TMA4240 Statistikk
Høst 2020

Norges teknisk-naturvitenskapelige universitet
Instituttt for matematiske fag

Skriftlig innlevering 4

Oppgave 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?

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?

Oppgave 2

Et apparat inneholder k like komponenter og fungerer bare dersom alle disse er i orden. Komponentenes levetider $T_1, T_2, \ldots T_k$ er uavhengige og eksponensielt fordelte med parameter β (> 0), dvs. sannsynlighetstettheten er

$$f(t) = \begin{cases} \dfrac{1}{\beta} e^{-t/\beta} & \text{for} \quad t \geq 0, \\ 0 & \text{for} \quad t < 0. \end{cases}$$

a) Finn den kumulative fordelingsfunksjonen for levetiden til en komponent. Hva blir $P(T_1 < 3)$ og $P(2 < T_1 < 4)$ når $\beta = 5$?

b) La X betegne apparatets levetid. Vis at X er eksponensielt fordelt med parameter β/k. Hva blir apparatets forventede levetid når $k = 4$ og $\beta = 5$?

Bedriften har laget flere utgaver av apparatet med forskjellig antall komponenter. Apparatet fungerer bedre med mange komponenter, men har samtidig kortere forventet levetid. La $X_1, X_2, \ldots, X_n$ være levetidene for n apparater med hhv. $k_1, k_2, \ldots, k_n$ komponenter. To estimatorer for β basert på apparatenes levetider er foreslått,

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} X_i k_i \qquad \text{og} \qquad \tilde{\beta} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} k_i^{-1}}.$$

c) Finn forventningsverdi og varians til begge estimatorene.

d) Vis at en av estimatorene i pkt. (b) er sannsynlighetsmaksimeringsestimatoren (SME) og vis at denne har varians som alltid er mindre enn eller lik variansen til den andre estimatoren.

Hint: Sett $r_i = 1/k_i$ og bruk resultatet $\frac{1}{n}\sum r_i^2 - (\frac{1}{n}\sum r_i)^2 \geq 0$.

Oppgave 3

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?

I resten av oppgaven skal vi anta at λ er en ukjent parameter. Ved hjelp av fabrikkens opptegnelser over tidligere feil på kabelen ønsker vi å estimere λ. Men det viser seg dessverre at fabrikken ikke har notert nøyaktig lengde på kabelen mellom hver feil, i stedet er det kun notert antall hele kilometer, M, med kabel mellom hver feil. Dvs, dersom $Z < 1.0$ har fabrikken notert seg $M = 0$, dersom $1.0 \leq Z < 2.0$ har fabrikken notert seg $M = 1$, dersom $2.0 \leq Z < 3.0$ har fabrikken notert seg $M = 2$, osv.

b) Vis at punktsannsynligheten for M blir

$$\mathrm{P}(M = m) = (1 - e^{-\lambda})e^{-\lambda m} \quad \text{for} \quad m = 0, 1, 2, \ldots.$$

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

Oppgave 4

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

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

Fasit

1. **a**) 0.159, 0.318, 0.158 **b**) $\mathrm{E}(\widehat{\mu}_A) = \mu_A$, $\mathrm{Var}(\widehat{\mu}_A) = \sigma^2$, $\mathrm{E}(\widehat{\mu}_B) = \mu_B$, $\mathrm{Var}(\widehat{\mu}_B) = \sigma^2$, $\mathrm{E}(\widetilde{\mu}_A) = \mu_A$, $\mathrm{Var}(\widetilde{\mu}_A) = \sigma^2/2$, $\mathrm{E}(\widetilde{\mu}_B) = \mu_B$, $\mathrm{Var}(\widetilde{\mu}_B) = \sigma^2/2$

2. **a**) 0.4512, 0.2210 **b**) 1.25

3. **a**) $0.607, 0.607$ **c**) $\widehat{\lambda} = \ln(n + \sum_{i=1}^n M_i) - \ln(\sum_{i=1}^n M_i)$