TMA4240 Statistikk
Eksamen desember 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

La den kontinuerlege stokastiske variabelen X ha fordelingsfunksjon (sannsynstettleik) gjeve ved

$$f(x) = \begin{cases} \theta \cdot x^{-(\theta+1)} & \text{for } x > 1 \\ 0 & \text{ellers,} \end{cases} \tag{1.1}$$

der $\theta > 0$. Denne fordelinga er ein populær model dersom X er eit normalisert mål på velstand i ein populasjon.

a) Finn den kumulative fordelingsfunksjonen til X, $F(x) = P(X \leq x)$.

La $\theta = 1.16$ og rekn ut $P(X \leq 2)$, $P(X > 4)$, og $P(X > 4 \mid X > 2)$.

La $X_1, X_2, \ldots, X_n$ vere eit tilfeldig utval (uavhengige og identisk fordelte stokastiske variablar) frå ein populasjon som føljer $f(x)$ i likning (1.1).

b) Utlei sannsynsmaksimeringsestimatoren (the maximum likelihood estimator) for θ.

Oppgave 2

a) La X og Y vere to uavhengige normalfordelte stokastiske variablar, der $\mathrm{E}(X) = 1$, $\mathrm{E}(Y) = 2$, $\mathrm{Var}(X) = 4$ og $\mathrm{Var}(Y) = 1$.

Finn $P(X \leq 0)$.

Kva fordeling har $X + Y$?

Finn $P(X + Y > 4)$, og $P(X - Y \leq -2)$.

La $X_1, X_2, \ldots, X_n$ vere eit tilfeldig utval frå ein normalfordelt populasjon med forventningsverdi μ_X og standardavvik σ. Vidare er $Y_1, Y_2, \ldots, Y_m$ eit tilfeldig utval frå ein normalfordelt populasjon med forventningsverdi μ_Y og standardavvik σ. Anta at dei to utvala er uavhengige av kvarandre.

Definer $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j$, $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ og $S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y})^2$.

b) Vi ønskjer å estimere σ^2, og ser på følgjande estimatorar:

$$S^2_{\text{pooled}} = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$
$$S^2_{\text{mean}} = \frac{1}{2}S_X^2 + \frac{1}{2}S_Y^2$$

Er estimatorane forventingsrette?

Finn variansen til dei to estimatorane.

La $n = 10$ og $m = 20$. Kven av de to estimatorane vil du då foretrekke?

Du kan bruke at $V_X = \frac{(n-1)S_X^2}{\sigma^2}$ er kjikvadratfordelt (khikvadratfordelt) med parameter $n-1$, og $V_Y = \frac{(m-1)S_Y^2}{\sigma^2}$ er kjikvadratfordelt (khikvadratfordelt) med parameter $m-1$.

c) Vi ønskjer å undersøkje om det er grunn til å tru at μ_X er større enn μ_Y.

Sett opp nullhypotese og alternativ hypotese.

Velg testobservator, og oppgje kva sannsynsfordeling denne har når nullhypotesa er sann.

Finn ein forkastningsregel når vi veljer signifikansnivå 0.05.

Kva blir resultatet av testen dersom vi har observert: $n = 129$, $\bar{x} = 75.2$, $s_X^2 = 174.6$, $m = 141$, $\bar{y} = 61.0$ og $s_Y^2 = 292.1$?

Tala er tatt frå ei nettbasert spørreundersøking i TMA4240 hausten 2015, der X og Y-verdien gjev svaret på spørsmålet: «Hvis du går på forelesning, eller ser forelesningen på video, hvor mange prosent av forelesningen mener du at du i gjennomsnitt forstår». Vidare er X-utvalet dei som svarte «Enig» eller «Svært enig» på påstanden «Akkurat nå synes jeg at Statistikk er et artig kurs» og Y-utvalet dei som svarte «Nøytral», «Uenig» eller «Svært uenig».

Korleis vil du oversette ditt svar på hypotesetesten (forkast eller ikkje forkast nullhypotesa) i denne situasjonen?

Oppgave 3

To robotar kommuniserer ved å sende binære signal til kvarandre. Dessverre er ikkje overføringa av signala perfekt: Det er eit sannsyn for at dersom 0 er sendt frå den eine roboten, så mottar den andre roboten 1, og dersom 1 er sendt frå den eine roboten, mottar den andre roboten 0.

a) La X vere ein diskret stokastisk variabel som seier kva signal som er sendt, og la Y vere ein diskret stokastisk variabel som seier kva signal som er mottatt. Moglege verdiar for X og Y er 0 og 1. Anta at $P(X = 1) = 0.2$. Anta vidare at kommunikasjonen er slik at $P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 0.9$.

Bruk lova om total sannsynlighet til å rekne ut $P(Y = 1)$.

Bruk Bayes' regel til å rekne ut sannsynet $P(X = 1|Y = 1)$.

