STK1110 Høsten 2021

Repetisjon om konfidensintervaller

Tilsvarer Avsnitt 8.1 og 8.2

Ingrid Hobæk Haff
Matematisk institutt
Universitetet i Oslo

Normalfordelte data med kjent varians

- Vi antar at vi har observasjoner $x_1, \ldots, x_n$ av de stokastiske variablene $X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$.
- Vi ønsker å konstruere et konfidensintervall for μ, som er ukjent, og vi antar først at σ^2 er kjent (selv om det er urealistisk).
- Vi vet at $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, slik at $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.
- For å lage et 95% konfidensintervall for μ, bruker vi

$$\mathrm{P}\left(-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

- Ved omforming av ulikhetene får vi

$$\mathrm{P}\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Konfidensintervall for μ

- Da inneholder det **stokastiske** intervallet
$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right),$$
μ med 95% sannsynlighet.
- Når vi setter inn de observerte verdiene $x_1, \ldots, x_n$, får vi et 95% **konfidensintervall** for μ, gitt ved:
$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right),$$
- Vi kan ikke lenger si at μ ligger innenfor intervallet med 95% sannsynlighet, for det er ikke lenger stokastisk.

Konfidensintervall for μ (forts.)

- I stedet må vi tolke det slik: dersom en gjentar øvelsen over for mange observerte datasett av samme størrelse, vil rundt 95% av de resulterende konfidensintervallene inneholde den sanne verdien av μ.
- Konfidensintervaller med andre **konfidensnivåer**, som 90%, 99%, eller mer generelt $100 \cdot (1-\alpha)\%$, får vi ved å bytte ut $1.96 = z_{0.05/2}$ med $z_{\alpha/2}$, der z_α er øvre $\alpha/2$-kvantilen i standard normalfordeling:

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

Generell framgangsmåte for å konstruere konfidensintervaller

- Anta at vi har stokastiske variable $X_1, \ldots, X_n$, og at vi vil finne et $100 \cdot (1-\alpha)\%$ konfidensintervall for en parameter θ.
- Anta videre at vi har en observator $h(X_1, \ldots, X_n; \theta)$, som avhenger av θ, men hvis fordeling ikke avhenger av θ.
- La a og b være henholdvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantilen i fordelingen til $h(X_1, \ldots, X_n; \theta)$.
- Da har vi:

$$\mathrm{P}(a \leq h(X_1, \ldots, X_n; \theta) \leq b) = 1 - \alpha.$$

Generell framgangsmåte for å konstruere konfidensintervaller (forts.)

- Så omformer vi ulikhetene:

$$a \leq h(X_1, \ldots, X_n; \theta) \leq b \iff L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n).$$

- Da er $\mathrm{P}(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n)) = 1 - \alpha$.
- For å få et $100 \cdot (1-\alpha)\%$ konfidensintervall for θ setter vi inn de observerte verdiene og får:

$$(L(x_1, \ldots, x_n), U(x_1, \ldots, x_n)).$$

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der μ er kjent. Vi ønsker å lage et konfidensintervall for σ^2.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} Bernoulli(p)$. Vi ønsker å lage et konfidensintervall for p.

Stort utvalg uif data med ukjent varians

- Vi antar at vi har observasjoner $x_1, \ldots, x_n$ av de stokastiske variablene $X_1, \ldots, X_n$, som er uif med forventning μ og varians σ^2, som begge er ukjent.
- I henhold til **sentralgrenseteoremet** er da $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{tiln.}}{\sim} N(0,1)$ når n er tilstrekkelig stor.
- Dersom σ^2 er kjent, kan vi lage konfidensintervall for μ på samme måte som før.
- Ellers kan vi bruke at (det kan vises) $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{\text{tiln.}}{\sim} N(0,1)$, med $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, når n er stor nok.

Stort utvalg uif data med ukjent varians (forts.)

- Vi har da:

$$\mathrm{P}\left(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq z_{\alpha/2}\right) \approx 1-\alpha.$$

- Ved omforming av ulikhetene får vi

$$\mathrm{P}\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right) \approx 1-\alpha.$$

- Det betyr at et tilnærmet $100 \cdot (1-\alpha)\%$ konfidensintervall for μ er gitt ved:

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right).$$