Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

Kontakt under eksamen:
Rupali Akerkar (40 33 20 37)

EKSAMEN I TMA4275 LEVETIDSANALYSE

Onsdag 30. mai 2012
Tid: 09:00 – 13:00

Tillatte hjelpemidler:
Statistiske tabeller og formler (Tapir forlag)
Godkjent kalkulator (HP30S)
K. Rottmann: Matematisk formelsamling
Ett gult ark (A4 med stempel) med dine egne formler og notater

Sensur: 20. juni 2012

Oppgave 1

Et tilfeldig utvalg på $n = 100$ enheter ble satt under test på et bestemt tidspunkt, og testen ble kjørt til $t_c = 1000$ timer, da $R = 20$ enheter hadde sviktet. De rapporterte feiltidene ble notert. Vi kan betegne disse ved $t_1, \cdots, t_{20}$. Basert på fysisk kunnskap om sviktmekanismen, tror ingeniører at en en-parameter eksponentialfordeling kan brukes til å beskrive feiltiden. Den eksponensielle kumulative fordelingsfunksjonen (CDF) er gitt ved
$F(t, \theta) = 1 - \exp(-t/\theta)$.

a) Skriv ned et uttrykk for likelihood-funksjonen for θ for de gitte dataene.

b) Utled et uttrykk for maksimum likelihood estimatoren for θ basert på de gitte dataene.

Oppgave 2

Det er gjennomført en klinisk studie for å evaluere effekten av kjemoterapi for en bestemt krefttype. Etter å ha nådd en tilstand av komplett remisjon (bortfall av kreftsymptomer) gjennom behandling, ble pasientene som deltok i studien randomisert i to grupper. Den første gruppen fikk vedlikeholdskjemoterapi, mens den andre gruppen (kalt kontrollgruppen) ikke fikk dette. For en foreløpig analyse i løpet av studien var dataene som følger:
Lengde av komplett remisjon (i uker).

Vedlikeholdsgruppe: 9, 13, 13^+, 23, 24^+, 34, 45^+, 55, 161^+
Kontrollgruppe: 5, 13, 13, 16^+, 20, 21, 43, 45
+ indikerer sensurert observasjon.

a) Hva betyr det (generelt) at en observasjon er sensurert?
Hva er typiske årsaker (generelt) til at et datasett inneholder sensurerte observasjoner?

b) La $R_M(t)$ og $R_C(t)$ være overlevelsesfunksjonene for lengden av remisjon for henholdsvis vedlikeholdsgruppen og kontrollgruppen. Beregn Kaplan-Meier (KM) estimatorer for $R_M(t)$ og $R_C(t)$ og tegn dem i den samme grafen. Sammenlign KM-kurvene. Hvilken gruppe synes å ha den beste remisjonslengden (kreftfri overlevelse)?

c) Forklar hvordan man kan bruke logrank-testen for å teste nullhypotesen at remisjonslengden for pasienter i vedlikeholdsgruppen og kontrollgruppen har samme overlevelsesfordeling. Det forventede totale antallet hendelser (slutt på remisjonstilstand) for vedlikeholdsgruppen er 7.73 og for kontrollgruppen 4.27. Bruk disse resultatene til å beregne testobservatoren og utfør testen med signifikansnivå 5 %. Hva konkluderer du om hvorvidt de to gruppene har samme overlevelsesfordeling for remisjonslengden?

Man vil bruke en Cox-modell for dataene.

La den eneste kovariaten være behandling, representert ved variabelen x med verdier 1 for vedlikeholdsgruppen og 0 for kontrollgruppen.

d) Skriv ned en enkel Cox-modell som leder til en test for denne situasjonen. Formuler en nullhypotese og en alternativ hypotese for å teste om de kreftfrie overlevelsestidene er de samme for de to gruppene.
(Du bes ikke om å utføre testingen eller skrive ned den partielle likelihooden).

Denne testen ble utført i R og resultatet er gitt på neste side:

```
coxph(formula = Surv(time, status) ~ trt, data = data)

     coef exp(coef) se(coef)     x      p
trt -1.06       0.347    0.636  -1.67   0.096

Likelihood ratio test = 2.92  on 1 df, p=0.0873  n= 17, number of events= 12
```

Bruk denne informasjonen til å konkludere om den kreftfrie overlevelse er den samme for de to gruppene.

Oppgave 3

Betrakt en komponent hvor feiltiden T har overlevelses(pålitelighets)funksjon

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

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

a) La den $p$te kvantilen $t_p(\alpha,\theta)$ for T være gitt ved $P(T \leq t_p(\alpha,\theta)) = p$ 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 logaritmen.
Utled et uttrykk for medianen i fordelingen til T.
Finn også uttrykket for den første og den tredje kvantilen, ($Q1 = t_{0,25}$ og $Q3 = t_{0,75}$).

b) Vis at hasardfunksjonen (feilrate-funksjonen) for T er gitt ved

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

Oppgave 4 er på neste side.

Oppgave 4

Man har registrert feil på en spesiell ventil til tid $\tau = 900$ dager etter start. Følgende data ble observert (feiltider i dager etter start):

$270, 520, 700, 810, 860$

Det antas at feiltidene følger en ikke-homogen Poisson-prosess (NHPP).

a) Hvordan ville du estimere den kumulative intensiteten $W(t)$ når det ikke gjøres parametriske antagelser? Tegn den estimerte kurven for $W(t)$.
Indikerer plottet en trend i feilintensiteten? Gjennomfør Laplace-testen for å undersøke om det er en trend i feiltidene.
Hva er din konklusjon hvis du bruker 5% signifikansnivå?

b) Anta at ROCOF $w(t)$ for feilprosessen er

$$w(t) = e^{\alpha+\beta t} \quad \text{for} \quad t > 0 \tag{2}$$

der $-\infty < \alpha < \infty$ og $-\infty < \beta < \infty$ er ukjente parametrer.
Hva kalles denne modellen? Hva er den kumulative ROCOF for denne modellen?
Hva er svarene på de to foregående spørsmålene hvis $\beta = 0$?

c) Skriv ned log-likelihooden for modellen når $w(t)$ er gitt ved ligning (2). Bruk log-likelihooden til å skrive ned de ligningene som vil bestemme maximum likelihood estimatene (MLE) for α og β. (Du trenger ikke å løse disse ligningene.)

d) La $N(\tau)$ være antall feil inntil tid $\tau = 900$ dager. Vis at MLE for α tilfredsstiller

$$\hat{\alpha} = \ln\Big(\frac{N(\tau)\cdot\hat{\beta}}{\exp(\hat{\beta}\cdot\tau) - 1}\Big)$$

der $\hat{\beta}$ er MLE for β. Beregn $\hat{\alpha}$ for det gitte datasettet når du får vite at $\hat{\beta} = 0.003$.
Tegn den estimerte kurven for forventet antall feil som funksjon av tid i den samme grafen som plottet i (a).
Virker det som modellen i (2) gir en rimelig tilpasning til dataene?