Norges teknisk–naturvitenskapelige universitet
Institutt for matematiske fag

English

TMA4245 STATISTIKK - Proposed Solution

5. august 2009

Oppgave 1

We know from the exercise that

$$T_{A1} \sim N(60, 5^2), \quad T_{B2} \sim N(65, 5^2)$$

where T_{A1} denote the result for Asbjørn if he uses skis 1 and T_{B2} Bernt's result if he uses skis 2.

a)

$$P(T_{A1} < 55) = P\left(\frac{T_{A1} - 60}{5} < \frac{55 - 60}{5}\right) = P(Z < -1) = 0.1587,$$

where Z is a standard normal random variable, $Z \sim N(0, 1)$.

$$\begin{aligned} P(55 < T_{A1} < 60) &= P\left(\frac{55 - 60}{5} < \frac{T_{A1} - 60}{5} < \frac{60 - 60}{5}\right) \\ &= P(-1 < Z < 0) = P(Z < 0) - P(Z < -1) \\ &= 0.5 - 0.1587 = 0.3413. \end{aligned}$$

b)

$$P(T_{A1} < T_{B2}) = P(T_{A1} - T_{B2} < 0).$$

We know that the random variable $Y_1 = T_{A1} - T_{B2}$ is normally distributed with mean

$$E(Y_1) = E(T_{A1}) - E(T_{B2}) = 60 - 65 = -5$$

and variance (T_{A1} and T_{B2} are independent)

$$V(Y_1) = V(T_{A1}) + V(T_{B2}) = 5^2 + 5^2 = 2 \times 5^2.$$

Then,

$$\begin{aligned} P(T_{A1} - T_{B2} < 0) = P(Y_1 < 0) &= P\left(\frac{Y_1 - (-5)}{\sqrt{2} \times 5} < \frac{0 - (-5)}{\sqrt{2} \times 5}\right) \\ &= P(Z < 1/\sqrt{2}) = 0.7611. \end{aligned}$$

Let T_{B1} denote Bernt's result if he uses skis 1 and T_{A2} the result for Asbjørn if he uses skis 2.

$$T_{B1} \sim N(62, 4^2), \quad T_{A2} \sim N(63, 7^2).$$

Define $Y_2 = T_{A2} - T_{B1}$. If we proceed the same way as before, we get

$$Y_2 \sim N(63 - 62, 7^2 + 4^2) \equiv N(1, 65)$$

c) Lets denote $(A - S1)$ as the event Asbjørn get skis 1 and $(A - S2)$ if he gets skis 2. $(B - S1)$ and $(B - S2)$ have the same meaning for Bernt.

Using Bayes theorem, we have that

$$\begin{aligned} & P(A - S1 | T_A - T_B < 0) = P(B - S2 | T_A - T_B < 0) \\ &= \frac{P(T_A - T_B < 0 | A - S1) P(A - S1)}{P(T_A - T_B < 0 | A - S1) P(A - S1) + P(T_A - T_B < 0 | A - S2) P(A - S2)} \\ &= \frac{P(T_{A1} - T_{B2} < 0) \times 0.5}{P(T_{A1} - T_{B2} < 0) \times 0.5 + P(T_{A2} - T_{B1} < 0) \times 0.5} \\ &= \frac{P(T_{A1} - T_{B2} < 0)}{P(T_{A1} - T_{B2} < 0) + P(T_{A2} - T_{B1} < 0)} \\ &= \frac{0.3806}{0.6067} = 0.6273 \end{aligned}$$

Oppgave 2

$$T \sim e(t; \beta)$$

a) $\beta = 4.0$

$$P(T > 8) = \int_8^\infty \frac{1}{\beta} e^{-t/\beta} dt = [(-1)e^{-t/\beta}]|_8^\infty = (-1)[0 - e^{-8/\beta}] = e^{-2} = 0.135$$

$$\begin{aligned} P((T_1 \text{ and/or } T_2) > 8) &= 1 - P((T_1 \text{ and } T_2) < 8) && (1) \\ &= 1 - P(T_1 < 8)P(T_2 < 8) = 1 - [1 - e^{-2}]^2 = 0.252. && (2) \end{aligned}$$

In (1) we used the fact that the probability of at least one out of two subsequent calls lasting longer than 8 minutes is equal to the complement of the probability of both call lasting less than 8 minutes.