Vi ser vidare på signal av lengde 5. La X_i vere ein diskret stokastisk variabel som gjev signalet som er sendt i posisjon i, og Y_i ein diskret stokastiske variabel som gjev signalet som er mottatt i posisjon i. Her er $i = 1, 2, \ldots, 5$, og dei moglege verdiane til X_i og Y_i er 0 og 1. Anta at signalet i kvar posisjon i blir overført korrekt med sannsyn $P(Y_i = 1 \mid X_i = 1) = P(Y_i = 0 \mid X_i = 0) = 0.9$ for alle $i = 1, 2, \ldots, 5$. Den stokastiske variabelen Y_i er kun avhengig av X_i og ikkje av andre Y_j eller X_j, for $j \neq i$.

b) Kva er sannsynet for at eit utsendt signal blir korrekt mottatt i minst 4 av dei 5 posisjonane?

I ein situasjon har vi at signala $(0, 0, 0, 0, 0)$ og $(1, 1, 1, 1, 1)$ begge har sannsyn 0.35 for å bli utsendt, medan kvart at dei resterande 30 moglege signala har sannsyn 0.01. Dersom mottatt signal er $(0, 0, 0, 0, 0)$. Kva er sannsynet for at det utsendte signalet er $(0, 0, 0, 0, 0)$?

Oppgave 4

På søndagar sist vinter selde Alexander kakao til skiløparar som passerte huset hans. Det skal han og gjere denne vinteren.

a) La Y vere talet på koppar kakao Alexander seljer ein gjeve dag. Anta at Y er Poisson-fordelt med forventningsverdi $\lambda = 18$.

Kva er sannsynet for at han sel 18 eller færre koppar kakao?

Kva er sannsynet for at han sel meir enn 10 koppar, gjeve at han sel 18 eller færre kopper?

Det er kjent at vi kan tilnærme ei Poisson-fordeling med ei normalfordeling når λ er stor, og vidare i denne oppgåva antar vi at talet på koppar med kakao som Alexander sel ein gjeve dag er normalfordelt.

Alexander observerte sist vinter at kakaosalet var avhengig av føreforholda i skiløypene. Han laga seg ein føreforholdindeks, x, der $x = 1$ var «dårlige forhold», $x = 2$ var «gode forhold», $x = 3$ var «veldig gode forhold» og $x = 4$ var «fantastiske forhold».

For kvar av 20 søndagar, $i = 1, \ldots, 20$, noterte han ned talet på koppar kakao han selte, y_i og tilhøyrande føreforholdindeks, x_i.

Vi set opp følgjande regresjonsmodell for salet:

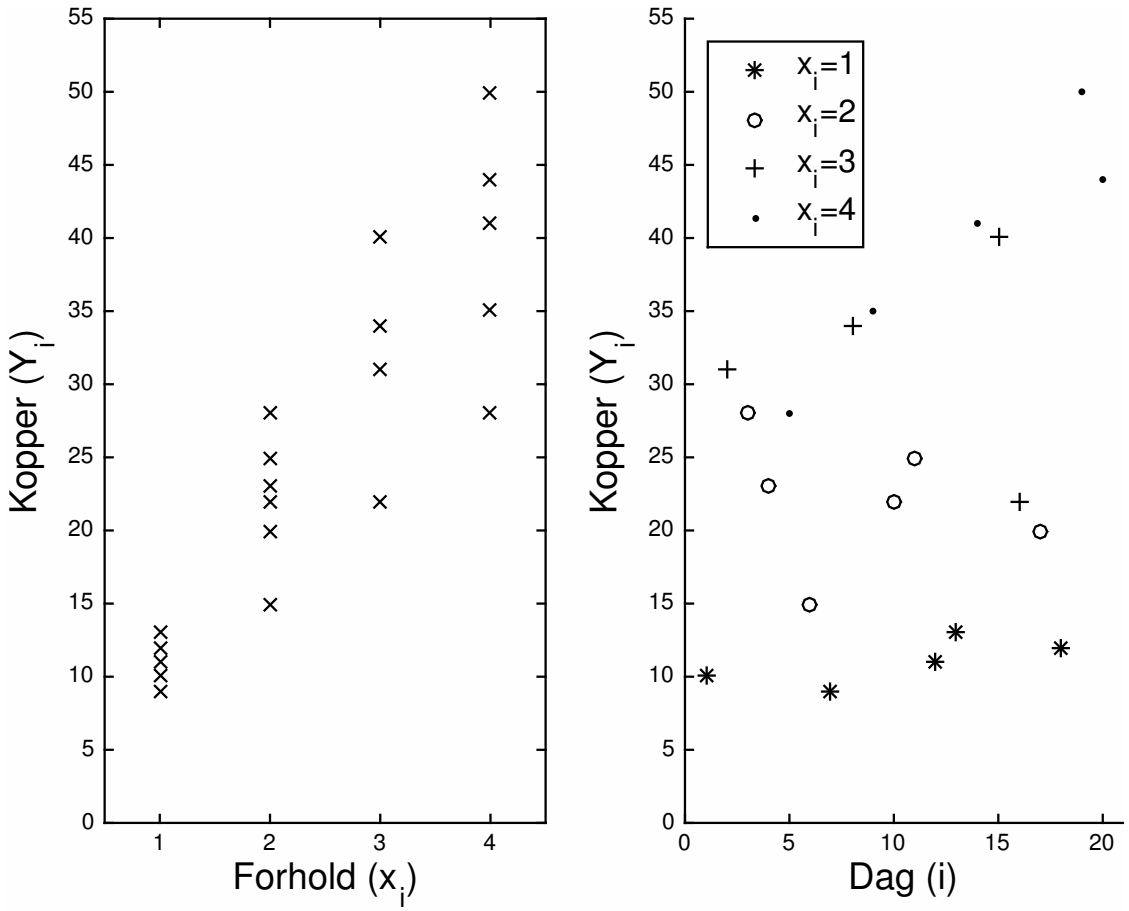
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \ldots, 20,$$

