Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Brukerkurs i statistikk, ST0103

7. desember, 2012

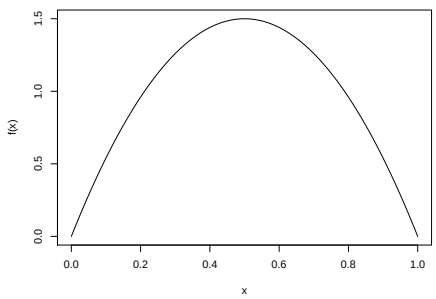
Løsningsforslag

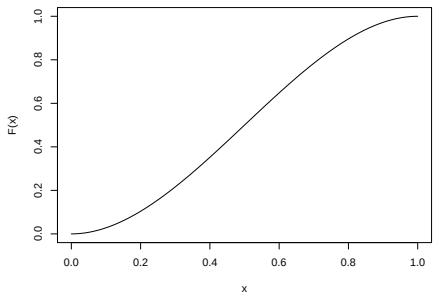
Oppgave 1

a) Kumulativ fordeling bli

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = 6\int_{0}^{x} t - t^2 dx = 6\left[\frac{1}{2}t^2 - \frac{1}{3}t^3\right]_0^x = 3x^2 - 2x^3 \tag{1}$$

for $x \geq 0$.





b)

$$P(X > 1/2) = 1 - F(1/2) = 1/2 \tag{2}$$

$$\begin{aligned} P(X > 3/4 | X > 1/2) &= \frac{P(X > 3/4 \cap X > 1/2)}{P(X > 1/2)} \\ &= \frac{P(X > 3/4)}{P(X > 1/2)} = \frac{1 - F(3/4)}{1 - F(1/2)} \\ &= \frac{1 - 27/32}{1 - 1/2} = 5/16. \end{aligned} \tag{3}$$

c) Siden $f(x)$ er symmetrisk omkring $x = 1/2$ er forventningen til X, $EX = 1/2$. Videre blir

$$E(X^2) = \int_0^1 x^2 6x(1-x)dx = 6\int_0^1 x^3 - x^4 dx = 6\left[\frac{1}{4}x^4 - \frac{1}{5}x^5\right]_0^1 = \frac{3}{10} \tag{4}$$

slik at $\operatorname{Var} X = E(X^2) - (EX)^2 = 3/10 - (1/2)^2 = 1/20$.

Oppgave 2

a) Antall ganger healeren velger rett svar X vil være binomisk fordelt med parametere $n = 30$ og p. Dette forutsetter også at forsøkene er uavhengige.

b) Vi ønsker å teste

$$H_0 : p = p_0 \tag{5}$$

versus

$$H_1 : p \neq p_0 \tag{6}$$

hvor $p_0 = 1/2$, altså nullhypotesen at healeren gjetter helt uavhengig av hvor Andreas har plassert sin hånd og dermed får rett med sannsynlighet $1/2$ versus alternativ hypotese at healeren på en eller annen måte påvirkes av plasseringen til Andreas' hånd ved at sannsynligheten for å gjette riktig enten forhøyes eller reduseres i forhold til $p_0 = 1/2$.

Under H_0 er testobservatoren

$$Z = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/30}} \tag{7}$$

tilnærmet $N(0,1)$ og vi forkaster H_0 hvis $Z < -z_{0.025} = -1.96$ eller $Z > 1.96$. Observert verdi av Z blir

$$Z = \frac{16/30 - 1/2}{\sqrt{1/2(1-1/2)/30}} = 0.36 \tag{8}$$

og vi beholder dermed H_0.

c) Testens p-verdi basert på normaltilnærming blir

$$P(Z \geq 0.36 \cup Z \leq -0.36) \approx 2G(-0.36) = 0.719 \tag{9}$$

og beregnet ut i fra at testobservatoren er en lineærtransformasjon av en diskret fordelt binomisk fordelt variabel blir p-verdien

$$\begin{aligned} P(Z \geq 0.036 \cup Z \leq -0.036) &= P(X \geq 16 \cup X \leq 14) \\ &= 1 - P(X = 15) \\ &= 1 - \binom{30}{15}(1/2)^{15}(1-1/2)^{30-15} \\ &= 0.856. \end{aligned} \tag{10}$$

Testens p-verdi er sannsynligheten for den observerte verdien eller mer ekstreme verdier av testobservatoren i favør av H_1 forutsatt at H_0 er sann. Siden denne sannsynligheten er hele 0.85 forutsatt av H_0 er riktig har vi med andre ord, dersom vi er skeptikere og tror på H_0, ikke noen grunn til å være særlig overrasket over utfallet av forsøket.

Oppgave 3

a) Fordi differansen Y er en lineærkombinasjon av normalfordelte variable U og V blir Y normalfordelt med forventning

$$E(Y) = E(U - V) = EU - EV = \mu - \mu = 0 \tag{11}$$

og varians

$$\mathrm{Var}(Y) = \mathrm{Var}(U - V) = \mathrm{Var}(U) + (-1)^2\,\mathrm{Var}(V) = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2. \tag{12}$$

Sannsynlighetstetthetsfunksjonen til Y er dermed

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}. \tag{13}$$

b) Betrakter vi σ^2 og ikke σ som den variable blir likelihoodfunksjonen

$$L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_i^2}{2\sigma^2}} = (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2), \tag{14}$$

log-likelihoodfunksjonen

$$\ln L(\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2, \tag{15}$$

og denne har sitt maksimum når

$$\begin{aligned} \frac{d}{d(\sigma^2)} \ln L &= 0, \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} y_i^2 &= 0, \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} y_i^2. \end{aligned} \tag{16}$$

Altså er sannsynlighetsmaksimeringsestimator av σ^2, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$.

c) Siden hver Y_i har forventning $EY_i = 0$ blir $E(Y_i^2) = E[(Y_i - 0)^2] = \operatorname{Var} Y_i = \sigma^2$ og

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^{n} E(Y_i^2) = \frac{1}{n} n\sigma^2 = \sigma^2 \tag{17}$$

slik at $\hat{\sigma}^2$ altså er forventningsrett for σ^2.

Estimatet av σ blir $\hat{\sigma}^2 = 0.096$.

d) Når $n\hat{\sigma}^2/\sigma^2$ er kji-kvadrat med n frihetsgrader har vi at

$$\begin{aligned} 1 - \alpha &= P\left(\chi^2_{1-\alpha/2,n} < \frac{n\hat{\sigma}^2}{\sigma^2} < \chi^2_{\alpha/2,n} \right) \\ &= P\left(\frac{1}{\chi^2_{1-\alpha/2,n}} > \frac{\sigma^2}{n\hat{\sigma}^2} > \frac{1}{\chi^2_{\alpha/2,n}} \right) \\ &= P\left(\frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n}} > \sigma^2 > \frac{n\hat{\sigma}^2}{\chi^2_{\alpha/2,n}} \right) \end{aligned} \tag{18}$$

slik at

$$\left(\frac{n\hat{\sigma}^2}{\chi^2_{\alpha/2,n}}, \frac{n\hat{\sigma}^2}{\chi^2_{1-\alpha/2,n}} \right) = \left(\frac{5 \cdot 0.096}{12.83}, \frac{5 \cdot 0.096}{0.83} \right) = (0.037, 0.577) \tag{19}$$

er et 95%-konfidensintervall for σ^2.