In (2) we used the fact that T_1 and T_2 are independent.

b) $\beta = 4.0$.

Lets define I_i to be a Bernoulli random variable that takes 1 if the call i last more than 8 minutes.

$$I_i = \begin{cases} 1 & P(T_i > 8) = e^{-2} \\ 0 & P(T_i \leq 8) = 1 - e^{-2} \end{cases} \qquad (3)$$

Then, $N = \sum_{i=1}^{10} I_i$ is the number of calls lasting longer than 8 minutes. Since N is the sum of 10 Bernoulli trials we have that N is binomial distributed.

$$N \sim Bin(10, e^{-2})$$

$$\begin{aligned} P(N \leq 2) &= \sum_{i=0}^{2} \binom{10}{i} (e^{-2})^i (1 - e^{-2})^{10-i} \\ &= (1 - e^{-2})^{10} + 10e^{-2}(1 - e^{-2})^9 + 45e^{-4}(1 - e^{-2})^8 \\ &= 0.234 + 0.366 + 0.257 = 0.857 \end{aligned}$$

c) The likelihood function is given by

$$l(\beta; t_1, ..., t_5) = \prod_{i=1}^{5} \frac{1}{\beta} e^{-t_i/\beta}$$

To find the maximum of the likelihood function, it is easier to maximize the log-likelihood, given by

$$ln\ l(\beta; \cdot) = -5ln\ \beta - \left(\frac{1}{\beta}\right) \sum_{i=1}^{5} t_i$$

We then set the first derivative of the log-likelihood equals to zero:

$$\frac{dlnl}{d\beta} = -5\frac{1}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{5} t_i = 0 \Rightarrow \hat{\beta} = \frac{1}{5} \sum_{i=1}^{5} t_i = \bar{t}$$

The maximum likelihood estimator is:
$\hat{\beta} = \frac{1}{5} \sum_{i=1}^{5} T_i$,
and a maximum likelihood estimate is given by:
$\hat{\beta} = \bar{t} = 4.3$

d) Since $w_{n_1}, w_{n_2}, ..., w_{n_5}$ are independent, the likelihood function is given by

$$l(\beta; w_{n_1}, w_{n_2}, ..., w_{n_5}) = \prod_{i=1}^{5} \frac{1}{\beta^{n_i} n_i!} w_{n_i}^{n_i-1} e^{-w_{n_i}/\beta}$$

To find the maximum of the likelihood function, it is easier to maximize the log-likelihood, given by

$$ln\ l(\beta, \cdot) = \sum_i n_i ln(\beta) + \sum_i (n_i - 1) ln(w_{n_i}) - \frac{1}{\beta} \sum_i w_{n_i}$$

We then set the first derivative of the log-likelihood equals to zero:

$$\frac{dln\ l(\cdot)}{d\beta} = -\frac{1}{\beta}\sum_i n_i + \frac{1}{\beta^2}\sum_i w_{n_i} = 0 \Rightarrow \tilde{\beta} = \frac{1}{\sum_i n_i}\sum_i w_{n_i}$$

The maximum likelihood estimator is:
$\tilde{\beta} = \frac{1}{\sum_i n_i}\sum_i W_{n_i}$
and a maximum likelihood estimate is given by:
$\tilde{\beta} = \frac{1}{\sum_i n_i}\sum_i w_{n_i} = 4.34$

$$E(\hat{\beta}) = \frac{1}{5}\sum_{i=1}^{5} E(T_i) = \frac{1}{5}\sum_i \beta = \beta$$
$$E(\tilde{\beta}) = \frac{1}{\sum_i n_i}\sum_{i=1}^{5} E(W_i) = \frac{1}{\sum_i n_i}\sum_i n_i\beta = \beta$$

Both estimators are unbiased.

$$Var(\hat{\beta}) = \frac{1}{5^2}\sum_{i=1}^{5} Var(T_i) = \frac{1}{5}\beta^2$$
$$Var(\tilde{\beta}) = \frac{1}{(\sum n_i)^2}\sum_i Var(W_i) = \frac{1}{(\sum_i n_i)^2}\sum_i n_i\beta^2 = \frac{\beta^2}{\sum_i n_i}$$

Since $\sum_i n_i > 5$, we have that $Var(\tilde{\beta}) < Var(\hat{\beta})$.

