Skriftlig Innlevering 4

1

Vi betrakter en skålvekt for veiing av prøver. Prøven med vekt μ plasseres på en av skålene og lodd plasseres på den andre skålen til vektens arm er tilnærmet horisontal. Den kjente vekten av loddene er da tilnærmet prøvens vekt. Vi antar at den målte vekt er normalfordelt med forventningsverdi μ og varians σ^2 og at to forskjellige målinger er uavhengige.

a) Anta i dette punktet at $\mu = 10$ gram og $\sigma^2 = 0.2^2$ gram^2.

Hva er sannsynligheten for at målt vekt skal være mer enn 10.2 gram? Hva er sannsynligheten for at målt vekt skal avvike fra μ med mer enn 0.2 gram? Hva er sannsynligheten for at gjennomsnittet av to uavhengige målinger skal avvike fra μ med mer enn 0.2 gram?

La X være den målte vekten

Ønsker å standardisere X,

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$$Z = \frac{10{,}2-10}{0{,}2} = 1$$

- $P(X>10{,}2) = 1 - P(X \leq 10{,}2)$

$= 1 - P(Z \leq 1)$

$= 1 - 0{,}8413$ tabell

$\underline{\underline{P(X>10{,}2) = 0{,}1587}}$

- $P(X<9{,}8 \cup X>10{,}2) = \underbrace{P(X<9{,}8)}_{Z = \frac{X-\mu}{\sigma} = \frac{9{,}8-10}{0{,}2} = -1} + P(X>10{,}2)$

$= P(Z<-1) + 0{,}1587$

kontinuerlig

$= P(Z \leq -1) + 0{,}1587$

$= 0{,}1587 + 0{,}1587$ tabell

- $\underline{\underline{P(X<9{,}8 \cup X>10{,}2) = 0{,}3174}}$

$$\bar{X} = \frac{1}{2}\sum_{i=1}^{2} X_i$$

- $P(\bar{X}<9{,}8 \cup \bar{X}>10{,}2) = P(\bar{X}<9{,}8) + P(\bar{X}>10{,}2)$

Sentralgrenseteoremet

$$\bar{X} = 9{,}8 \Rightarrow Z = \frac{9{,}8-10}{\sqrt{\frac{0{,}2^2}{2}}} = -\sqrt{2} = -1{,}41$$

$$X = 10{,}2 \Rightarrow Z = \frac{10{,}2-10}{\sqrt{\frac{0{,}2^2}{2}}} = \sqrt{2} = 1{,}41$$

$= P(Z<-1{,}41) + P(Z>1{,}41)$

kontinuerlig

$= P(Z \leq -1{,}41) + 1 - P(Z \leq 1{,}41)$

tabell

$= 0{,}0793 + 1 - 0{,}9207$

$\Rightarrow \underline{\underline{P(\bar{X}<9{,}8 \cup \bar{X}>10{,}2) = 0{,}1586}}$

Det er kommet inn to prøver som begge skal veies, prøve A og prøve B. La μ_A og μ_B betegne sann vekt av prøvene. Det er foreslått å estimere vekten av prøve A og B på to forskjellige måter:

1. Vei først prøve A og la X_1 betegne resultatet. Vei deretter prøve B og la X_2 betegne resultatet. Bruk følgende estimatorer for μ_A og μ_B,
$$\hat{\mu}_A = X_1, \qquad \text{og} \qquad \hat{\mu}_B = X_2. \tag{1.1}$$

2. Vei først summen av prøvene $A + B$ ved å plassere begge prøvene på samme vektskål og la Y_1 betegne resultatet. Vei deretter differansen av prøvene $A - B$ ved å plassere prøvene på hver sin vektskål og Y_2 betegne resultatet. Bruk følgende estimatorer for μ_A og μ_B,
$$\tilde{\mu}_A = (Y_1 + Y_2)/2, \qquad \text{og} \qquad \tilde{\mu}_B = (Y_1 - Y_2)/2. \tag{1.2}$$

b) Finn forventningsverdi og varians til $\hat{\mu}_A$, $\hat{\mu}_B$, $\tilde{\mu}_A$ og $\tilde{\mu}_B$. Hvilken fremgangsmåte vil du foretrekke til å estimere vekten av prøve A og B, og hvorfor?

