UNIVERSITETET I OSLO

Matematisk Institutt

Emnekode:	**STK 1110 – Statistiske metoder og dataanalyse 1**
	Avsluttende eksamen
Dato for eksamen:	**Mandag 6. desember 2004, kl. 9:00–12:00**
Hjelpemidler:	**«Formelsamling til STK 1100 og STK 1110»,**
	Kalkulator

Dette oppgavesettet utgjør den andre av kursets to eksamener. Det inneholder fire oppgaver og er på tre sider. Kursets første eksamen ble arrangert 14. oktober d.å.

Oppgave 1

Vi skal se på noe i en homogen Poisson-skog. λ noe $\{$av den bestemte typen$\}$ skog G μ Hensis$\}$ en godtatt slik at $\{a\}$ antall π λ $N(A)$ inneholdt et område A med areal $a(A)$ $\{$målt i kvadratmeter$\}$ en Poisson-fordelt med parameter $\lambda a(A)$; $\{b\}$ antall π λ $N(A)$ og $N(B)$ inneholdt ikke-overlappende områder A og B er stokastisk uavhengige. Her er λ en ukjent parameter som altså svarer til den gjennomsnittlige tre-tettheten pr. kvadratmeter.

$\{a\}$ Gå et tilfeldig utgangspunkt i skogen, la A π λ en sirkel med radius r. Hva er sjansen for at det skal π λ utpeltig tre π λ inneholdt A? Og hva er sjansen for at det ikke skal π λ noen π λ der?

$\{b\}$ La R π λ distansen fra dette utgangspunktet til λ nærmeste tre. Vis at R har sannsynlighetstettheten

$$f(r, \lambda) = \quad^{-\lambda\pi r^2} 2\lambda\pi r \quad \text{for } r \geq 0.$$

$\{c\}$ I dette punktet kan en benytte deg av formlene

$$\int_0^\infty r^2 \quad^{-\pi r^2} dr = \frac{\sqrt{\pi}}{4\sqrt{\pi}} \quad \text{og} \quad \int_0^\infty r^3 \quad^{-\pi r^2} dr = \frac{1}{2\pi^2},$$

som holder for π positiv. Vis at

$$ER = \frac{\frac{1}{2}}{\sqrt{\lambda}},$$

og finn tetthetsfunksjonen til R.

(d) I tillegg til M, distansen fra utgangspunktet til det n -neste tre, kan man observere Z, distansen fra utgangspunktet til det nest n -neste tre. Finn en formel for sannsynlighetsfordelingen til Z.

Oppgave 2

Vi ser nå på [en denne bestemte typen] en det i skogen — en velger da at det er n forskjellige stamtreer i skogen, alle uavhengig spesifiserte, og for disse måler man distansen til det n -neste tre. Dette gir målingene $M_1, \ldots, M_n$.

(a) I det følgende skal vi se på at $M_1, \ldots, M_n$ er uavhengige med den samme sannsynlighetstetthet, nemlig den $f(x, \lambda)$ vi fant i Oppgave 1(b). Diskuter hvor om disse antagelsene spres ulike eller like.

(b) Gjennomsnittet $\bar{M} = \frac{1}{n}\sum_{i=1}^{n} M_i$ vil konvergere i sannsynlighet når n øker. — et bevis gjennom Bernoulli snart.

(c) Finn momentestimatoren for λ, og kall den λ^*. Vis at den er konsistent, altså at λ^* konvergerer i sannsynlighet mot λ når n øker.

(d) Vis så at — $\mathbb{U}$-estimatoren (som ikke blir maksimalsannsynlighetsestimatoren) blir

$$\lambda^{**} = \frac{1}{n}\frac{1}{\bar{M}}, \qquad \text{der } \bar{M} = \frac{1}{n}\sum_{i=1}^{n} M_i^2.$$

(e) — en konkret $n = 25$ slike målinger, og fant disse tallene, i meter:

9.55	4.32	7.85	1.27	2.91	6.77	2.94	3.39	
4.29	4.14	6.16	0.15	4.94	5.59	2.53	9.15	
3.71	4.42	2.34	7.36	0.35	2.21	1.38	0.94	2.70

Fra disse målingene finner man (blant annet)

$$\frac{1}{n}\sum_{i=1}^{n} M_i = 4.122, \qquad \frac{1}{n}\sum_{i=1}^{n} M_i^2 = 28.455, \qquad \text{empirisk standardavvik } 2.503.$$

Dette oppfyller vår en å lage et konfidensintervall for λ med konfidensgrad rifle — mest 95%.

Oppgave 3

En biolog undersøker 300 kanoner, og klassifiserer dem av dem i bestemte kategorier 1, 2, 3. Hun har en genetisk-biologisk hypotese som etter nye sannsynlighetsberegning har konkludert som at sannsynlighetene for at kanonene skal — ne av type 1, type 2, type 3 kan uttrykkes som

$$p_1 = \theta^2, \quad p_2 = 2\theta(1-\theta), \quad p_3 = (1-\theta)^2,$$

for en passende parameter θ. — r kanonen 300 kanoner viser $N_1 = 10$ seg å — ne av type 1, $N_2 = 80$ av type 2, og $N_3 = 210$ av type 3. Hun vil gjenne teste sin tenk.

(a) Gjelder kun de observasjoner som skal til for at likelihoodfunksjonen for data, under denne modell, blir

$$L(\theta) = \frac{300!}{101!\,99!\,100!}\theta_0(\theta)^{101}\theta_1(\theta)^{99}\theta_2(\theta)^{100}.$$

uten å det tilgjengelige at disse betingelsen er oppfylt.

(b) Finn et estimat for θ.

(c) Bruk data til å teste om modellen denne holder.

Oppgave 4

En ny medicinske mål, eller en ikke NI skal se på den gjeldende modellen for $M_0, \ldots, M_{99}$ er uavhengige observasjoner fra den samme normalfordelingen $(\theta, 1)$, og den man altså vil teste $H_0 : \theta = 0$ mot alternativet at $\theta \neq 0$.

(a) Sett opp likelihoodfunksjonen $L_{99}(\theta)$ for modellen, og vis at L-estimatoren under en polynom-restriksjonen er $\hat{\theta} = \bar{M}_{99}$, gjennomsnittet av observasjonene.

(b) Sett opp et generelt uttrykk for GLR_{99}, den generaliserte likelihood-ratio-testen for H_0, og finn en så enkel formel for

$$W_{99} = -2\log \mathrm{GLR}_{99},$$

i denne situasjonen, som mulig.

(c) Hva sier den generelle teorien for prøven om fordelingen til $W = -2\log \mathrm{GLR}$, når n nærmer seg ∞? Hvordan bruker dette approksimasjonsresultatet for situasjonen i denne oppgaven?

(d) Hva denne testen basert på W_{99} blir bedt til å teste $H_0 : \theta = 0$, med forkastnings-grense svarende til at testens nivå blir 0.05. Hvilken verdi man har benytte sannsynligheten for at testen faktisk forkaster H_0, når den riktige θ er lik 0.03.