TMA4245 Statistikk
Vår 2016

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Innlevering 3, blokk II

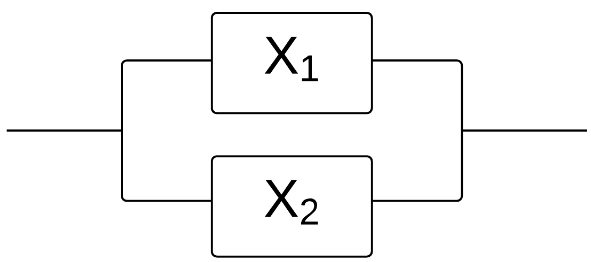
Dette er den første av to innleveringer i blokk 2. Denne øvingen skal oppsummere pensum forelest i 7.-9. forelesningsuke. Øvingen handler om funksjoner av stokastiske variabler, ordningsvariabler, estimering og konfidensintervall.

Oppgave 1

Betrakt et parallellsystem av 2 uavhengige komponenter, se Figur 1. Levetiden til hver komponent er eksponensialfordelt med parameter λ, dvs

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

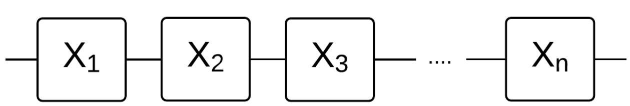
La V være levetiden til systemet. Finn fordelingen til V, samt $\mathrm{E}(V)$.



Figur 1: Parallellsystem med to komponenter

Oppgave 2

Betrakt et seriesystem sammensatt av n komponenter, se Figur 2.



Figur 2: Seriesystem med n komponenter

Levetiden til hver komponent følger en fordeling med kumulativ fordelingsfunksjonen

$$F_X(x) = \begin{cases} 1 - e^{-(\lambda x)^\alpha} & \text{for } x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

hvor $\lambda > 0$ og $\alpha > 0$. Dette kalles en Weibull-fordeling med skalaparameter λ og formparameter α. (Parametriseringen er litt anderledes enn i læreboka.)

Finn den kumulative fordelingsfunksjonen for levetiden til systemet. Hva kalles denne sannsynlighetsfordelingen?

Oppgave 3

Bronse er en legering der kobber og tinn er hovedbestanddelene. Vi studerer kobberinnholdet i bronsebolter av en gitt dimensjon som er laget av en spesiell type bronselegering.

Ved bedriften *Bronsespesialisten* produseres det bronsebolter, og det er tatt stikkprøver av $n = 10$ bronsebolter fra produksjonen. Kobberinnholdet, X_i, $i = 1, ..., n$, er målt. Vi antar at $X_1, X_2, ..., X_n$ er uavhengige og normalfordelte stokastiske variabler med $\mathrm{E}(X_i) = \mu_x$ og $\mathrm{Var}(X_i) = \sigma^2$.

a) Vi antar at forventningen er $\mu_x = 85$ gram og variansen er $\sigma^2 = 1$ gram^2 (kun i dette punktet).

Hva er sannsynligheten for at kobberinnholdet i en tilfeldig valgt bronsebolt er mindre enn 84 gram?

Finn et tall, k, slik at sannsynligheten er 0.01 for at kobberinnholdet i en tilfeldig valgt bronsebolt er større enn k.

Vi ser på kobberinnholdet i to tilfeldig valgte og uavhengige bronsebolter. Hva er sannsynligheten for at kobberinnholdet i de to bronseboltene avviker med mer enn 1.5 gram fra hverandre?

b) Anta at $X_1, \ldots, X_n$ er et tilfeldig utvalg fra en normalfordelt populasjon der forventningsverdien μ_x er kjent, men variansen σ^2 er ukjent. Finn uttrykket for sannsynlighetsmaksimeringsestimatoren for σ^2.

Vi antar i resten av oppgaven at både μ_x og σ^2 er ukjente parametere.

I det neste punktet kan du bruke uten bevis at $\sum_{i=1}^n (X_i - \overline{X})^2/\sigma^2$ er χ^2-fordelt med henholdsvis $n-1$ frihetsgrader (se formelsamlingen side 27).

c) Hvilke egenskaper kjennetegner en god estimator?

To aktuelle estimatorer for σ^2 er $\widehat{\sigma}^2$ og S^2

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \end{aligned}$$

Finn forventningsverdien og variansen til de to estimatorene, ved å bruke relasjonen til kji-kvadrat fordeling. Kommenter resultatet.

Oppgave 4

På Botanisk forskningsstasjon er det plantet et felt med en sjelden grasart. På et bestemt tidspunkt er lengden X målt i cm av et tilfeldig valgt grasstrå eksponentielt fordelt, dvs. at X har sannsynlighetstetthet f gitt ved at $f(x) = \frac{1}{\beta} e^{-x/\beta}$ for $x \geq 0$ og kumulativ fordelingsfunksjon F gitt ved at $F(x) = 1 - e^{-x/\beta}$ for $x \geq 0$.

a) Anta (bare i dette punktet) at $\beta = 10$. Regn ut

$$P(X \leq 4), \quad P(X > 7) \quad \text{og} \quad P(X > 7 \mid X > 4).$$

Vi antar at lengdene av stråene er uavhengige. Planen er å måle lengdene i et tilfeldig utvalg av stråene for å estimere β.

b) Utled sannsynlighetsmaksimeringsestimatoren (SME) for β basert på lengdene i et tilfeldig utvalg på k strå.

