UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK4290 — Parametrisk levetidsmodellering

Eksamensdag: Torsdag 12. juni 2014

Tid for eksamen: 14.30 – 18.30

Oppgavesettet er på 5 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Godkjent kalkulator

Kontroller at oppgavesettet er komplett før
du begynner å besvare spørsmålene.

Oppgave 1 Log-Laplace fordelingen

La $\Psi(w)$ og $\psi(w)$ være, henholdsvis, den kumulative fordelingsfunksjon og sannsynlighetstettheten for en stokastisk variabel W med verdiområde $(-\infty, \infty)$. Anta at $\psi(w) > 0$ for alle $w \in (-\infty, \infty)$.

1a

Forklar hvordan $\Psi(w)$ og $\psi(w)$ brukes til å definere en *log-lokasjon-skala* familie av fordelinger for en levetid T, indeksert av parametrer $-\infty < \mu < \infty$ og $\sigma > 0$. Hva kalles de to parametrene?

Finn den kumulative fordelingsfunksjon $F(t)$ og sannsynlighetstettheten $f(t)$ for T uttrykt ved funksjonene Ψ og ψ, og parametrene μ og σ.

I resten av oppgaven antas at W har fordeling gitt ved en standard *Laplace-fordeling* (også kalt *dobbel eksponensialfordeling*), dvs. det antas at

$$\psi(w) = \frac{1}{2} e^{-|w|} \text{ for } -\infty < w < \infty. \tag{1}$$

Du kan i det følgende bruke uten å vise det at den kumulative fordelingsfunksjon for W er

$$\Psi(w) = \begin{cases} \frac{1}{2} e^{w} & \text{for } w \leq 0 \\ 1 - \frac{1}{2} e^{-w} & \text{for } w > 0 \end{cases}$$

og at den momentgenererende funksjon for W er

$$M_W(s) \equiv E(e^{sW}) = \frac{1}{1 - s^2} \text{ for } |s| < 1.$$

(Fortsettes på side 2.)

1b

Den resulterende familie av fordelinger for T kalles *Log-Laplace fordelingen.* Vis at denne fordelingen kan reparametriseres med $\alpha > 0, \theta > 0$ slik at den kumulative fordelingsfunksjonen kan skrives

$$F(t) = \begin{cases} \frac{1}{2}\left(\frac{t}{\theta}\right)^{\alpha} & \text{for } 0 \le t \le \theta, \\ 1 - \frac{1}{2}\left(\frac{t}{\theta}\right)^{-\alpha} & \text{for } t > \theta. \end{cases} \tag{2}$$

Uttrykk α og θ som funksjoner av μ og σ.

1c

Finn uttrykk for tettheten $f(t)$ og hasardraten $z(t)$ for T under parametriseringen (2).

Skisser funksjonene $f(t)$ og $z(t)$ når $\alpha = \theta = 1$.

1d

Finn uttrykk for forventning og varians i fordelingen (2). Vis at forventningen er definert bare hvis $\alpha > 1$, mens variansen er definert bare hvis $\alpha > 2$.

(*Vink:* Du kan for eksempel først bruke den momentgenererende funksjon for W til å finne en formel for $E(T^r)$ for generell r.)

1e

Vis at medianen i fordelingen (2) er θ.

La t_p være $100p$-persentilen for fordelingen (2), slik at $F(t_p) = p$, for $0 < p < 1$. Finn t_p som funksjon av p.

(*Vink:* Skill mellom tilfellene $p \le 1/2$ og $p > 1/2$.

Oppgave 2 TTT-plott

La $T_1, \ldots, T_n$ være uavhengige (og usensurerte) observasjoner av en levetid T med kumulativ fordelingsfunksjon $F(t)$ og tetthet $f(t)$. Det antas at $f(t) > 0$ for alle $t > 0$, slik at $F(t)$ er strengt voksende overalt. Det antas også at $\mu = E(T)$ eksisterer.

La de ordnede levetidene være

$$T_{(1)} < T_{(2)} < \cdots < T_{(n)}.$$

Den empiriske fordelingsfunksjon $F_n(t)$ er definert ved

$$F_n(t) = \frac{i}{n} \text{ for } T_{(i)} \le t < T_{(i+1)};\ i = 0, 1, \ldots, n,$$

der $T_{(0)} = 0,\ T_{(n+1)} = \infty$.

(Fortsettes på side 3.)

2a

La $\mathcal{T}(t)$ være "Total Time on Test" til og med tid t, for $t > 0$.

Forklar kort hvorfor vi kan skrive, for $i = 1, \dots, n$,

$$\mathcal{T}(T_{(i)}) = nT_{(1)} + (n-1)(T_{(2)} - T_{(1)}) + \ldots + (n-i+1)(T_{(i)} - T_{(i-1)}).$$

Kan du finne et enklere uttrykk for $\mathcal{T}(T_{(n)})$?

