

Oppgave 1 Pinnekjøtt

- a) La X være Poissonfordelt med punktsannsynlighet $P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$, $x = 0, 1, \dots$, og anta parameter $\mu = 3$.

Regn ut sannsynligheten for at X er mindre eller lik 3.

Regn ut $P(X \leq 3 | X > 0)$.

- b) En null-trunkert Poissonfordeling med parameter μ er karakterisert med punktsannsynlighet $f(y) = P(X = y | X > 0)$, der X er Poissonfordelt med parameter μ .

Vis at punktsannsynligheten til Y er

$$f(y) = \frac{\mu^y}{y!} \frac{e^{-\mu}}{1 - e^{-\mu}}, \quad y = 1, 2, \dots$$

Finn forventningen $E(Y)$ og regn den ut for $\mu = 3$.

Karin skal ha gjester til pinnekjøttmiddag i julen. Hun antar at antallet kjøttstykker som hver person spiser er null-trunkert Poissonfordelt med parameter μ og uavhengig for ulike personer. Karin bruker data fra julebordet i pensjonistlaget til å estimere parameteren μ .

- c) Under julebordet observerer hun antall kjøttstykker; $y_1, \dots, y_{21}$, som de 21 gjestene spiser.

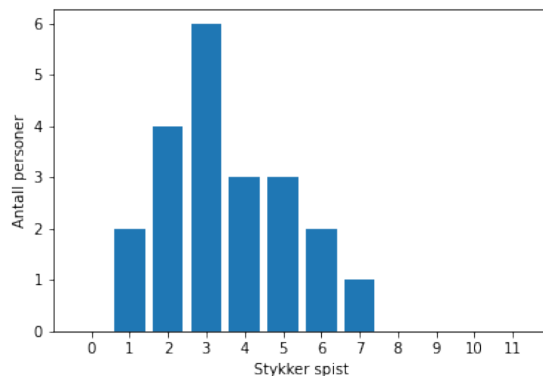
Vis at log-rimelighetsfunksjonen (log likelihood function) for μ blir

$$l(\mu) = \ln(\mu) \sum_{i=1}^{21} y_i - \sum_{i=1}^{21} \ln y_i! - 21\mu - 21 \ln(1 - e^{-\mu})$$

Bruk første-orden Taylorutvikling: $\ln(1 - e^{-\mu}) \approx \ln(1 - e^{-3}) + \frac{e^{-3}}{1 - e^{-3}}(\mu - 3)$ til å finne en approksimasjon til sannsynlighetsmaksimeringsestimatoren $\hat{\mu}$.

Gjestene spiser tilsammen 74 stykker. Regn ut estimatet fra oppgitte tall.

Figur 1 viser histogrammet av antall stykker julebordsdeltakerne spiser. Bruk punktsannsynligheten for null-trunkert Poissonfordeling til å diskutere om modellen stemmer bra med histogrammet sin oppsummering av data for 21 personer. (Du kan bruke tabell for Poissonfordeling som del av utregningen, med parameter nærmes den du fant ved sannsynlighetsestimering.)



Figur 1: Datainnnsamling av spiste stykker i forbindelse med julebord. Det er 21 julebordsdeltakere.

Oppgave 2 Lyspærer

- a) La stokastisk variabel X være eksponensialfordelt slik at sannsynlighetstettheten er $f(x) = \frac{1}{\beta} \exp(-x/\beta)$, $x > 0$. Vi setter her $\beta = 1$.

Regn ut $P(X < 3)$.

Finn teoretisk median m til X , altså verdien av konstanten m slik at $P(X \leq m) = 0.5$.

- b) Et toalett har tre lyspærer. Man antar at levetidene til lyspærene, X_1 , X_2 og X_3 , er uavhengige og eksponensialfordelte med parameter $\beta = 1$ år. Det vil være lys i rommet så lenge minst en av lyspærene virker. Vi definerer Y som tiden til toalettet blir helt mørkt.

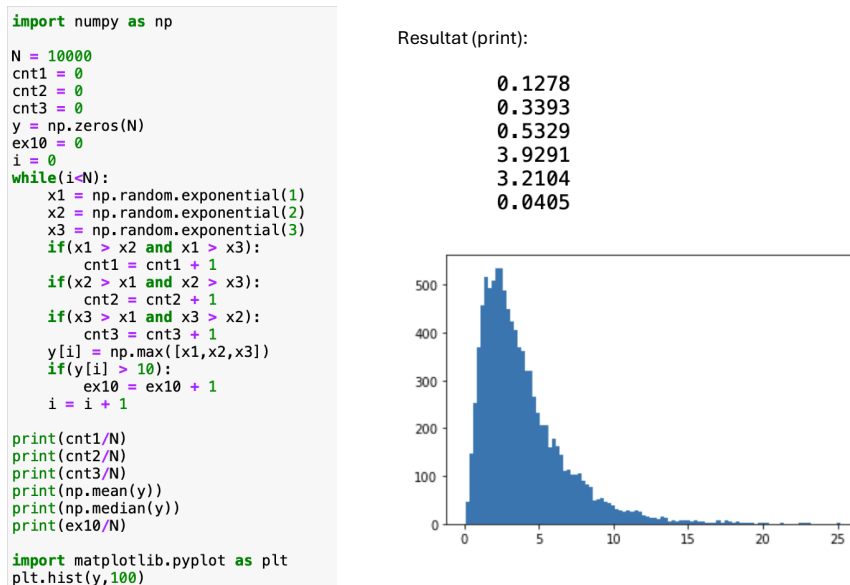
Hvordan kan hendelsen $Y \leq y$ uttrykkes som en hendelse som involverer X_1 , X_2 og X_3 ?