Ved en misforståelse slår vaktmesteren graset samme dag som målingene skal gjøres, og pløyer deretter opp feltet. Forskeren planlegger nå å måle lengden Y på et utvalg av stråene som ligger i grasklipperens oppsamler, og basere estimeringen på disse lengdene. Alle stråene ble klipt i samme høyde c, og bare de som var høyere enn klippehøyden ble klipt. Gitt at et strå hadde lengde $X > c$, ligger det altså i oppsamleren, og har lengde $Y = X - c$.

c) Vis at $P(X > c) = e^{-c/\beta}$.

Finn $P(Y > y)$, der $y > 0$.

Hvilken kjent fordeling har Y?

Dessverre viser det seg at stråene i oppsamleren har tørket inn og krympet, slik at det ikke er mulig å måle lengdene. Men det er mulig å telle stråene, og oppsamleren var tom da vaktmesteren startet slåingen. Dessuten kjenner forskeren antall strå som ble plantet ut i feltet og som var i live denne dagen (døde strå ble fjernet fra feltet fortløpende). Forskeren vet at stråene som ble klipt hadde høyde på minst c centimeter, så ved å se på antall strå som ble klipt, sammenlignet med antall strå som var i livet på feltet, vil han kunne lære noe om strålengden til denne grasarten til tross for uhellet med grasklippingen.

d) La Z være antall strå i oppsamleren. Anta at det var n strå i live på feltet. Hvilke betingelser må være oppfylt for at Z skal være binomisk fordelt?

Anta at Z er binomisk fordelt. Uttrykk suksessannsynligheten ved β og c.

Foreslå en estimator for β basert på Z.

Oppgave 5

En metallurg har vært med på å utvikle en ny legering, og skal presentere ulike egenskaper ved legeringen til sine kolleger. På grunn av mindre variasjoner i sammensetningen av legeringen, målefeil og lignende, kan gjentatte målinger av smeltepunktet til legeringen antas å være realisasjoner av uavhengige og normalfordelte variabler med forventningsverdi μ og kjent standardavvik $\sigma = 2°$ C.

a) Anta først (kun i dette punktet) at forventningsverdien er $\mu = 1468$°C.

Metallurgen tar en måling av smeltepunktet. Hva er sannsynligheten for at observert smeltepunkt er lavere enn 1467°C?

Anta så at metallurgen tar åtte uavhengige målinger av smeltepunktet, $X_1, X_2, ..., X_8$. Hva er sannsynligheten for at gjennomsnittet av de åtte målingene av smeltepunktet, $\bar{X} = \frac{1}{8}\sum_{i=1}^{8} X_i$, ligger mellom 1467°C og 1469°C?

Anta i resten av oppgaven at forventningsverdien til legeringens smeltepunkt, μ, er ukjent, men at standardavviket er kjent og lik $\sigma = 2°$ C.

b) Utled et 90% konfidensintervall for μ basert på n uavhengige målinger av smeltepunktet, $X_1, X_2, ..., X_n$.

Hvor stor må n minst være for at lengden på intervallet ikke overstiger 3°C? Kall dette antallet n_0 og bruk de n_0 første observasjonene gitt i tabell 1 til å bestemme intervallet numerisk.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1467.4	1468.0	1471.6	1468.6	1468.8	1471.9	1469.4	1466.0

Tabell 1: *Data over smeltepunkt i °C. Det oppgis at gjennomsnittet av de* 8 *observasjonene er* 1469.0°*C.*

Oppgave 6

Bremselengde for bil med to ulike dekktyper skal undersøkes. En bremseprøve utføres ved at man begynner å bremse når bilen kjører i $80km/t$ og bremselengde måles. For dekktype 1 utføres n slike prøver. La $X_1, X_2, \ldots, X_n$ betegne bremselengdene målt i disse prøvene. Helt tilsvarende utføres m bremseprøver for dekktype 2. La $Y_1, Y_2, \ldots, Y_m$ betegne bremselengdene målt her.

Anta at $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ alle er uavhengige og normalfordelt. Anta videre at X_1, X_2, $\ldots$, X_n har ukjent forventningsverdi μ_1, at Y_1, $Y_2, \ldots, Y_m$ har ukjent forventningsverdi μ_2 og at X_1, $X_2 \ldots, X_n$, Y_1, $Y_2, \ldots, Y_m$ alle har samme kjente varians σ_0^2.

Utled et $(1-\alpha)\cdot 100\%$ konfidensintervall for differansen $\mu_1 - \mu_2$.

Regn også ut intervallet numerisk når $\alpha = 0.05$, $n = m = 10$, $\sigma_0^2 = 2^2$ og observerte bremselengder er som gitt i følgende tabell

i	1	2	3	4	5	6	7	8	9	10
x_i	33.0	30.8	28.0	28.7	28.9	26.6	27.9	28.9	27.8	27.4
y_i	23.4	25.3	25.0	28.9	26.7	25.9	24.4	26.8	28.8	25.5

Det oppgis at $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 28.80$ og $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i = 26.07$.

Fasit

3. **a**) 0.1587, 87.33, 0.2892

4. **a**) 0.33, 0.50, 0.74

5. **a**) 0.3085, 0.8414 **b**) $[1467.41, 1470.35]$

6. $[0.977, 4.483]$