Vis at for hver $i = 1, 2, \dots, n$ har vi

$$\frac{1}{n}\mathcal{T}(T_{(i)}) = \int_0^{F_n^{-1}(i/n)} (1 - F_n(t))dt.$$

(*Vink:* For en generell kumulativ fordelingsfunksjon $H(x)$ defineres

$$H^{-1}(u) = \inf\{x : H(x) \geq u\} \text{ for } 0 \leq u \leq 1,$$

slik at $F_n^{-1}(i/n) = T_{(i)}$ for $i = 1, \dots, n$.)

2b

Anta at $n \to \infty$ og at samtidig $i \to \infty$ på en slik måte at $i/n \to u$, der $0 \leq u \leq 1$. Det kan vises (du skal ikke gjøre det) at vi da har konvergensen

$$\int_0^{F_n^{-1}(i/n)} (1 - F_n(t))dt \to \int_0^{F^{-1}(u)} (1 - F(t))dt \equiv G_F(u)$$

med sannsynlighet 1, uniformt i $0 \leq u \leq 1$.

TTT-plottet basert på observasjonene $T_1, \dots, T_n$ er som kjent gitt ved punktene

$$\left(\frac{i}{n}, \frac{\mathcal{T}(T_{(i)})}{\mathcal{T}(T_{(n)})}\right), \; i = 1, \dots, n.$$

Gjør rede for at TTT-plottet, når $n \to \infty$ og $i/n \to u$, konvergerer mot kurven definert ved

$$\left(u, \frac{G_F(u)}{\mu}\right), \quad 0 \leq u \leq 1.$$

2c

Anta i dette punktet at T er eksponensialfordelt med forventning $\theta > 0$, dvs. at T har kumulativ fordelingsfunksjon

$$F(t) = 1 - e^{-t/\theta} \text{ for } t > 0.$$

Regn ut funksjonen $G_F(u)$. Hvilken kurve vil TTT-plottet dermed konvergere mot når observasjonene kommer fra en eksponensialfordeling?

Hvorfor er dette et nyttig resultat ved praktisk bruk av TTT-plott?

(Fortsettes på side 4.)

2d

La igjen T ha en generell fordeling. Vis ved å derivere $G_F(u)$ at

$$G'_F(F(t)) = \frac{1}{z(t)} \text{ for } t > 0, \tag{3}$$

der $z(t)$ er hasardraten til T.

(*Vink:* Du kan bruke formelen for derivasjon av den inverse til en funksjon $k(x)$:

$$[k^{-1}]'(x) = \frac{1}{k'(k^{-1}(x))} \text{).}$$

Gjør rede for hvordan resultatet (3) kan brukes til å fortolke TTT-plott med hensyn på voksende eller avtagende hasardrate.

Oppgave 3 Reparerbart system

Vi betrakter et reparerbart system som opererer fra tid $t = 0$. Ved feil blir systemet reparert og satt tilbake i drift umiddelbart.

La $N(t)$ for $t > 0$ betegne antall feil i tidsintervallet $(0, t]$. Det antas at $N(t)$ er en ikke-homogen Poisson-prosess (NHPP) med intensitetsfunksjon $w(t)$ og kumulativ intensitetsfunksjon $W(t) = \int_0^t w(u)du$.

3a

Hvilke egenskaper definerer $N(t)$ som en NHPP?

Gjør kort rede for hva det vil si at en reparasjon er henholdsvis minimal og perfekt. Hvilken av disse to typene av reparasjoner modelleres ved NHPP? Begrunn svaret.

3b

Anta nå at funksjonene $w(t)$ og $W(t)$ modelleres ved parametriske funksjoner $w(t; \theta)$ og $W(t; \theta)$, og at man ønsker å gjøre statistisk inferens om θ.

Systemet observeres på tidsintervallet $[0, \tau]$, der det kun registreres antall feil i disjunkte tidsintervaller definert ved oppdelingen

$$h_0 = 0 < h_1 < h_2 < \cdots < h_r = \tau.$$

La D_i være antall observerte feil i intervallet $(h_{i-1}, h_i]$, for $i = 1, 2, \ldots, r$. Forklar hvorfor $D_1, \ldots, D_r$ er uavhengige og Poisson-fordelte, med

$$E(D_i) = W(h_i; \theta) - W(h_{i-1}; \theta).$$

Bruk dette til å vise at likelihood-funksjonen for θ når observasjonene er gitt som $d_1, \ldots, d_r$, kan skrives

$$L(\theta; d_1, \ldots, d_r) = \left\{ \prod_{i=1}^{r} \frac{[W(h_i; \theta) - W(h_{i-1}; \theta)]^{d_i}}{d_i!} \right\} e^{-W(\tau; \theta)}. \tag{4}$$

(Fortsettes på side 5.)

3c

Anta til slutt at man likevel har registrert de eksakte feiltidspunktene $s_1, s_2, \ldots, s_n$ i intervallet $[0, \tau]$.

Vis hvordan likelihood-funksjonen i (4) leder til følgende uttrykk for likelihood-funksjonen for dette tilfellet:

$$L(\theta; s_1, \ldots, s_n) = \left\{ \prod_{\ell=1}^{n} w(s_\ell; \theta) \right\} e^{-W(\tau;\theta)}.$$

SLUTT