Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

Kvalitetsavdelingen i en fabrikk som produserer klokker ønsker å se nøyere på de defekte klokkene som av og til kommer fra produksjonen. De bestemmer seg for å bruke $k = 3$ defekte klokker i inspeksjonen. Fra en produksjonslinje kommer det en kontinuerlig strøm av klokker og hver klokke som produseres har en sannsynlighet p for å være defekt, uavhengig av hverandre.

La X benevne det minste antall klokker en må inspisere fra produksjonsstrømmen fra linjen for å identifisere eksakt $k = 3$ defekte klokker. Vi vet da at den tilfeldige variablen X er negativ-binomisk fordelt med sannsynlighetsfordeling

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \qquad ; \qquad x = k, k+1, \ldots$$

a) Anta i dette punktet at defektsannsynligheten $p = 0.1$ og regn ut sannsynlighetene

$$P(X > 3),$$
$$P(X < 6),$$
$$P(X \geq 6 | X > 3).$$

Defektsannsynligheten p antas nå ukjent og skal estimeres. Kvalitetsavdelingen gjentar forsøket med å identifisere $k = 3$ defekte klokker n ganger, og får et tilfeldig utvalg: $X_1, \ldots, X_n$. Basert på dette tilfeldige utvalget ønsker en å estimere p.

b) Utled sannsynlighetsmaksimeringsestimatoren $\hat{p}$ for p, basert på det tilfeldige utvalget.

Anta så at klokkefabrikken faktisk har to separate produksjonslinjer, benevnt henholdsvis A og B og med ulike defektsannsynligheter p_A og p_B

Statistikeren i avdelingen mottar en observasjon $X = x$ på antallet klokker som må inspiseres før $k = 3$ defekte er identifisert fra en av de to linjene. Han vet ikke om observajonen er hentet fra produksjonslinje A eller B, så han antar derfor i utgangspunktet sannsynlighet 0.5 for hver av mulighetene.

c) Bruk Bayes' regel til å utlede et uttrykk for sannsynligheten for at observasjonen er hentet fra produksjonslinje A gitt at $X = x$.

La så $p_A = 0.1$ og $p_B = 0.2$, samt $x = 5$, og regn ut tallsvaret for sannsynligheten for at observasjonen er hentet fra produksjonslinje A.

Oppgave 2

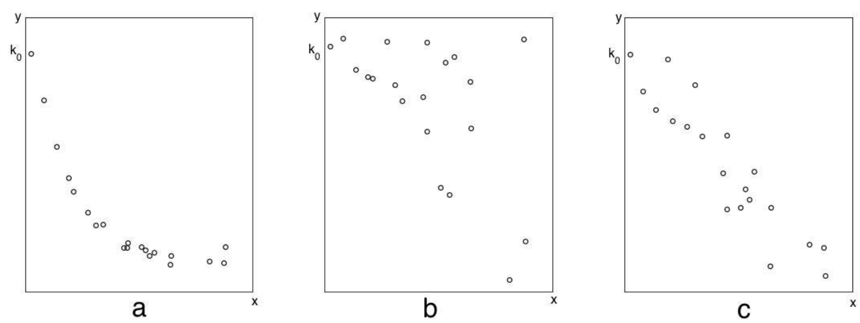
En bilprodusent vil evaluere slitasjen på bremseklossene på bilene som produseres. En definerer en enkel lineær regresjonsmodell,

$$Y = k_0 - \beta x + \epsilon,$$

hvor responsvariablen Y er tykkelsen på bremseklossene, forklaringsvariablen x er antall kilometer kjørt, k_0 er klosstykkelsen for en ny bil og β er slitasjeraten. Feilleddet ϵ antas å være normalfordelt, $n(\epsilon; 0, \sigma)$. Vi antar at k_0 er kjent, mens raten β og variansen σ^2 er ukjente modellparametre som skal estimeres.

Bilprodusenten designer et forsøk for å estimere β og σ^2. En gruppe av n testsjåfører kjører ulike biler over varierende antall kilometer og deretter måles klosstykkelsen. Dette definerer et tilfeldig utvalg fra modellen, $(x_1, Y_1), \ldots, (x_n, Y_n)$.

I Figur 1 presenteres tre plott av mulige utfall $(x_1, y_1), \ldots, (x_n, y_n)$ for $n = 20$.



Figur 1: Tre mulige utfall av forsøket i Oppgave 2.

a) For hvilket av disse tre plottene i Figur 1 synes den enkle lineære regresjonsmodellen definert over å være en god modell? Begrunn svaret.

Hvorfor er ikke modellen god for de to andre plottene?