For $\hat{\mu}_A$ og $\hat{\mu}_B$ er det logisk at:

Metode 1
$$\begin{cases} E[\hat{\mu}_A] = E[X_1] = \mu_A \\ Var[\hat{\mu}_A] = Var[X_1] = \sigma^2 \\ E[\hat{\mu}_B] = E[X_2] = \mu_B \\ Var[\hat{\mu}_B] = Var[X_2] = \sigma^2 \end{cases}$$

I tillegg er variansen for alle målinger lik (kommer fra vekta)

$\Rightarrow Var[Y_1] = Var[Y_2] = \sigma^2$

Metode 2

$$E[\tilde{\mu}_1] = E\left[\frac{Y_1+Y_2}{2}\right] = \frac{1}{2}\left(E[Y_1]+E[Y_2]\right) = \frac{1}{2}\left(\mu_A + \cancel{\mu_B} + \mu_A - \cancel{\mu_B}\right) = \underline{\mu_A}$$

uavhengige målinger, lineær funksjon

$$Var[\tilde{\mu}_1] = Var\left[\frac{Y_1+Y_2}{2}\right] = \frac{1}{4}\left(Var[Y_1]+Var[Y_2]\right) = \frac{1}{4}\left(\sigma^2+\sigma^2\right) = \underline{\frac{\sigma^2}{2}}$$

$$E[\tilde{\mu}] = E\left[\frac{Y_1-Y_2}{2}\right] = \frac{1}{2}\left(E[Y_1]-E[Y_2]\right) = \frac{1}{2}\left(\cancel{\mu_A} + \mu_B - (\cancel{\mu_A} - \mu_B)\right) = \underline{\mu_B}$$

uavhengige målinger, lineær funksjon

$$Var[\tilde{\mu}] = Var\left[\frac{Y_1-Y_2}{2}\right] = \frac{1}{4}\left(Var[Y_1]+Var[Y_2]\right) = \frac{1}{4}\left(\sigma^2+\sigma^2\right) = \underline{\frac{\sigma^2}{2}}$$

$Var[X-Y] = Var[X] + Var[-1 \cdot Y] = Var[X] + (-1)^2 Var[Y] = Var[X] + Var[Y]$

Jeg ville valgt fremgangsmåte 2, begge er forventningsrette, men variansene i metode 2 er lavere.

2

En fabrikk produserer kabel og en gang i blant oppstår det feil på den produserte kabelen. La Z betegne lengden (i kilometer) på kabelen mellom to etterfølgende feil. Vi skal anta at feilene oppstår uavhengig av hverandre, dvs. at påfølgende observasjoner av Z langs kabelen, $Z_1, Z_2, Z_3, \ldots$, er uavhengige stokastiske variable.

Av erfaring vet en at lengden mellom to etterfølgende feil er eksponensialfordelt med parameter λ, dvs. Z har sannsynlighetstetthet

$$f(z;\lambda) = \begin{cases} \lambda e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers} \end{cases}$$

og kumulativ fordelingsfunksjon

$$F(z;\lambda) = \begin{cases} 1 - e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers.} \end{cases}$$

a) Anta i dette punktet at $\lambda = 0.05$.

Hva er sannsynligheten for at lengden mellom to etterfølgende feil er mer enn 10 kilometer?

