UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1100 — Sannsynlighetsregning og statistisk modellering.

Eksamensdag: Fredag 9. juni 2017.

Tid for eksamen: 9.00 – 13.00.

Oppgavesettet er på 3 sider.

Vedlegg: Ingen.

Tillatte hjelpemidler: Godkjent kalkulator.
Formelsamling for STK1100 og STK1110.

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

Oppgave 1

Vi minner om at hvis en stokastisk variabel V er gamma-fordelt med formparameter α og skalaparameter β, som vi skriver $V \sim \text{gamma}(\alpha, \beta)$, så er sannsynlighetstettheten til V gitt ved

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

a) Anta at $V \sim \text{gamma}(\alpha, \beta)$. Vis at

$$E\left(V^r\right) = \beta^r \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \quad \text{hvis} \quad r > -\alpha. \tag{1}$$

Du får bruk for dette resultatet senere 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.

b) Vis at den kumulative sannsynlighetsfordelingen til X er gitt ved

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

c) Vis at median inntekt er $\widetilde{\mu} = k2^{1/\theta}$ og forklar hva medianen gir uttrykk for.

(Fortsettes på side 2.)

d) Vis at $2\theta \ln(X/k) \sim \text{gamma}(1, 2)$.

For å estimere θ, gjør vi $n \geq 3$ observasjoner av inntektene i den aktuelle befolkningsgruppen. Du kan anta at de observerte inntektene $x_1, x_2, \dots, x_n$ er observerte verdier av stokastiske variabler $X_1, X_2, \dots, X_n$ som er uavhengige og identisk fordelte med sannsynlighetstettheten (2). En mulig estimator for θ er da

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(X_i/k)}. \tag{3}$$

e) Vis at $2n\theta/\widehat{\theta} \sim \text{gamma}(n, 2)$. (Du kan regne som kjent at en sum av uavhengige gamma-fordelte variabler som alle har samme skalaparameter, selv er gamma-fordelt.)
Bruk dette resultatet og (1) til å vise at $E(\widehat{\theta}) = [n/(n-1)]\theta$.

f) Estimatoren (3) er ikke forventningsrett. Foreslå en estimator som er forventningsrett og bestem variansen til estimatoren.

g) Utled et 95% konfidensinterval for θ. Intervallet skal uttrykkes ved hjelp av n, $\widehat{\theta}$, a og b. Her er a og b henholdsvis 2.5% og 97.5% persentilene for gamma-fordelingen med formparameter $\alpha = n$ og skalaparameter $\beta = 2$.

Bestem også et 95% konfidensinterval for median inntekten $\widetilde{\mu}$.

Oppgave 2

Sigdcelleanemi er en farlig blodsykdom som særlig rammer mennesker av afrikansk opprinnelse. Et barn får sykdommen dersom det får et bestemt recessivt gen (a) fra både mor og far. Barnet får ikke sykdommen hvis det får det dominante genet (A) fra minst én av foreldrene.

Vi ser først på et par der både kvinnen og mannen har genotypen Aa. Ingen av dem lider av sigdcelleanemi, men de er bærere av sykdommen og kan føre den videre til sine barn. Et barn de får sammen, arver fra mor det recessive genet a med sannsynlighet $1/2$ og det dominante genet A med sannsynlighet $1/2$. Det samme gjelder for arven fra far. Videre er arven fra de to foreldrene uavhengig.

a) Forklar at sannsynligheten er $1/4$ for at barnet får genotypen aa og dermed vil lide av sigdcelleanemi. Forklar også at sannsynligheten er $1/2$ for at barnet får genotypen Aa og dermed vil være en frisk bærer av sykdomsgenet.

Vi ser så på et par der mannen er bærer av sigdcelleanemi (dvs. har genotypen Aa). Kvinnen lider ikke av sigdcelleanemi, men vi vet ikke om hun er bærer av sykdommen eller ikke. Sannsynligheten for at hun er bærer er 8% (svarende til andelen bærere i den afroamerikanske befolkningen), mens sannsynligheten er 92% for at hun ikke er bærer.

b) Hva er sannsynligheten for at et barn paret får, vil lide av sigdcelleanemi?

c) Paret får tre barn og ingen av dem lider av sigdcelleanemi. Hva er sannsynligheten for at kvinnen er bærer av sykdommen?

(Fortsettes på side 3.)

Oppgave 3

Vi betrakter i denne oppgaven stokastiske variabler X og Y som har simultantetthet av formen:

$$f_{X,Y}(x,y) = \begin{cases} K & \text{hvis } x \geq 0,\ y \geq 0,\ x + y \leq \theta, \\ 0 & \text{ellers.} \end{cases}$$

Her er $\theta > 0$ en parameter, mens K er en konstant som bestemmes slik at $f_{X,Y}$ blir en sannsynlighetstetthet.

a) Vis at $K = 2\theta^{-2}$.

b) Vis at marginaltettheten til X er:

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

Finn også marginaltettheten til Y. Er X og Y uavhengige? Begrunn svaret.

c) Vis at:

$$E(X^r) = \frac{2\theta^r}{(r+1)(r+2)}.$$

Bruk dette resultatet til å finne forventningen $E(X)$ og variansen $V(X)$.

d) Finn den betingede fordelingen for Y gitt $X = x$ uttrykt ved tettheten $f_{Y|X=x}(y)$. Hvilken kjent fordeling er dette? Bestem også $E(Y|X = x)$.

e) Finn sannsynligheten $P(X^2 + Y^2 \leq \theta^2/2)$.

SLUTT