Plan for dagens undervisning

- Avslutte eksemplet fra forrige uke: Sentralgrenseteorem og diskrete fordelinger
- Regne på eksempler/situasjoner som inkluderer
  - estimere lyshastigheten i luft basert på Michelsons data fra 1879
  - regne ut SME i normalfordeling
  - sjekke om estimatorer er forventningsrette, finne varians til estimatorer
  - SME i en situasjon hvor det ikke fungerer å derivere og sette lik null

Sentralgrenseteoremet for poissonfordelte variabler

- Anta $X_1, X_2, \dots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

Sentralgrenseteoremet for poissonfordelte variabler

- ⋆ Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- ⋆ Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

- ⋆ $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y = n\overline{X} = \sum_{i=1}^{n} X_i$
- ⋆ Oppgave: For $\lambda = 1$ og $n = 40$ finn

$$P\left(\overline{X} < \frac{37}{40}\right) = P(Y < 37) = P(Y \leq 36)$$

Sentralgrenseteoremet for poissonfordelte variabler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y = n\overline{X} = \sum_{i=1}^{n} X_i$
- Oppgave: For $\lambda = 1$ og $n = 40$ finn

$$P\left(\overline{X} < \frac{37}{40}\right) = P(Y < 37) = P(Y \leq 36)$$

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $Var[X_i] = \sigma^2$. La

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er gjennomsnittet av X_i-ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \to \infty$.

Sentralgrenseteoremet for poissonfordelte variabler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y = n\overline{X} = \sum_{i=1}^{n} X_i$
- Oppgave: For $\lambda = 1$ og $n = 40$ finn

$$P\left(\overline{X} < \frac{37}{40}\right) = P(Y < 37) = P(Y \leq 36)$$

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $Var[X_i] = \sigma^2$. La

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er gjennomsnittet av X_i-ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \to \infty$.

- I vår situasjon:

$$Z = \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \approx N(0, 1)$$

Sentralgrenseteoremet for poissonfordelte variabler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y = n\overline{X} = \sum_{i=1}^{n} X_i$
- Oppgave: For $\lambda = 1$ og $n = 40$ finn

$$P\left(\overline{X} < \frac{37}{40}\right) = P(Y < 37) = P(Y \leq 36)$$

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $Var[X_i] = \sigma^2$. La

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er gjennomsnittet av X_i-ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \to \infty$.

- I vår situasjon:

$$Z = \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \approx N(0,1) \quad \Leftrightarrow \quad Y = n\overline{X} = n(\sqrt{\frac{\lambda}{n}}Z + \lambda) \approx N(n\lambda, n\lambda)$$

Sentralgrenseteoremet for poissonfordelte variabler

- Anta $X_1, X_2, \ldots, X_n$ uavhengige og identisk poissonfordelte, $X_i \sim \text{Poisson}(\lambda)$
- Har da at

$$\mu = E[X_i] = \lambda \qquad \text{og} \qquad \sigma^2 = \text{Var}[X_i] = \lambda$$

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Y = n\overline{X} = \sum_{i=1}^{n} X_i$
- Oppgave: For $\lambda = 1$ og $n = 40$ finn

$$P\left(\overline{X} < \frac{37}{40}\right) = P(Y < 37) = P(Y \leq 36)$$

Teorem (Sentralgranseteoremet)

La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $Var[X_i] = \sigma^2$. La

$$Z = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er gjennomsnittet av X_i-ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \to \infty$.

- I vår situasjon:

$$Z = \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \approx N(0,1) \quad \Leftrightarrow \quad Y = n\overline{X} = n(\sqrt{\frac{\lambda}{n}}Z + \lambda) \approx N(n\lambda, n\lambda)$$

Sannsynlighetsmaksimeringsprinsippet og SME

Definisjon (Sannsynlighetsmaksimeringsprinsippet)

Dersom man skal estimere verdien til en parameter θ ut fra observerte verdier $x_1, x_2, \ldots, x_n$, skal man velge som estimat den verdi av θ som gjør det mest sannsynlig å observere de verdiene man faktisk har observert.

Definisjon (Rimelighetsfunksjonen)

La $X_1, X_2, \ldots, X_n$ være stokastiske variabler med simultan sannsynlighetsfordeling $f(x_1, x_2, \ldots, x_n; \theta)$, der θ er en skalar parameter eller en vektor av parametre. Anta at en formel for $f(x_1, x_2, \ldots, x_n; \theta)$ er kjent, men at verdien til θ er ukjent. Hvis man har observerte verdier $x_1, x_2, \ldots, x_n$ for de stokastiske variablene $X_1, X_2, \ldots, X_n$ er rimelighetsfunksjonen gitt som

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta).$$

Definisjon (Log-rimelighetsfunksjonen)

La $L(\theta; x_1, x_2, \ldots, x_n)$ være en rimelighetsfunksjon. Log-rimelighetsfunksjonen er da

$$\ell(\theta; x_1, x_2, \ldots, x_n) = \ln L(\theta; x_1, x_2, \ldots, x_n)$$

for de verdier av θ hvor $L(\theta; x_1, x_2, \ldots, x_n) > 0$.

Regneregler for forventingsverdi og varians

- Basis regneregler for lineære funksjoner

$$
\begin{aligned}
E[aX] &= aE[X] & \quad \mathrm{Var}[aX] &= a^2\mathrm{Var}[X] \\
E[b] &= b & \quad \mathrm{Var}[b] &= 0 \\
E[X+Y] &= E[X]+E[Y] & \quad \mathrm{Var}[X+Y] &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X,Y]
\end{aligned}
$$

- Regneregler for summer

$$
E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]
$$

$$
\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathrm{Cov}[X_i, X_j]
$$

- For uavhengige stokastiske variabler $X_1, X_2, \ldots, X_n$

$$
\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]
$$

- Andre nyttige regneregler

$$
\begin{aligned}
\mathrm{Var}[X] &= E[X^2] - E[X]^2 \\
\mathrm{Cov}[X,Y] &= E[X\cdot Y] - E[X]\cdot E[Y]
\end{aligned}
$$

- Hvis X og Y er uavhengige stokastiske variabler

$$
\begin{aligned}
E[X\cdot Y] &= E[X]\cdot E[Y] \\
\mathrm{Cov}[X,Y] &= 0
\end{aligned}
$$

Noen egenskaper til normalfordeling og χ^2-fordeling

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. La $a_1, a_2, \ldots, a_n$ og b være konstanter, og

$$Y = \sum_{i=1}^{n} a_i X_i + b = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b.$$

Da er

$$Y \sim N\left(\sum_{i=1}^{n} a_i \mu_i + b, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

- Spesielt: $X \sim N(\mu, \sigma^2) \;\Leftrightarrow\; \frac{X-\mu}{\sigma} \sim N(0,1)$

Teorem

Anta at $X \sim N(0,1)$. Da er $X^2 \sim \chi_1^2$.

Teorem

La $X_1, X_2, \ldots, X_n$ være uavhengige og $X_i \sim \chi_{\nu_i}^2$. La $Y = \sum_{i=1}^{n} X_i$. Da er Y kjikvadratfordelt med $\sum_{i=1}^{n} \nu_i$ frihetsgrader.