TMA4240 Statistikk
Eksamen desember 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

La den kontinuerlige stokastiske variabelen X ha fordelingsfunksjon (sannsynlighetstetthet) gitt 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 fordelingen er en populær model dersom X er et normalisert mål på velstand i en populasjon.

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

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

La $X_1, X_2, \ldots, X_n$ være et tilfeldig utvalg (uavhengige og identisk fordelte stokastiske variabler) fra en populasjon beskrevet av $f(x)$ i ligning (1.1).

b) Utled sannsynlighetsmaksimeringsestimatoren (the maximum likelihood estimator) for θ.

Oppgave 2

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

Finn $P(X \leq 0)$.

Hvilken fordeling har $X + Y$?

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

La $X_1, X_2, \ldots, X_n$ være et tilfeldig utvalg fra en normalfordelt populasjon med forventningsverdi μ_X og standardavvik σ. Videre er $Y_1, Y_2, \ldots, Y_m$ et tilfeldig utvalg fra en normalfordelt populasjon med forventningsverdi μ_Y og standardavvik σ. Anta at de to utvalgene er uavhengige av hverandre.

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 ønsker å estimere σ^2, og ser på følgende estimatorer:

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

Er estimatorene forventingsrette?

Finn variansen til de to estimatorene.

La $n = 10$ og $m = 20$. Hvilken av de to estimatorene vil du da 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 ønsker å undersøke om det er grunn til å tro at μ_X er større enn μ_Y.

Sett opp nullhypotese og alternativ hypotese.

Velg testobservator, og angi hvilken sannsynlighetsfordeling denne har når nullhypotesen er sann.

Finn en forkastningsregel når vi velger signifikansnivå 0.05.

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

Tallene er tatt fra en nettbasert spørreundersøkelse i TMA4240 høsten 2015, der X og Y-verdien gir 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». Videre er X-utvalget de som svarte «Enig» eller «Svært enig» på påstanden «Akkurat nå synes jeg at Statistikk er et artig kurs» og Y-utvalget de som svarte «Nøytral», «Uenig» eller «Svært uenig».

Hvordan vil du oversette ditt svar på hypotesetesten (forkast eller ikke forkast nullhypotesen) til denne situasjonen?

Oppgave 3

To roboter kommuniserer ved å sende binære signaler til hverandre. Dessverre er ikke overføringen av signaler perfekt: Det er en sannsynlighet for at hvis 0 er sendt fra den ene roboten så mottas 1 av den andre roboten, og hvis 1 er sendt fra den ene roboten mottas 0 av den andre roboten.

a) La X være en diskret stokastisk variabel som angir signalet som er sendt, og la Y være en diskret stokastisk variabel som angir signalet som er mottatt. Mulige verdier for X og Y er 0 og 1. Anta at $P(X = 1) = 0.2$. Anta videre at kommunikasjonen er slik at $P(Y = 1 \mid X = 1) = P(Y = 0 \mid X = 0) = 0.9$.

Bruk loven om total sannsynlighet til å regne ut $P(Y = 1)$.

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

Vi ser videre på signaler av lengde 5. La X_i være en diskret stokastisk variabel som angir signalet som er sendt i posisjon i, og Y_i en diskret stokastiske variabel som angir signalet som er mottatt i posisjon i. Her er $i = 1, 2, \ldots, 5$, og de mulige verdiene til X_i og Y_i er 0 og 1. Anta at signalet i hver posisjon i overføres korrekt med sannsynlighet $P(Y_i = 1 \mid X_i = 1) = P(Y_i = 0 \mid X_i = 0) = 0.9$ for alle $i = 1, 2, \ldots, 5$. Den stokastiske variablen Y_i er kun avhengig av X_i og ikke av andre Y_j eller X_j, for $j \neq i$.

b) Hva er sannsynligheten for at et utsendt signal blir korrekt mottatt i minst 4 av de 5 posisjonene?

I en situasjon har vi at signalene $(0, 0, 0, 0, 0)$ og $(1, 1, 1, 1, 1)$ begge har sannsynlighet 0.35 for å bli utsendt, mens hver at de resterende 30 mulige signalene har sannsynlighet 0.01. Dersom mottatt signal er $(0, 0, 0, 0, 0)$. Hva er sannsynligheten for at utsendt signal er $(0, 0, 0, 0, 0)$?

Oppgave 4

På søndager sist vinter solgte Alexander kakao til skiløpere som passerte huset hans. Det skal han gjøre også denne vinteren.

a) La Y være antall kopper kakao Alexander selger en gitt dag. Anta at Y er Poisson-fordelt med forventningsverdi $\lambda = 18$.

Hva er sannsynligheten for at han selger 18 eller færre kopper kakao?

Hva er sannsynligheten for at han selger mer enn 10 kopper, gitt at han selger 18 eller færre kopper?

Det er kjent at vi kan tilnærme en Poisson-fordeling med en normalfordeling når λ er stor, og videre i denne oppgaven antar vi at antall kopper med kakao som Alexander selger en gitt dag er normalfordelt.

Alexander observerte sist vinter at kakaosalget var avhengig av føreforholdene i skiløypene. Han laget seg en 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 hver av 20 søndager, $i = 1, \ldots, 20$, noterte han ned antall kopper kakao han solgte, y_i og tilhørende føreforholdindeks, x_i.

