Institut for Matematik
AARHUS UNIVERSITY

Introduktion til Sandsynlighedsteori og Statistik
Markus Kiderlen
November 3, 2023

UGESEDDEL 11

Forelæsningerne i Uge 9: Vi har gennemgået til og med Afsnit 8.2.3, men mangler Eksempel 8.10.2 og 8.11.

Forelæsningerne i Uge 10:

1. **Forelæsning 19 (7. november):** Vi går videre med Kapitel 8.2.4 og 8.3, hvor jeg regner med at vi når til og med Theorem 8.3 på side 463 (Afsnit 8.3.3).

2. **Forelæsning 20 (9. november):** Jeg regner med at vi når at gennemgå til og med Afsnit 8.4.2 (side 480).

Teoretiske øvelser i Uge 11 (13.-19. november): Opgaverne 1.-3. i Del 1 og 1.-2. i Del 2 kan regnes efter forelæsningen tirsdag d. 7. november og de resterende opgaver kan regnes efter forelæsningen torsdag d. 9. november.

Del 1:

1. Sektion 8.6: Øvelse 11.

2. Sektion 8.6: Øvelse 13.

3. Sektion 8.6: Øvelse 14.

Del 2:

1. Øvelse A (nedenfor).

2. ~~Eksamen, Sommer 2018 (reeksamen), Opgave 6.~~
   Eksamen, Vinter 2022/2023, Opgave 5
   (dog ikke sidste del i 5.3 vdr. hypotesetest).

3. Sektion 8.6: Øvelse 16.

4. Sektion 8.6: Øvelse 17.

Øvelse A: Lad $X_1, \dots, X_n$ betegne en stikprøve hvor $X_i \sim Exponential(\lambda)$ og $\lambda > 0$ er en ukendt parameter. Som sædvanligt betegner $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ stikprøvegennemsnittet. I denne opgave vil vi vise at et $(1-\alpha)100\%$ konfidensinterval for λ er givet ved

$$\left[\frac{g_{1-\alpha/2,n}}{\overline{X}}, \frac{g_{\alpha/2,n}}{\overline{X}}\right] \tag{1}$$

hvor $g_{p,n}$ betegner $(1-p)$-fraktilen for en $Gamma(n,n)$ fordeling. Med andre ord er $g_{p,n} = F_X^{-1}(1-p)$, hvor F_X^{-1} angiver den inverse fordelingsfunktionen for $X \sim Gamma(n,n)$.

(a) Find fordelingen for $\overline{X}$.

Hint: Sætning B, ugeseddel 8.

(b) Lad $Q = \lambda\overline{X}$ og vis at Q er en pivot.

(c) Vis at $P(g_{1-\alpha/2,n} \leq Q \leq g_{\alpha/2,n}) = 1 - \alpha$.

(d) Vis at (1) er et $(1-\alpha)100\%$ konfidensinterval for λ.

Afleveringsopgave 11: I denne opgave antager vi at levetiden (målt i år) på telefoner er eksponentialfordelt, samt at vi har observeret levetiden på 10 telefoner. Vi er interesseret i at teste nulhypotesen at middellevetiden μ for telefoner er 2 år, mod den alternative hypotese at den gennemsnitlige levetid er skarpt mindre end 2 år.

Hint: Benyt Sætning B. på ugeseddel 8.

(a) Opstil den ovenfor beskrevne model og de tilhørende hypoteser matematisk, samt relater μ til parameteren i eksponentialfordelingen.

Lad W betegne gennemsnittet af vores 10 observationer og lad $c > 0$ være et positivt tal. Vi vil nu lave et eksakt (dvs. ikke asymptotisk) test for nulhypotesen som følger: Accepter nulhypotesen hvis $W \geq c$, og forkast ellers.

(b) Find værdien af c der gør at vores test har signifikansniveau $\alpha = 0.05$.

(c) Antag nu at vi har observeret levetiden på 1000 telefoner. Find værdien af c der gør at vores test har igen signifikansniveau $\alpha = 0.05$, og sammenlign resultatet med det du fik i spørgsmål (b).

Ugens udfordring: Denne uges udfordring beskæftiger sig med Cramér–Rao's ulighed der giver en nedre grænse på variansen for enhver unbiased estimator. Betragt en klasse af fordelinger hvis PDF er givet ved $f(x;\theta)$ hvor $\theta \in \mathbb{R}$ er en ukendt parameter. (Notationen indikerer at PDF'en afhænger af parameteren θ). *Fisher informationen* for fordelingen med hensyn til parameteren θ er givet ved

$$\mathcal{I}(\theta) = E\left[(h(X))^2\right],$$

hvor den stokastiske variable X har PDF $f(x;\theta)$ og $h(x) = \frac{\partial \log f(x;\theta)}{\partial \theta}$. Fisher informationen angiver hvor meget information om θ der er i den stokastiske variabel X.

(a) Udregn Fisher information for normalfordelingen $N(\mu, \sigma^2)$, hvor μ er vores ukendte parameter (svarer til θ ovenfor), og $\sigma^2 > 0$ er en kendt parameter.

Lad $X_1, \ldots, X_n$ betegne en random sample med n observationer. Ifølge Cramér–Rao's ulighed gælder der at enhver unbiased estimat $\hat{\Theta}$ for θ opfylder uligheden:

$$\mathrm{Var}(\hat{\Theta}) \geq \frac{1}{n\mathcal{I}(\theta)}, \tag{2}$$

vor $\mathcal{I}(\theta)$ betegner Fisher informationen for θ. Hvis der gælder lighed i ligning (2) siges estimatoren $\hat{\Theta}$ at være *efficient*; dvs. den har så lav en varians som det er muligt.

(b) Vis at gennemsnittet $\overline{X}$ er et efficient estimat for middelværdien μ for en normaltfordelt stikprøve $X_1, \ldots, X_n$ med $X_i \sim N(\mu, \sigma^2)$ med kendt varians $\sigma^2 > 0$ og ukendt middelværdi μ.

Cramér–Rao's ulighed er relateret til Heisenbergs ubestemthedsrelation ('uncertainty principle') i kvantefysik.