Oppgave 3

Lets denote the weight loss after cycling for t minutes as ΔV.

$$[\Delta V|t] \sim N(\cdot; \beta t, \sigma^2)$$

a) β can be interpreted as average weight loss per minute training.

The likelihood function is given by

$$l(\beta; (\Delta v, t)_i; i = 1, ..., 12) = \prod_{i=1}^{12} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}(\frac{\Delta v_i - \beta t_i}{\sigma})^2 \right\}$$

To find the maximum of the likelihood function, it is easier to maximize the log-likelihood, given by

$$ln\ l(\beta; \cdot) = -nln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_i(\Delta v_i - \beta t_i)^2$$

We then set the first derivative of the log-likelihood equals to zero:

$$\frac{dln\ l(\beta; \cdot)}{d\beta} = -\frac{1}{\sigma^2}\sum_i(\Delta v_i - \beta t_i)(-t_i) = 0 \Rightarrow \hat{\beta} = \frac{\sum_i \Delta v_i t_i}{\sum_i t_i^2}$$

The maximum likelihood estimator is:

$\hat{\beta} = \frac{\sum_i \Delta v_i t_i}{\sum_i t_i^2}$

and a maximum likelihood estimate is given by:

$\hat{\beta} = \frac{12\times 60.9}{12\times 2426.9} = 0.025$

b)

$$H_0 : \beta = 0.027$$
$$H_1 : \beta < 0.027$$

Test statistic is $\hat{\beta} = \frac{\sum_i t_i \Delta V_i}{\sum_i t_i^2} \sim N(\beta, \sigma^2 / \sum_i t_i^2)$ and we reject H_0 if $\hat{\beta} - 0.027 < k$.

$$Prob(reject\ H_0 | H_0\ is\ true) < 0.05$$
$$Prob(\hat{\beta} - 0.027 < k | \beta = 0.027) < 0.05$$
$$Prob\left(\frac{\hat{\beta}}{\sigma/\sqrt{\sum_i t_i^2}} < \frac{k}{\sigma/\sqrt{\sum_i t_i^2}}\middle| \beta = 0.027\right) < 0.05$$
$$Prob(Z < \frac{k}{\sigma/\sqrt{\sum_i t_i^2}}) < 0.05$$
$$k = \frac{-z_{0.05}\sigma}{\sqrt{\sum_i t_i^2}} = \frac{-1.65 \times 0.1}{170.7} = -0.001$$

Since $\hat{\beta} - 0.027 = -0.002 < k = -0.001$, we reject H_0 at the 5% level of confidence.

c) Predictor: $\widehat{\Delta V} = \hat{\beta} \times 50$

Prediction considering the prediction error:

$$\begin{aligned}[\Delta V - \widehat{\Delta V}] &= \beta \times 50 + \epsilon - \hat{\beta} \times 50 \\ &= 50 \times (\beta - \hat{\beta}) + \epsilon \sim N\left(\cdot; 0, \left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]\sigma^2\right)\end{aligned}$$

$$Prob\left(-z_{0.025} < \frac{\Delta V - \hat{\Delta V}}{\left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]^{1/2} \sigma} < z_{0.025}\right) = 0.95$$

$$Prob\left(\hat{\Delta V} - \left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]^{1/2} \sigma \times z_{0.025} < \Delta V < \hat{\Delta V} + \left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]^{1/2} \sigma \times z_{0.025}\right) = 0.95$$

0.95 prediction interval:

$$\left[50\hat{\beta} - \left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]^{1/2} \sigma \times z_{0.025}, 50\hat{\beta} + \left[1 + 50^2 \frac{1}{\sum_i t_i^2}\right]^{1/2} \sigma \times z_{0.025}\right]$$

Using numerical values we get

$$\left[50 \times 0.025 - \left[1 + 50^2 \frac{1}{(2426.9 \times 12)}\right]^{1/2} 0.1 \times 1.96, 50 \times 0.025 + \left[1 + 50^2 \frac{1}{(2426.9 \times 12)}\right]^{1/2} 0.1 \times 1.96\right]$$

$$\Rightarrow [1.045761, 1.454239]$$