Dersom man har observert at de første 10 kilometrene er feilfrie, hva er da sannsynligheten for at også de neste 10 kilometrene er feilfrie?

$$P(Z>10) = 1 - P(Z \leq 10)$$
$$= 1 - F(10; 0{,}05)$$
$$= 1 - (1 - e^{-0{,}05 \cdot 10})$$
$$= e^{-0{,}5} \approx 0{,}607$$

$$\underline{\underline{P(Z>10) = 0{,}607}}$$

$$P(Z>20 \mid Z>10) = \frac{P(Z>20 \cap Z>10)}{P(Z>10)} = \frac{P(Z>20)}{P(Z>10)}$$
$$= \frac{e^{-0{,}05 \cdot 20}}{e^{-0{,}05 \cdot 10}} = \frac{e^{-1}}{e^{-0{,}5}} = e^{-0{,}5} \approx 0{,}607$$

$$\underline{\underline{P(Z>20 \mid Z>10) = 0{,}607}}$$

Av erfaring vet en at lengden mellom to etterfølgende feil er eksponensialfordelt med parameter λ, dvs. Z har sannsynlighetstetthet

$$f(z;\lambda) = \begin{cases} \lambda e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers} \end{cases}$$

og kumulativ fordelingsfunksjon

$$F(z;\lambda) = \begin{cases} 1 - e^{-\lambda z} & \text{for } z > 0, \\ 0 & \text{ellers.} \end{cases}$$

a) Anta i dette punktet at $\lambda = 0.05$.

Hva er sannsynligheten for at lengden mellom to etterfølgende feil er mer enn 10 kilometer?

Dersom man har observert at de første 10 kilometrene er feilfrie, hva er da sannsynligheten for at også de neste 10 kilometrene er feilfrie?

$$
\begin{aligned}
P(M=m) &= P(m \leq Z < m+1) \\
&= P(Z < m+1) - P(Z < m) \\
&= F(m+1\,;\lambda) - F(m\,;\lambda) \\
&= \cancel{1} - e^{-\lambda(m+1)} - \left(\cancel{1} - e^{-\lambda m}\right) \\
&= e^{-\lambda m} - e^{-\lambda(m+1)} \\
&= e^{-\lambda m} - e^{-\lambda m} \cdot e^{-\lambda}
\end{aligned}
$$

$$
\underline{\underline{P(M=m) = \left(1 - e^{-\lambda}\right) e^{-\lambda m}}} \quad \square
$$

c) Bestem sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på n observasjoner $M_1, M_2, \ldots, M_n$.

Z_i er uavhengige variabler $\Rightarrow$ M_i er uavhengige variabler

Simultanfordelingen: $f(m_1, m_2, \ldots m_n\,; \lambda) = \prod_{i=1}^{n} f_M(m_i\,; \lambda) = \left(1 - e^{-\lambda}\right)^n e^{-\lambda(m_1 + m_2 + \ldots + m_n)}$

Rimelighetsfunksjonen

$$
L(\lambda) = f(m_1, m_2, \ldots m_n\,; \lambda) = \left(1 - e^{-\lambda}\right)^n e^{-\lambda(m_1 + m_2 + \ldots + m_n)}
$$

$$
\ln L(\lambda) = n \ln\left(1 - e^{-\lambda}\right) + \ln\left(e^{\lambda m_1} \cdot e^{\lambda m_2} \cdot \ldots \cdot e^{\lambda m_n}\right)
$$

$$
\ln L(\lambda) = n \ln\left(1 - e^{-\lambda}\right) - \lambda \sum_{i=1}^{n} m_i
$$

$$
\frac{d}{d\lambda} \ln L(\lambda) = \frac{n \cdot e^{-\lambda}}{1 - e^{-\lambda}} - \sum_{i=1}^{n} m_i = 0
$$

$$
\frac{n \cdot e^{-\lambda}}{1 - e^{-\lambda}} = \sum_{i=1}^{n} m_i
$$

$$
n \cdot e^{-\lambda} = \sum_{i=1}^{n} m_i - e^{-\lambda} \sum_{i=1}^{n}
$$

$$
e^{-\lambda}\left(n + \sum_{i=1}^{n} m_i\right) = \sum_{i=1}^{n} m_i
$$

$$e^{-\lambda} = \frac{\sum_{i=1}^{n} m_i}{n + \sum_{i=1}^{n} m_i} \quad / \ln$$

$$-\lambda \overbrace{\ln e}^{1} = \ln\left(\frac{\sum_{i=1}^{n} m_i}{n + \sum_{i=1}^{n} m_i}\right) = \ln\left(\sum_{i=1}^{n} m_i\right) - \ln\left(n + \sum_{i=1}^{n} m_i\right)$$

$$\underline{\lambda = \ln\left(n + \sum_{i=1}^{n} m_i\right) - \ln\left(\sum_{i=1}^{n} m_i\right)}$$

