UNIVERSITETET I OSLO

Matematisk Institutt

EKSAMEN I:	**STK 1110 – Statistiske metoder og dataanalyse 1**
	Midtsemestereksamen
TID FOR EKSAMEN:	**Torsdag 14. oktober 2004, kl. 8:00–11:00**
HJELPEMIDLER:	**«Formelsamling til STK 1100 og STK 1110», kalkulator**

Dette oppgavesettet utgjør den første av kursets to eksamener. Det inneholder fire oppgaver og er på tre sider. Kursets andre eksamen arrangeres mandag 6. desember s.å.

Oppgave 1

ENHVER DATAMASKIN er idag utstyrt med en såkalt «random-generator», som produserer tall som kan regnes som uavhengige og uniformt fordelte over $[0,1]$, altså med sannsynlighetstetthet $f(x) = 1$ over dette intervallet. La $X_1, \dots, X_n$ være slike uniformt fordelte tall.

(a) La

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{og} \quad s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}.$$

Forklar hvorfor $\bar{X}_n$ og s_n konvergerer i sannsynlighet mot henholdsvis $\frac{1}{2}$ og $\frac{1}{\sqrt{12}}$, når n vokser, gjerne med henvisning til generelle resultater fra pensum.

(b) Se deretter på $M_n = \max\{X_1, \dots, X_n\}$. Hva skjer med M_n når n vokser? Vis at M_n vil konvergere i sannsynlighet.

(c) Hvilken sannsynlighetstetthet har $Y_i = \sqrt{X_i}$? Finn grensen i sannsynlighet for $\frac{1}{n}\sum_{i=1}^{n}\sqrt{X_i}$.

Oppgave 2

DEN MOMENTGENERERENDE FUNKSJON for en stokastisk variabel X er som kjent definert som $M_X(t) = \mathrm{E}\, e^{tX}$.

(a) Hvis X er Poisson-fordelt med parameter θ, vis at

$$M_X(t) = e^{\theta(e^t - 1)}.$$

(b) Bruk dette resultatet til å finne forventningsverdien for X.

(c) Vis så, fortsatt ved hjelp av momentgenererende funksjoner, at en sum av uavhengige Poisson-fordelte variable er Poisson-fordelt.

(d) Man arbeider av og til med den såkalte avkortede Poisson-fordeling, som er fordelingen til en Poisson-variabel når nullmålinger ikke kan registreres. Dette leder til fordelingen med sannsynligheter

$$P\{X = x\} = \frac{e^{-\theta}}{1 - e^{-\theta}} \frac{\theta^x}{x!} \quad \text{for } x = 1, 2, 3, \ldots.$$

Finn den momentgenererende funksjonen for en slik fordeling.

Oppgave 3

UAVHENGIGE OBSERVASJONER $X_1, \ldots, X_{200}$ foreligger av et visst fenomen, som antas å følge en fordeling med sannsynlighetstetthet

$$f(x, \theta) = \tfrac{1}{6}\theta^4 x^3 e^{-\theta x} \quad \text{for } x > 0.$$

Her er θ en ukjent positiv parameter.

(a) Finn forventningsverdien for X_i. Bruk for eksempel formelen $\int_0^\infty u^m e^{-u}\, du = m!$, som holder for alle naturlige tall m.

(b) Finn informasjonstallet $I(\theta)$ for denne modellen, altså det som kalles «informasjonen i en enkeltobservasjon».

(c) Vis at maximum likelihood-estimatoren blir

$$\widehat{\theta} = \frac{4}{\bar{X}_n},$$

der $\bar{X}_n$ er gjennomsnittet av de $n = 200$ observasjonene.

(d) De 200 observasjonene viser seg å ha gjennomsnitt $\bar{X}_n = 0.3796$ og empirisk standardavvik $s_n = 0.1801$. Lag et konfidensintervall for θ med konfidensgrad tilnærmet 95%.

(e) Forklar i passende detalj hvordan man kan konstruere et tilnærmet 95%-nivå konfidensintervall for θ ved såkalt parametrisk bootstrapping, der man simulerer pseudodatasett fra den estimerte modellen.

Oppgave 4

HVOR FORT SVØMMER EN HVAL? Det kommer selvsagt an på hvalen og dens mangslungne gjøremål, og er dessuten vanskelig å måle. Amerikanske og norske forskere har imidlertid gjort systematiske observasjonsstudier av blant annet grønlandshvalen *(Balanea mysticetus)*, og estimert hvalers hastighet. Et tilfeldig utvalg av disse fartsmålingene, i kilometer pr. time, er som følger:

```
4.440 7.479 1.390 6.090 4.170 5.319 0.870 0.450 5.470 3.270
```

Kaller man disse målingene $X_1, \ldots, X_{10}$ finner man

$$\bar{X} = 3.8948 \quad \text{og} \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 50.2284,$$

der $\bar{X}$ som vanlig betegner gjennomsnittet.

Din oppgave nå er å lage 90% konfidensintervall for μ og for σ, under den modellantagelse at målingene over er uavhengige og stammer fra den samme normalfordeling $\mathrm{N}(\mu, \sigma^2)$.

Til dette formål kan du bruke passende tall fra tabellen under. Den har tre kolonner med χ^2-kvantiler, for henholdsvis 9, 10, 11 frihetsgrader (degrees of freedom), og deretter tre kolonner med t-kvantiler, igjen for henholdsvis 9, 10, 11 frihetsgrader.

	χ^2-kvantiler:			t-kvantiler:		
quant	df=9	df=10	df=11	df=9	df=10	df=11
0.05	3.325	3.940	4.575	-1.833	-1.812	-1.796
0.10	4.168	4.865	5.578	-1.383	-1.372	-1.363
0.15	4.817	5.570	6.336	-1.100	-1.093	-1.088
0.20	5.380	6.179	6.989	-0.883	-0.879	-0.876
0.25	5.899	6.737	7.584	-0.703	-0.700	-0.697
0.30	6.393	7.267	8.148	-0.543	-0.542	-0.540
0.35	6.876	7.783	8.695	-0.398	-0.397	-0.396
0.40	7.357	8.295	9.237	-0.261	-0.260	-0.260
0.45	7.843	8.812	9.783	-0.129	-0.129	-0.129
0.50	8.343	9.342	10.341	0.000	0.000	0.000
0.55	8.863	9.892	10.920	0.129	0.129	0.129
0.60	9.414	10.473	11.530	0.261	0.260	0.260
0.65	10.006	11.097	12.184	0.398	0.397	0.396
0.70	10.656	11.781	12.899	0.543	0.542	0.540
0.75	11.389	12.549	13.701	0.703	0.700	0.697
0.80	12.242	13.442	14.631	0.883	0.879	0.876
0.85	13.288	14.534	15.767	1.100	1.093	1.088
0.90	14.684	15.987	17.275	1.383	1.372	1.363
0.95	16.919	18.307	19.675	1.833	1.812	1.796