Plan:

1. Utvalg og estimatorer
2. Beskrivelse av datasett

Mer om dette neste gang.

Repeter Forelesning 1 / Kap. 2 !

Pensum: [L] 6.1 – 6.2, 2.1 – 2.5

Repetisjon:

a) Normalfordelte variabler

$X \sim N(\mu, \sigma)$: parametre $\mu = E(X)$, $\sigma^2 = Var(X)$

X normalfordelt

X, Y normalfordelte $\Rightarrow$ $V = aX + bY$ eksakt normalfordelt (a, b konstanter)

$$\begin{cases} \mu_V = a\mu_X + b\mu_Y \\ \sigma_V^2 = a^2 \cdot \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \cdot Cov(X, Y) \end{cases}$$

b) Noen viktige formler

$X_1, X_2, \dots, X_n$ uavhengige variabler,

$\mu = E(X_i)$

$\sigma^2 = Var(X_i)$

i) $X = \sum_{i=1}^{n} X_i = X_1 + \dots + X_n$

$E(X) = n\mu$ $Var(X) = n \cdot \sigma^2$ $\boxed{\sigma_X = \sqrt{n} \cdot \sigma}$

ii) $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + \dots + X_n}{n}$

$E(\bar{X}) = \mu$ $Var(\bar{X}) = \sigma^2 / n$ $\boxed{\sigma_{\bar{X}} = \sigma / \sqrt{n}}$

X normalfordelt

$\Downarrow$

$\bar{X}$ normalfordelt

c) Sentralgrenseteoremet

$X_1, \dots, X_n$ uavhengige, identisk fordelte

$E(X_i) = \mu$ $Var(X_i) = \sigma^2$

$$\boxed{\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) \rightarrow N(\mu, \sigma/\sqrt{n}) \text{ når } n \to \infty}$$

$\mathcal{N}(\mu, \sigma/\sqrt{n})$

$n \geq 30$: god tilnærming

d) Tilnærming til binomisk fordeling

X : binomisk fordelt med parametre n, p

n stor ⇒ X tilnærmet normalfordelt

$X \longrightarrow X^{norm}$

binomisk (n,p) normal $\mu = np \quad \sigma^2 = np(1-p)$

$E(X) = np$

$Var(X) = npq = np(1-p)$

bedre tilnærming jo større n er, og jo nærmere p er 1/2.

$\sigma^2 = np(1-p) \geq 5$: god tilnærming

Tilnærming med heltallskorreksjon:

$$p(a \leq X \leq b) \simeq p(a - 1/2 \leq X^{norm} \leq b + 1/2)$$

binomisk (diskret)

$p(18) + p(19) + p(20) = p(18 \leq X \leq 20) = p(18) \cdot 1 + p(19) \cdot 1 + p(20) \cdot 1$

= summen av de tre søylene

≃ arealet under Gauss-kurven mellom $x = 17{,}5$ og $x = 20{,}5$

$17{,}5 = 18 - \frac{1}{2}$, $20{,}5 = 20 + \frac{1}{2}$

Heltallskorreksjon gir mer nøyaktige tilnærminger når vi tilnærmer en diskret variabel.

Oppgaveark 8

6. $X \sim N(2,3)$ $\quad E(X)=2 \quad Var(X)=9 \quad \mu_X=2 \quad \sigma_X=3$

$Y \sim N(5,4)$ $\quad E(Y)=5 \quad Var(Y)=16 \quad \mu_Y=5 \quad \sigma_Y=4$

$Cov(X,Y)=5$

a) $X+Y$: $N(7,\sqrt{35})$

$E(X+Y)=E(X)+E(Y)=2+5=\underline{7}$

$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

$= 9+16+10=\underline{35}$

$2X+3Y$: $N(19,\sqrt{240})$

$E(2X+3Y)=2\cdot 2+3\cdot 5=\underline{19}$

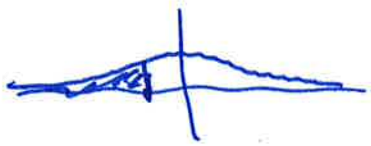
$Var(2X+3Y)=2^2\cdot 9+3^2\cdot 16$

$+2\cdot 2\cdot 3\cdot 5$

$=36+144+60=\underline{240}$

b) $P(X+Y\leq 5)=P\left(Z\leq \frac{5-7}{\sqrt{35}}\right)=P\left(Z\leq \frac{-2}{\sqrt{35}}\right)$