Sjekker om det er maks

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = n \cdot \frac{-e^{-x}(1-e^{-x}) - (e^{-x} \cdot e^{-x})}{(1-e^{-x})^2} = n \cdot \frac{-e^{-x} + \cancel{(e^{-x})^2} - \cancel{(e^{-x})^2}}{(1-e^{-x})^2}$$

$$= \frac{-ne^{-\lambda}}{(1-e^{-\lambda})^2} < 0 \text{ for alle } \lambda$$

$\Rightarrow$ Fant maksimumspunkt

Dermed er SME:

$$\underline{\hat{\lambda} = \ln\left(n + \sum_{i=1}^{n} M_i\right) - \ln\left(\sum_{i=1}^{n} M_i\right)}$$

3

I en Poissonprosess med parameter λ, la X_1 være tiden til første hendelse, og la X_2 være tiden mellom første og andre hendelse. Da er X_1 og X_2 uavhengige og eksponentialfordelte med forventningsverdi $1/\lambda$ (du behøver ikke vise dette). La Y betegne tiden til andre hendelse, det vil si

$$Y = X_1 + X_2.$$

Vis at Y er gammafordelt med parametre $\alpha = 2$ og $\beta = 1/\lambda$, altså at sannsynlighetstettheten til Y er

$$f(y) = \begin{cases} \lambda^2 y e^{-\lambda y} & \text{for } y > 0, \\ 0 & \text{ellers.} \end{cases}$$

X_1 og X_2 er eksponentialfordelt:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

$$F_Y(y) = P(Y \leq y) = P(X_1 + X_2 \leq y),$$

$$= \iint_{x_1 + x_2 \leq y} f(x_1, x_2)\, dx_1\, dx_2$$

Siden X_1 og X_2 er uavhengige, er $f(x_1, x_2) = f_X(x_1) \cdot f_X(x_2)$

$$= \int_0^y \int_0^{y - x_2} (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2})\, dx_1\, dx_2 \quad , \quad x_1 \cap x_2 \geq 0$$

Øvre grense "dx_1":
$x_1 + x_2 \leq y$
$\Rightarrow x_1 \leq y - x_2$

$$= \lambda^2 \int_0^y e^{-\lambda x_2} \int_0^{y - x_2} e^{-\lambda x_1}\, dx_1\, dx_2$$

$$= \lambda^{\cancel{2}} \int_0^y e^{-\lambda x_2} \left[-\frac{1}{\cancel{\lambda}} e^{-\lambda x_1} \right]_0^{y - x_2} dx_2$$

$$= \lambda \int_0^y e^{-\lambda x_2} \left(1 - e^{-\lambda y + \lambda x_2}\right) dx_2$$

$$= \lambda \int_0^y e^{-\lambda x_2} - e^{-\lambda y}\, dx_2$$

$$= \lambda \left[-\frac{1}{\lambda} e^{-\lambda x_2} - x_2 e^{-\lambda y} \right]_0^y$$

$$= \lambda \left(-\frac{1}{\lambda} e^{-\lambda y} - y e^{-\lambda y} - \left(-\frac{1}{\lambda} - 0\right)\right)$$

$$F_Y(y) = 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}$$

$$f(y) = \frac{d}{dy} F_Y(y) = 0 - \cancel{(-\lambda) e^{-\lambda y}} - \lambda\left(\cancel{e^{-\lambda y}} - y \lambda e^{-\lambda y}\right) = \lambda^2 y e^{-\lambda y} \quad , \quad y > 0 \quad \square$$