Den enkle lineære regresjonsmodellen definert over må nødvendigvis være approksimativ og gyldig bare for et intervall av forklaringsvariablen x. Forklar kort hvorfor.

b) Bruk enten minste kvadraters metode eller sannsynlighetsmaksimeringsprinsippet til å utlede en estimator $\hat{\beta}$ for β basert på det tilfeldige utvalget. Vis at estimatoren blir

$$\hat{\beta} = \frac{k_0 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i Y_i}{\sum_{j=1}^{n} x_j^2}.$$

Utled uttrykk for forventningsverdien og variansen til $\hat{\beta}$.

Som estimator for σ^2 basert på det tilfeldige utvalget er det rimelig å benytte

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[Y_i - (k_0 - \hat{\beta} x_i) \right]^2.$$

Videre oppgis det at $\hat{\beta}$ er normalfordelt, at

$$V = \frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

er kji-kvadratfordelt med $(n-1)$ frihetsgrader, og at $\hat{\beta}$ og V er uavhengige tilfeldige variabler.

c) Utled et $100(1-\alpha)\%$-konfidensintervall for β.

Beskriv kort hvordan konfidensintervallet kan benyttes til å teste om slitasjeraten er eksakt lik β_0.

Oppgave 3

En bonde fra Sogn dyrker epler. Han pakker og selger eplene i det som betegnes '3-kilo-poser'. Det er selvsagt et helt antall epler i hver pose, så posene varierer nødvendigvis i vekt. En tilfeldig pose veier X kilogram, hvor X er normalfordelt med forventning μ og varians σ^2. Betrakt μ som ukjent og la $\sigma^2 = 0.4^2$. Det er selvsagt ønskelig at forventningen μ er 3 kilogram.

Lageret til Rema 1000 på Sandmoen mottar et stort billass med '3-kilo-poser' med epler fra bonden. Innkjøpsavdelingen på Rema 1000 ønsker å kontrollere at posene er tunge nok. De tar et tilfeldig utvalg på $n = 3$ poser fra billasset, veier disse posene, og registrerer følgende vekter: X_1, X_2, X_3. En rimelig estimator for forventet vekt μ er

$$\hat{\mu} = \frac{1}{3}\sum_{i=1}^{3} X_i = \bar{X}.$$

a) Estimatoren $\hat{\mu}$ er normalfordelt, forklar kort med ord hvorfor.

Utled uttrykk for forventningen og variansen til denne normalfordelingen.

Er estimatoren $\hat{\mu}$ forventningsrett? Begrunn svaret.

Beskriv kort med ord hva det innebærer at en estimator er forventningsrett.

Innkjøpsavdelingen ønsker å sikre seg at forventet vekt av posene, μ, er minst 3 kilogram. Statistikeren i avdelingen formulerer vektkontrollen som et hypotesetestingsproblem,

$$H_0 : \mu = 3 \text{ mot } H_1 : \mu < 3$$

og bruker signifikansnivå $\alpha = 0.05$ i en test med estimatoren $\hat{\mu}$ som testobservator.

b) Utled forkastningsområdet for $\hat{\mu}$ med hensyn til hypotesene definert over.

c) Utled et uttrykk for styrkefunksjonen for testen som ble definert i punkt **b)**.

Skisser grafisk hvordan styrkefunksjonen ser ut.

Dersom forventet vekt μ er på kun 2.9 kilogram, så ønsker statistikeren å avsløre at posene veier for lite med sannsynlighet minst 0.9. Regn ut hvor mange poser n det da må være i det tilfeldige utvalget som hentes fra billasset.

Statistikeren fortsetter å leke seg litt med problemet etter arbeidstid. Han betrakter den ordnede versjonen av det tilfeldige utvalget, $X_{(1)}, X_{(2)}, X_{(3)}$ i stigende orden. Deretter definerer han en alternativ estimator for μ,

$$\tilde{\mu} = X_{(2)}$$

d) Utled et uttrykk for sannsynlighetsfordelingen til estimatoren $\tilde{\mu}$.

Vis at $\tilde{\mu}$ er en forventningsrett estimator for μ.

Fasit

1. **a**) 0.999, 0.00856, 0.9924 **b**) $\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i}$ **c**) 0.1366

2. **b**) $\mathrm{E}\left[\widehat{p}\right] = \beta$, $\mathrm{Var}\left[\widehat{\beta}\right] = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

3. **a**) $\mathrm{E}\left[\widehat{\mu}\right] = \mu$, $\mathrm{Var}\left[\widehat{\mu}\right] = \frac{\sigma^2}{n}$, $\widehat{\mu}$ er forventningsrett **b**) Forkast H_0 dersom $\widehat{\mu} < 2.62$ **c**) 138 **d**) $f_{X_{(2)}}(x) = 6F_X(x)f_X(x)\left[1 - F_X(x)\right]$