Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Bo Lindqvist 975 89 418

EKSAMEN I FAG TMA4275 LEVETIDSANALYSE

Fredag 26. mai 2006

Tid: 09:00–13:00

Tillatte hjelpemidler:
Alle trykte og håndskrevne hjelpemidler. Alle kalkulatorer tillatt.

Sensur: 16. juni 2006.

BOKMÅL

Oppgave 1

Levetiden T for en komponent har fordeling gitt ved pålitelighetsfunksjonen (*reliability function*)

$$R(t;\alpha,\theta) = \frac{1}{1+\left(\frac{t}{\theta}\right)^{\alpha}} \qquad \text{for } t > 0, \tag{1}$$

der $\alpha > 0$ og $\theta > 0$ er parametre.

a) La p-kvantilen $t_p(\alpha,\theta)$ for T være gitt ved $P(T \le t_p(\alpha,\theta)) = p$ for $0 < p < 1$.

Vis at

$$\ln t_p(\alpha,\theta) = \ln\theta + \frac{1}{\alpha}\ln\frac{p}{1-p}$$

der ln er den naturlige logaritme.

Hva blir medianen i fordelingen til T?

Finn også uttrykk for første og tredje kvartil (dvs. Q1=$t_{0.25}$ og Q3=$t_{0.75}$). Hva er den praktiske fortolkning av disse?

b) Vis at hasardfunksjonen (*failure rate function*) til T kan skrives

$$z(t;\alpha,\theta) = \frac{\alpha t^{\alpha-1}}{\theta^\alpha + t^\alpha} \qquad \text{for } t > 0.$$

Vis at $z(t;\alpha,\theta)$ er avtagende i t for $\alpha \leq 1$.

For hvilke verdier av t er $z(t;\alpha,\theta)$ voksende, respektive avtagende, når $\alpha > 1$? For hvilke praktiske situasjoner vil et slikt forløp være rimelig?

c) Vis at (1) definerer en log-lokasjon-skala familie, dvs. at $Y = \ln T$ har kumulativ fordelingsfunksjon (*distribution function*) på formen

$$F_Y(y) = P(Y \leq y) = \Phi_0\left(\frac{y-\mu}{\sigma}\right) \qquad \text{for } -\infty < y < \infty$$

for en kumulativ fordelingsfunksjon Φ_0 og konstanter μ (lokasjon) og σ (skala).

Finn funksjonen Φ_0, og finn konstantene μ og σ uttrykt ved α og θ.

Hva kalles fordelingen med kumulativ fordelingsfunksjon Φ_0? Under hvilket navn er dermed fordelingen til T kjent?

Oppgave 2

Tolv motorer av samme type ble satt på test samtidig. En av motorene ble tatt ut av testen etter 200 kcycles fordi den skulle brukes i et annet forsøk. Tre motorer feilet ved 212, 445 og 792 kcycles, mens de gjenværende åtte motorene gikk uten å feile til 1100 kcycles, da testen ble avsluttet. (1 kcycle er 1000 cycles).

La T (enhet kcycles) være tidspunktet for feil for en motor av denne typen.

a) La $R(t) = P(T > t)$ være pålitelighetsfunksjonen for T. Beregn Kaplan-Meier (KM) estimatet for $R(t)$ ut fra de gitte resultatene og tegn den estimerte kurven i et diagram.

Hvilke forutsetninger bygger KM-estimatoren på? Er det rimelig å anta at disse er oppfylt her?

Finn et estimat for første kvartil (Q1) i fordelingen til T.

Hvorfor kan medianen og tredje kvartil (Q3) i fordelingen til T ikke estimeres ved hjelp av KM-estimatet?

I resten av oppgaven antas at T har fordelingen gitt ved (1) i Oppgave 1, der parametrene α og θ er ukjente.

Du kan bruke i det følgende at maximum likelihood estimatene for α og θ er gitt ved

$$\begin{aligned}\hat{\alpha} &= 1.2868 \\ \hat{\theta} &= 2286.5\end{aligned}$$

mens den inverse av den observerte informasjonsmatrise er gitt ved

$$\left\{I(\hat{\alpha},\hat{\theta})\right\}^{-1} = \begin{bmatrix} 0.44512 & -750.54 \\ -750.54 & 2.6638 \cdot 10^6 \end{bmatrix}$$

b) Bruk resultatene ovenfor til å lage tilnærmede 95% konfidensintervaller for α og θ. Begrunn valg av metode.

Estimer også første kvartil (Q1) ved å bruke de estimerte parameterverdiene og sammenlign med det ikke-parametriske estimatet i punkt (a) i denne oppgaven.

