Første obligatoriske oppgave i STK3100/STK4100
Høst 2007

Utlevering: Fredag 21. September

Innleveringsfrist: Fredag 5. oktober, kl. 14:30

Besvarelsen innleveres ved ekspedisjonen i 7. etasje, Niels Henrik Abels hus

Dette er det første settet med obligatoriske innlevering i STK31000 / STK4100 høsten 2007. Oppgavesettet består av en oppgave. Det er valgfritt om du vil skrive besvarelsen for hånd eller om du vil bruke et tekstbehandlingsprogram. Der du bruker R (eller et annet program), må utskrifter legges ved eller limes inn. Hvis flere studenter samarbeider om å løse oppgavene, må likevel hver student levere sin selvstendige besvarelse. Det må gå fram av besvarelsen hvem du har samarbeidet med. Se ellers "Regelverk for obligatoriske oppgaver" som er gitt pa kursets hjemmeside.

Obligatorisk oppgave:

Anta at Y er Pareto-fordelt, dvs. har tetthet

$$f(y;\theta) = \begin{cases} \frac{\theta}{y^{\theta+1}} & \text{for } y \geq 1 \\ 0 & \text{for } y < 1 \end{cases}$$

I STK3100-Oppgave 8 (H07) ble det vist at vi kan uttrykke $f(y;\theta) = \exp(a(y)b(\theta) + c(\theta))$ med $a(y) = \log(y)$, $b(\theta) = -(\theta+1)$ og $c(\theta) = \log(\theta)$.

a) Anta $Y \sim f(y;\theta_0)$. Bruk representasjon som eksponensiell klasse til vise at $\mathrm{E}[\log(Y)] = 1/\theta_0$.

b) Vis at

$$l_0(\theta) = \mathrm{E}[\log(f(Y;\theta)] = \log(\theta) - \frac{\theta+1}{\theta_0}$$

$$u_0(\theta) = \mathrm{E}[\frac{\partial \log(f(Y;\theta)}{\partial\theta}] = \frac{1}{\theta} - \frac{1}{\theta_0}$$

samt at $l_0(\theta)$ maksimeres i θ_0 og at $u_0(\theta_0) = 0$.

c) Vis at Y har kumulativ fordelingsfunksjon $F(y;\theta_0) = 1 - y^{-\theta_0}$ og foreslå en `R`-kommando for å trekke uavhengige Y_i fra Pareto-fordelingen.

d) Anta $Y_i \sim f(y;\theta_0), i = 1,\ldots,n$. Finn uttrykk for log-likelihood $l(\theta)$ og score-funksjonen $U(\theta)$.

e) Plot $l_0(\theta)$ med $\theta_0 = 0.5$. Simuler $n = 10$ Pareto-fordelte variable med $\theta_0 = 0.5$ og tegn inn $l(\theta)/n$. Gjenta simuleringen 9 ganger og tegn påny inn $l(\theta)/n$.

f) Plott også $u_0(\theta)$ og de tilhørende gjennomsnittlige scorefunksjonene $U(\theta)/n$.

g) Gjenta simuleringene og lag tilsvarende plott med $n = 100$.

h) Finn MLE for θ samt dens tilnærmede fordeling.

i) Simuler $n = 10$ Pareto-fordelte-variable med $\theta_0 = 0.5$ og beregn MLE. Gjenta simuleringen 999 ganger og tegn histogram over estimatene. Er MLE tilnærmet normalfordelt?

j) Gjenta simulering med $n = 100$. Er normaltilnærmingen bedre nå?