Vi setter opp følgende regresjonsmodell for salget:

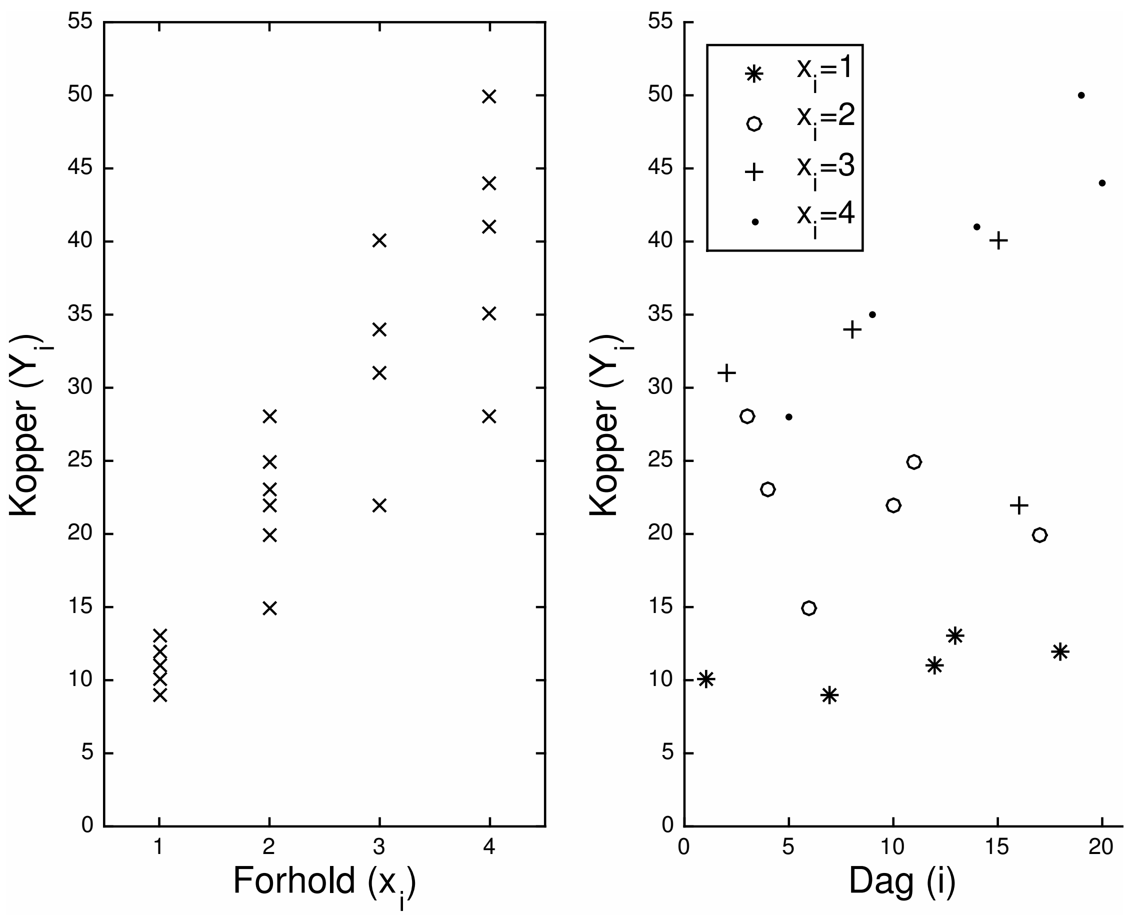
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \ldots, 20,$$

der $\epsilon_1, \ldots, \epsilon_{20}$ er uavhengige og identisk normalfordelte variabler med forventningsverdi 0 og varians σ^2, og parametrene β_0 og β_1 er ukjente regresjonsparametere.

Fra 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 $\sum_{i=1}^{20} (x_i - \bar{x}) y_i = 237.15$. Videre er $s^2 = \frac{1}{18} \sum_{i=1}^{20} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 5.65^2$ et estimat av variansen σ^2, der $\hat{\beta}_0$ og $\hat{\beta}_1$ er minste kvadratsums estimat for henholdsvis β_0 og β_1.

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

Alexander vil predikere antall kopper han selger en dag da føreforholdene er «fantastiske». Regn ut et 90 prosent prediksjonsintervall for antallet solgte kopper kakao denne



Figur 1: Venstre: Antall kopper kakao som er solgt (y-akse) for ulike føreforhold (x-akse). Høyre: Antall kopper kakao som er solgt (y-akse) for alle dager (x-akse). Føreforholdindeksen er gitt ulike plottesymbol for ulike verdier.

dagen.

I resten av oppgaven kan du bruke følgende resultater: La Y være en normalfordelt stokastisk variabel med forventningsverdi μ og varians σ^2. Dermed er sannsynlighetstettheten til Y gitt ved $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y-\mu)^2}{2\sigma^2})$. Da gjelder $\int_{-\infty}^{a} yf(y)dy = \mu\Phi((a-\mu)/\sigma) - \sigma\phi((a-\mu)/\sigma)$, der $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})$ er sannsynlighetstettheten til en standard normalfordelt stokastisk variabel og $\Phi(z)$ er den kumulative fordelingsfunksjonen til en standard normalfordelt stokastisk variabel.

c) Alexander har noen søndager gått tom for kakao. Anta at predikert antall solgte kopper med kakao en gitt dag er normalfordelt med forventningsverdi 20 og varians 5^2, men at maksimum antall kopper Alexander kan selge er de $n = 25$ han har laget på forhånd. Regn ut forventet antall kopper som blir solgt denne dagen.

Alexander kjøper inn og lager kakao til 5 kroner per kopp. Han selger en kopp for 20 kroner. Hvis han ikke får solgt alle de n koppene med kakao som han har forberedt, tømmer han ut de han ikke har solgt (uten inntekt). Han vil maksimere forventet fortjeneste, dvs. inntekter minus utgifter. Finn antall kopper n det vil være optimalt å forberede. Hint: prøv deg fram til optimal løsning 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, Var $\left(S^2_{pooled}\right) = \frac{2\sigma^4}{n+m-2}$, 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