Københavns Universitet
Eksamen ved Det naturvidenskabelige Fakultet sommeren 1996

Matematik 2 SS

Opgaver til besvarelse i 3 timer.

Alle hjælpemidler er tilladt.

Opgave 1

I denne opgave ser vi på en primitiv model for udbredelse af smitte i en population. Hvert individ repræsenteres ved en kugle, der enten er sort (individet er smittet) eller hvid (individet er ikke smittet). Kuglerne anbringes i en urne og til hvert tidspunkt $t = 0, 1, 2, \cdots$ udtrækkes tilfældigt to kugler fra urnen. Hvis kuglerne enten begge er sorte eller begge hvide, lægges de uændrede tilbage i urnen, men hvis den ene er sort og den anden hvid, farves den hvide kugle sort, hvorefter de to kugler (der nu begge er sorte) lægges tilbage i urnen. På denne måde bestemmes indholdet af urnen på det efterfølgende tidspunkt. Det antages, at urnen på begyndelsestidspunktet $t = 0$ (umiddelbart før første udtrækning) indeholder n sorte og m hvide kugler, hvor $n \in \mathbb{N}$, $m \in \mathbb{N}$.

(i) Bestem sandsynligheden for at urnen til tidspunktet $t = 1$ indeholder et uændret antal sorte og hvide kugler, samt sandsynligheden for at urnen til tidspunktet $t = 1$ indeholder én sort kugle mere end til tidspunktet $t = 0$.

(ii) Lad T være første tidspunkt, hvor urnen indeholder én sort kugle mere end til tidspunktet $t = 0$. Figuren illustrerer en situation, hvor $n = 5$, $m = 8$ og $T = 3$.

$$[5,8] \overset{\bullet\bullet}{\curvearrowright} [5,8] \overset{\circ\circ}{\curvearrowright} [5,8] \overset{\bullet\circ}{\curvearrowright} [6,7] \curvearrowright \cdots .$$

Bestem T's fordeling og middelværdi.

(iii) Lad T^* betegne første tidspunkt, hvor urnen indeholder lutter sorte kugler (alle er smittet). Find et udtryk for T^*'s middelværdi og vis herved, at såfremt $n = 1$, er

$$E(T^*) = m \sum_{\nu=1}^{m} \frac{1}{\nu} \qquad (\approx m \log m) .$$

Vejledning. Opfat T^* som en sum af ventetider og udnyt resultatet af (ii). Ved udregningen kan man evt. have glæde af omskrivningen

$$\frac{n+m}{(n+\nu)(m-\nu)} = \frac{1}{n+\nu} + \frac{1}{m-\nu} .$$

(iv) I tilfældet $n = 1$, $n = 2$ skal man bestemme fordelingen af den stokastiske variabel T^*, der blev indført ovenfor.

Opgavesættet fortsættes på side 2

Opgave 2

Lad funktionerne $f : \mathbb{R} \to \mathbb{R}$ og $g : \mathbb{R} \to \mathbb{R}$ være givet ved

$$f(x) = \begin{cases} 0\,; & x < 1 \\ 12(x^{-3} - 2x^{-4} + x^{-5})\,; & x \geq 1 \end{cases}$$

og

$$g(x) = \begin{cases} 0\,; & x < 1 \\ 12(x^{-4} - x^{-5})\,; & x \geq 1\,. \end{cases}$$

(i) Vis, at såvel f som g er tæthedsfunktioner.

I det følgende antages, at X har f som tæthedsfunktion og at Y har g som tæthedsfunktion.

(ii) Bestem $E(X)$, $E(Y)$ og $\sigma^2(Y)$. Undersøg, om X har endelig varians.

(iii) Idet φ betegner funktionen defineret på $]1, \infty[$ ved

$$\varphi(t) = \frac{t}{t-1}\,; \qquad t > 1\,,$$

skal man vise, at X har samme fordeling som $\varphi(Y)$ og at Y har samme fordeling som $\varphi(X)$.

(iv) Bestem middelværdi og varians af den stokastiske variabel $\frac{1}{X-1}$.

Opgave 3

I denne opgave ses på en statistisk model svarende til n uafhængige observationer $x_1, x_2, \cdots, x_n$ af en stokastisk variabel med tæthedsfunktion givet ved

$$f(x) = \frac{(\theta^2 + \theta)(x-1)}{x^{\theta+2}}\,; \qquad x > 1\,,$$

hvor θ er en positiv parameter.

(i) Bestem likelihoodfunktionen hørende til modellen og vis, at stikprøvefunktionen givet ved produktet $x_1 x_2 \cdots x_n$ er sufficient.

(ii) Vis, at maksimaliseringsestimatoren $\hat{\theta}$ (i noterne også kaldet maksimumlikelihood-estimatoren) er bestemt ved ligningen

$$\frac{1}{\hat{\theta}} + \frac{1}{\hat{\theta} + 1} = \frac{1}{n} \log(x_1 x_2 \cdots x_n)\,.$$

(iii) Det antages nu, at $n = 2$. Vis, at for enhver værdi af θ_1 med $\theta_1 > 1$ er testet med kritisk område (også kaldet forkastelsesområde) givet ved

$$C = \{(x_1, x_2) \mid x_1 x_2 \leq \frac{5}{2}\}$$

et bedste test af nulhypotesen $H_0 : \theta = 1$ mod den alternative hypotese $H_1 : \theta = \theta_1$.

(iv) Bestem signifikansniveauet for testet i (iii).

Bemærkning. Til kontrol af regningerne anføres, at signifikansniveauet bliver omtrent 4%.