der $\epsilon_1, \ldots, \epsilon_{20}$ er uavhengige og identisk normalfordelte variablar med forventningsverdi 0 og varians σ^2, og parametra β_0 og β_1 er ukjende regresjonsparametre.

Frå data har vi $\bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = 2.45$, $\bar{y} = \frac{1}{20}\sum_{i=1}^{20} y_i = 25.65$, $\sum_{i=1}^{20}(x_i - \bar{x})^2 = 24.95$ og $\sum_{i=1}^{20}(x_i - \bar{x})y_i = 237.15$. Vidare er $s^2 = \frac{1}{18}\sum_{i=1}^{20}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 5.65^2$ eit estimat av variansen σ^2, der $\hat{\beta}_0$ og $\hat{\beta}_1$ er minste kvadratsums estimat for høvesvis β_0 og β_1.

b) Datasettet er presentert i figur 1. Vurder modellantakingane i regresjonsmodellen ved å studere figuren.

Alexander vil predikere kor mange kopper han sel ein dag då føreforholda er «fantastiske». Rekn ut eit 90 prosent prediksjonsintervall for talet på koppar kakao selt denne



Figur 1: Venstre: Talet på koppar kakao som er selt (y-akse) for ulike føreforhold (x-akse). Høgre: Talet på koppar kakao som er selt (y-akse) for alle dagar (x-akse). Føreforholdindeksen er gjeve ulike plottesymbol for ulike verdier.

dagen.

I resten av oppgåva kan du bruke føljande resultat: La Y vere ein normalfordelt stokastisk variabel med forventningsverdi μ og varians σ^2. Dermed er sannsynstettleiken til Y gjeve ved $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y-\mu)^2}{2\sigma^2})$. Då gjeld $\int_{-\infty}^{a} y f(y) dy = \mu\Phi((a-\mu)/\sigma) - \sigma\phi((a-\mu)/\sigma)$, der $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})$ er sannsynstettleiken til ein standard normalfordelt stokastisk variabel og $\Phi(z)$ er den kumulative fordelingsfunksjonen til ein standard normalfordelt stokastisk variabel.

c) Alexander har nokre søndagar gått tom for kakao. Anta at det predikerte talet på selde koppar kakao ein gjeve dag er normalfordelt med forventningsverdi 20 og varians 5^2, men at maksimum talet på koppar Alexander kan selje er dei $n = 25$ han har laga på førehand. Rekn ut det forventa talet på koppar som blir selt denne dagen.

Alexander kjøpar inn og lagar kakao til 5 kroner per kopp. Han seljer ein kopp for 20 kroner. Hvis han ikkje får seld alle dei n koppane med kakao som han har forbereid, tømmer han ut dei han ikkje har selt (uten inntekt). Han vil maksimere forventa forteneste, dvs. inntekter minus utgifter. Finn det talet n på kopper det vil være optimalt å forbereide. Hint: prøv deg fram til optimal løysing for n.

Fasit

1. **a**) $F(x) = 1 - \frac{1}{x^\theta}$ når $x > 1$, 0.55, 0.20, 0.44

2. **a**) 0.3085, 0.3264, 0.3264 **b**) Begge er forventningsrette, $\text{Var}\left(S^2_{pooled}\right) = \frac{2\sigma^4}{n+m-2}$, $\text{Var}\left(S^2_{mean}\right) = \frac{\sigma^4}{2}\left(\frac{1}{n-1} + \frac{1}{m-1}\right)$, foretrekker S^2_{pooled} **c**) $H_0 : \mu_X = \mu_Y$, $H_1 : \mu_X > \mu_Y$, testobservator: $T_0 = (\overline{X} - \overline{Y})/(S_{pooled}(1/n + 1/m)^{1/2})$, forkaster H_0 hvis $T_0 > t_{0.05,n+m-2}$, forkaster H_0

3. **a**) 0.26, 0.69 **b**) 0.92, 0.98

4. **a**) 0.56, 0.946 **b**) $(29.9, 50.9)$ **c**) 19.58, 23