$$\frac{\widehat{\rho} - \rho}{\sqrt{\rho/n}}$$

er tilnærmet standard normal fordelt. Når er denne tilnærmelsen god?

(Fortsettes side 4.)

c) Zoologen regner med at tettheten i innsjøen av den aktuelle plankton arten ikke kan være større enn 20 dyr pr. liter. Hvor mange vannprøver må hun under denne antagelsen ta for at hun skal være 90% sikker på at $\widehat{\rho}$ ikke skal avvike mere enn 3 fra den sanne tettheten ρ?

I studier av forekomsten av dyreplankton i innsjøer finner en ofte at variasjonen fra vannprøve til vannprøve er større enn det Poisson fordelingen skulle tilsi. En mulig forklaring på dette er at plankton tettheten ikke er den samme i hele innsjøen. Dette vil føre til at parameteren ρ i Poisson fordelingen vil variere fra prøve til prøve. En slik variasjon kan modelleres ved å la tettheten være en Gamma fordelt stokastisk variabel med sannsynlighetstetthet

$$f(\rho) = \frac{\lambda^r}{\Gamma(r)}\rho^{r-1}e^{-\lambda\rho} \qquad \text{for } \rho > 0.$$

Gitt verdien av ρ er så den betingede punktsannsynligheten for X som angitt i punkt a).

d) Forklar hvorfor (den ubetingede) punktsannsynligheten for X blir

$$\mathrm{P}(X = x) = \int_0^\infty \frac{\rho^x}{x!}e^{-\rho}f(\rho)\mathrm{d}\rho \qquad \text{for } x = 0, 1, 2, \ldots .$$

Vis at dette medfører at

$$\mathrm{P}(X = x) = \binom{x+r-1}{x}\theta^r(1-\theta)^x$$

med $\theta = \lambda/(\lambda+1)$, dvs. X er negativt binomisk fordelt.

Du kan i dette punktet fritt benytte følgende resultater om gamma funksjonen: For r et positivt reelt tall og x et ikke-negativt heltall gjelder

$$\begin{array}{lccl}
\text{(i)} & \Gamma(r) & = & \int_0^\infty u^{r-1}e^{-u}\mathrm{d}u \\
\text{(ii)} & \Gamma(x+1) & = & x! \\
\text{(iii)} & \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} & = & \binom{x+r-1}{x}
\end{array}$$

SLUTT