Hvorfor kan medianen og tredje kvartil (Q3) nå estimeres, og hva blir estimatene?

c) Man vil gjøre en grafisk sjekk av den parametriske modellen ved hjelp av et sannsynlighetsplott (*probability plot*).

Vis at punktene

$$\left(\ln t,\ \ln\left(\frac{1-R(t;\alpha,\theta)}{R(t;\alpha,\theta)}\right)\right)$$

ligger på en rett linje når t varierer, og sett opp ligningen for denne linjen.

Bruk dette til å tegne et sannsynlighetsplott basert på de gitte dataene, der den nevnte linjen er inntegnet ved hjelp av estimerte verdier for parametrene.

Kommenter resultatet kort.

Oppgave 3

En fotballkamp går over $\tau = 90$ minutter. La N være totalt antall scorede mål for de to lagene, og anta at scoringene skjer på tidspunkter $S_1, S_2, \ldots, S_N$ (enhet minutter).

En mulig modell er å anta at $S_1, S_2, \ldots$ danner en ikke-homogen Poisson-prosess (NHPP) med intensitetsfunksjon (*ROCOF*) $w(t)$.

Vi vil nå studere runde 5 i årets norske Tippeliga (spilt 3.-4. mai). Resultatene, med tidspunkter for scoringer, er gitt i tabellen nedenfor. (Til orientering er tidspunkter for scoringer for bortelaget merket med ^{b}. I oppgaven skilles imidlertid ikke mellom mål scoret av hjemme- og bortelaget).

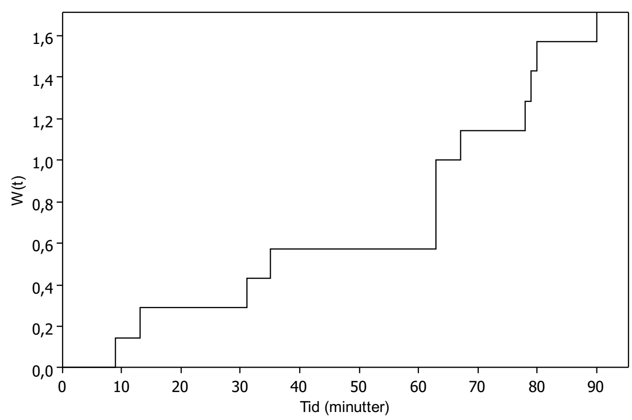
Det ses bort fra eventuell tilleggstid utover 90 minutter i kampene.

Hjemmelag-Bortelag	*Resultat*	*Scoringstidspunkter*			
HamKam-Lillestrøm	1-2	9	63^b	78^b	
Molde-Brann	0-2	31^b	90^b		
Sandefjord-Lyn	2-2	35	63^b	67	80^b
Stabæk-Rosenborg	1-1	63^b	79		
Start-Fredrikstad	0-0				
Tromsø-Odd Grenland	0-0				
Vålerenga-Viking	1-0	13			

Anta at scoringene i hver av de 7 kampene følger uavhengige NHPP med samme intensitetsfunksjon $w(t)$. La $W(t)$ være den tilsvarende kumulative intensitetsfunksjonen, $W(t) = \int_0^t w(u)du$.

a) Forklar hvordan funksjonen $W(t)$ kan estimeres fra de gitte dataene ved hjelp av Nelson-Aalen estimatoren. Du trenger ikke gjøre alle beregningene.

Et MINITAB-plott av den estimerte kurven for $W(t)$ er gitt i figuren nedenfor.



Hvilken fortolkning har $W(45)$?

Sett opp det nøyaktige estimatet for $W(45)$.

Beregn det tilhørende estimerte standardavvik.

b) Det blir påstått at intensiteten av scoringer er økende i løpet av en fotballkamp.

I hvilken grad gir MINITAB-plottet holdepunkt for en slik påstand?

Du skal nå bruke de gitte dataene til å undersøke holdbarheten av påstanden ved hjelp av en "pooled" Military Handbook-test.

Sett opp null-hypotesen og den relevante alternative hypotese i denne situasjonen.

Regn ut testobservatoren og bruk statistiske tabeller til å gi en konklusjon når signifikansnivået settes til 5%.

Gjør kort rede for at den TTT-baserte Military Handbook-testen i dette tilfellet vil gi den samme testobservatoren som "pooled"-versjonen beregnet ovenfor. Hvorfor blir det slik?

Anta til slutt at $w(t)$ ikke varierer med tiden, slik at vi kan sette $w(t) = \lambda$ der $\lambda > 0$ er en ukjent parameter.

c) Finn maksimum likelihood estimatet for λ basert på de gitte dataene, og beregn det tilhørende estimerte standardavvik.