Første obligatoriske oppgave i STK3100 Høst 2010

Utlevering: Mandag 20. September

Innleveringsfrist: Torsdag 7. oktober, kl. 14:30

Besvarelsen leveres i kassa for obligatoriske oppgaver i gangen i 7. etasje, Niels Henrik Abels hus.

Dette er det første settet med obligatoriske innlevering i STK31000 høsten 2010. 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 på kursets hjemmeside.

Obligatorisk oppgave:

Denne oppgaven dreier seg om testing av tosidige helspesifiserte hypoteser for en skalar parameter basert på likelihooden. Det finnes flere prinsipper: Sannsynlihhetskvote test, Wald test og scoretest. Oppgaven går ut på å sammenligne dem for i en modell for en binomisk fordelt variabel.

Anta at $Y \sim \mathrm{Bin}(n, p)$, dvs. binomisk fordelt med n forsøk og suksess-sannsynlighet p.

a) Sett opp loglikelihood $l(p)$, scorefunksjon $U(p)$ og forventet informasjon $\mathcal{J}(p)$ basert på Y.

b) Angi tilnærmet fordeling for MLE $\hat{p} = Y/n$ og argumenter for at test-observator for Wald-testen for $\mathrm{H}_0 : p = p_0$ gis ved

$$Z_W(p_0) = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{n}$$

(slik at testobservatoren er tilnærmet standardnormalfordelt under nullhypotesen).

c) Vis at scoretesten for samme nullhypotese gis ved testobservator

$$Z_U(p_0) = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}.$$

og argumenter for at denne er tilnærmet standardnormalfordelt under H_0.

d) Vis at testobservator for likelihood ratio testen av $\mathrm{H}_0 : p = p_0$ kan skrives
$$Z_{LR}^2(p_0) = 2[Y \log(\frac{\hat{p}}{p_0}) + (n - Y) \log(\frac{1-\hat{p}}{1-p_0})]$$
Angi tilnærmet fordelingen for $Z_{LR}^2(p_0)$

e) Med $n = 100$ og $Y = 30$ utfør og sammenlign testene av $\mathrm{H}_0 : p = 0.5$.

f) Med $n = 100$ og $Y = 5$ utfør og sammenlign testene av $\mathrm{H}_0 : p = 0.15$.

g) Vi kan for hver av observatorene i b)-d) generere tilnærmede 95% konfidensintervall for p som intervallet $\{p_0 : Z^2(p_0) < 3.84\}$. Forklar hvorfor.

h) Vis at med $\mathrm{se}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$ blir konfidensintervallet basert på Wald-observatoren $Z_W(p_0)$ lik $\hat{p} \pm 1.96\mathrm{se}(\hat{p})$.

i) Vis at intervallet basert på score-observatoren $Z_U(p_0)$ gis som løsning av en 2.gradslikning. (Fasit:
$$< \hat{p}_L, \hat{p}_U >= \frac{y + 1.92}{n + 3.84} \pm \frac{1.96\sqrt{n\hat{p}(1-\hat{p}) + 3.84/4}}{n + 3.84}.)$$

j) Beregn konfidensintervallene med $n = 100$ og $Y = 30$.

(For intervallet basert på likelihood ratio observatoren $Z_{LR}^2(p_0)$ finnes ikke eksplisitte formler for intervallgrensene, men du kan lese dem av et plott over $(p_0, Z_{LR}^2(p_0))$).

k) Beregn konfidensintervallene med $n = 100$ og $Y = 5$.

Vi kan alternativt parametrisere binomiske fordelinger med den kanoniske parameteren $\theta = \log(p/(1-p))$ (altså log-odds).

l) Vis at score- og likelihood ratio testene ikke påvirkes av parametrisering.

m) Angi den tilnærmede fordeling for MLE $\hat{\theta}$ og vis at standardfeilen for $\hat{\theta}$ kan uttrykkes som $\sqrt{1/Y + 1/(n-Y)}$.

n) Sett opp (det tilnærmede) 95% Wald-intervallet $< \hat{\theta}_L, \hat{\theta}_U >$ for θ. Forklar hvorfor
$$< \frac{\exp(\hat{\theta}_L)}{1 + \exp(\hat{\theta}_L)}, \frac{\exp(\hat{\theta}_U)}{1 + \exp(\hat{\theta}_U)} >$$
blir et (tilnærmet) 95% konfidensintervall for p.

o) Beregn dette siste intervallet med $n = 100$, $Y = 30$ og $Y = 5$. Sammenlign med de andre konfidensintervallene.