Vis at $P(Y \leq y) = (1 - e^{-y})^3$, $y > 0$.

Regn ut sannsynlighetstettheten til Y .

- c) Anta nå at de tre lyspærene er av ulike typer. De har eksponensialfordelte uavhengige levetider X_1 med parameter $\beta = 1$, X_2 med parameter $\beta = 2$ og X_3 med parameter $\beta = 3$.

Man kan bruke Monte Carlo simulering til å forstå de statistiske egenskapene til denne situasjonen. Forklar hvordan dette gjøres ved hjelp av Pythonkoden i Figur 2. Hva viser resultatene og plott til høyre i figuren?



Figur 2: Monte Carlo simulering av levetider for tre lyspærer. Pythonkode (venstre) og resultat av koden (høyre).

Oppgave 3 Skigåing

- a) X er normalfordelt med forventning $\mu = 10$ og standardavvik $\sigma = 0.5$.

Regn ut $P(X < 9)$.

Hva er c gitt at $P(X > c) = 0.99$?

Johannes er en dedikert skiløper. En konkurrent sier at han bruker 10 minutter opp en bakke, under gjennomsnittlig dagsform og føreforhold. Johannes tenker at han er raskere enn sin konkurrent i denne bakken. For å undersøke dette, går han på ski opp bakken ti ulike dager, under ulike føreforhold og avhengig av dagsform. La x_i = tiden på dag $i = 1, \dots, 10$. Tidene hans er 10.1, 9.3, 9.75, 9.9, 10.15, 9.55, 9.8, 9.95, 9.45, 9.6. Det oppgis at $\sum_{i=1}^{10} x_i = 97.55$ og at $\sum_{i=1}^{10} x_i^2 = 952.3$

- b) Anta at tida det tar å gå opp bakken på ski er normalfordelt med forventning μ og ukjent varians σ^2 . Vi antar videre at tidene er uavhengige.

Formuler Johannes sin situasjon som en hypotesetest.

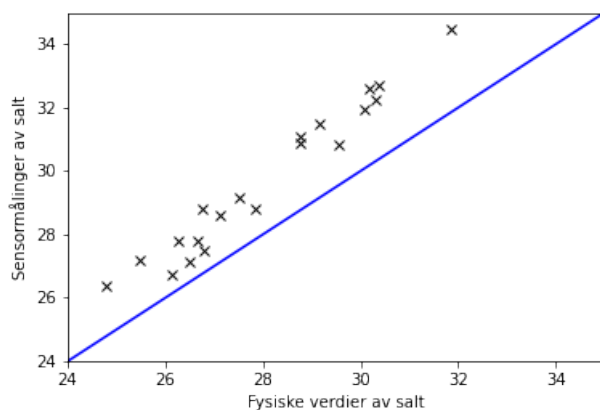
Gjennomfør hypotesetesten med signifikansnivå 0.05.

Det er vanskelig å undersøke om data faktisk er normalfordelt her. Gjennomfør i stedet en fortegnstest for hypotesen, med signifikansnivå 0.05.

Oppgave 4 Saltmålinger

En fysisk modell brukes for å forstå variabilitet i saltinnhold i Trondheimsfjorden. I tillegg kan saltinnholdet måles med en sensor på utvalgte steder. Slike målinger brukes blant annet til å kalibrere den fysiske modellen. Vi lar x_i og y_i være henholdsvis resultat av fysisk modell og sensormåling. Her indikerer $i = 1, \dots, n$ steder i fjorden der man på gitte tidspunkt har tilgjengelig begge variable.

Man foreslår regresjonsmodellen $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \dots, n$, der det antas at $\epsilon_i \sim N(0, \sigma^2)$ og uavhengige. Utfra $n = 20$ målinger (Figur 3) har vi gjennomsnitt $\bar{x} = 28.0$ og $\bar{y} = 29.6$, og $\sum_{i=1}^n (x_i - \bar{x})^2 = 64.6$ og $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 77.2$.



Figur 3: Plott av fysiske modellberegninger (førsteaksen) og sensormålinger (andreaksen). Den heltrukne blå linja viser $y = x$.

- a) Regn ut estimat for β og α .

Utfra residualene estimeres σ^2 med $s^2 = 0.48^2$. Regn ut et 90 % konfidensintervall for β .

- b) Et sted gir den fysiske modellen $x_0 = 30$. Finn predikert sensormåling der.

For hvilket resultat i den fysiske modellen vil usikkerheten i predikert sensormåling være minst?

Uten å regne, hva kan du si om vi stedet skulle predikert en sensormåling for $x_0 = 20$?