Ekstraoppgaver til STK1100

Sjette sett – april 2005

Oppgave E11

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte stokastiske variabler (dvs. alle har samme fordeling) med forventningsverdi μ og varians σ^2. Tenk deg at vi *ikke* kjenner σ^2, og at vi derfor vil anslå (eller estimere) verdien til σ^2 ut fra observerte verdier av X_i-ene. Vi skal i denne oppgaven se hvordan det kan gjøres.

a) La $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. Vis at $E(\hat{\sigma}^2) = \sigma^2$.
(Vink: Husk definisjonen av varians.)

Resultatet i punkt a viser at vi i det lange løp "i gjennomsnitt" vil anslå variansen riktig ved å bruke $\hat{\sigma}^2$.

Men for at vi skal kunne regne ut verdien til $\hat{\sigma}^2$ når vi har observert X_i-ene, må vi kjenne forventingsverdien μ. I praktiske situasjoner vil det bare unntaksvis være tilfellet at μ er kjent. Når μ ikke er kjent, er det rimelig å bruke $\overline{X} = \sum_{i=1}^{n} X_i/n$ til å anslå (estimere) μ. Setter vi $\overline{X}$ inn for μ i uttrykket for $\hat{\sigma}^2$, får vi $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Vi vil finne forventningen til $\tilde{\sigma}^2$.

b) Vis at $\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$.

c) Vis at $E(X_i^2) = \sigma^2 + \mu^2$.

d) Vis at $E(\overline{X}^2) = \sigma^2/n + \mu^2$.
(Vink: Vis først at $E(\overline{X}) = \mu$ og $\mathrm{Var}(\overline{X}) = \sigma^2/n$.)

e) Vis at $E\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = (n-1)\sigma^2$.
(Vink: Bruk resultatene i punktene b, c og d.)

f) Vis at $E(\tilde{\sigma}^2) = (n-1)\sigma^2/n$.

Resultatet i punkt f viser at vi i det lange løp "i gjennomsnitt" vil anslå variansen litt for lavt hvis vi bruker $\tilde{\sigma}^2$. For å rette opp dette bruker vi i stedet den *empiriske variansen* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

g) Vis at $E(S^2) = \sigma^2$.

Resultatet i punkt g viser at vi i det lange løp "i gjennomsnitt" vil anslå variansen riktig ved å bruke S^2. Det er en viktig grunn til at vi dividerer med $n-1$ i stedet for n i uttrykket for empirisk varians og empirisk standardavvik (jf. notatet om beskrivende statistikk).