$=G(-2/\sqrt{35})\simeq \underline{\underline{0.368}}$



7) $X = x_1 + x_2 + \dots + X_{79}$

$p(X \le 300)$

$\simeq p(X^{norm} \le \underset{+0.5}{300})$

$= p\left(Z \le \frac{300{,}5 - 276.5}{15{,}18}\right)$

$\simeq p(Z \le 1{,}58)$

$\simeq \underline{\underline{0.943}}$

$\left(\underline{\text{Uten heltallskorreksjon}} \quad \simeq p\left(Z \le \frac{300 - 276{,}5}{15{,}18}\right) \simeq 0{,}939 \right)$

$\mu_X = E(x_1 + \dots + x_{79}) = 79 \cdot 3{,}5$

$= \underline{276.5}$

$\sigma_X^2 = Var(x_1 + \dots + x_n) = 79 \cdot Var(x_i)$

$= 79 \cdot 2.92 \simeq 230{,}4$

$\sigma_X = \underline{15{,}18}$

X tilnærmet normalfordelt pga sentralgrenseteoremet

X_i uniform

x_i	$p(x_i)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$E(x_i) = 3{,}5$

$Var(x_i) =$

$E(x_i^2) - 3{,}5^2$

$\frac{91}{6} - 3{,}5^2$

$\underline{2.92}$

$X_1, \dots, X_{79}$ identisk fordelt

— " — uavhengige

$\bar{X}_n = \frac{1}{n}(x_1 + \dots + x_n)$ normalfordelt

$n = 79 > 30$: tilnærmet normalfordelt

$X = x_1 + \dots + x_n$ tilnærmet normalfordelt.

3). R_i uniform på $[-0{,}5, 0{,}5)$
$[a, b]$

a) R_i: $E(R_i) = \frac{a+b}{2} = \underline{0}$

$Var(R_i) = \frac{(b-a)^2}{12} = \frac{1}{12} \Rightarrow \sigma_{R_i} = \sqrt{1/12}$

$\simeq \underline{\underline{0.289}}$

e) $R = R_1 + \dots + R_{100}$

$E(R) = 100 \cdot E(R_i) = 100 \cdot 0 = \underline{\underline{0}}$

$Var(R) = Var(R_1) + \dots + Var(R_{100}) = 100 \cdot \frac{1}{12} = 100/12$

$\sigma_R = \sqrt{\frac{100}{12}} \approx \underline{\underline{2.89}}$

4). $R = R_1 + R_2 + \dots + R_{100}$

R_i uniform på $[-0{,}005, 0{,}005]$

$E(R_i) = 0$

$Var(R_i) = \frac{0.010^2}{12}$

$E(R) = 100 \cdot 0 = \underline{\underline{0}}$

$Var(R) = 100 \cdot \frac{0.01^2}{12} = \frac{0.01}{12}$

$\sigma_R = \sqrt{\frac{0{,}01}{12}} \approx \underline{\underline{0{,}0289}}$

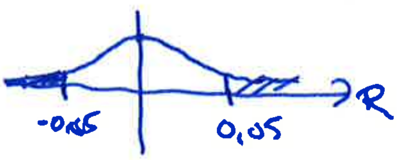
$\boxed{R \sim N(0, 0{,}0289)}$

a) $p(|R| > 0{,}05) = p(R > 0{,}05) + p(R < -0{,}05)$

$= 2 \cdot p(R < -0{,}05)$

$\simeq 2 \cdot p\left(Z \leq \frac{-0{,}05 - 0}{0{,}0289}\right) = 2 \cdot G\left(\frac{-0.05}{0{,}0289}\right)$

$\simeq \underline{\underline{0.084}}$



1 Estimatorer og utvalg

Forelesning 2-8 (Kap 3,4,5):

Sannsynlighetsteori/-regning

- Kjenner statistisk modell og dens parametre

"Anta at $X \sim N(2,5)$"

- Regner ut forskjellige ting (sannsynligheter) basert på dette.

Forelesning 1: (Kap. 2)

Beregninger basert på et datasett (utvalg)

$\bar{x}$, s_x, s_{xy}, r etc.

Statistikk

Vanlig fremgangsmåte:

- velger en statistisk modell, og forsøker å estimere parametrene.
- innhenter data fra et tilfeldig utvalg.

$X_1, X_2, \ldots, X_n$ trukket fra fordelingen til X

- antar $X_1, X_2, \ldots, X_n$ har samme fordeling som X, og uavhengige ← stokastiske variabler observasjoner:

$x_1 = 23$
$x_2 = 78$
$x_3 = 34$
$x_4 = 144$
$x_5 = 12$

(små bokstaver)
"
observasjoner

stokastiske variabler:

X_1
X_2
X_3
X_4
X_5

(store bokstaver)
"
stokastiske variabler

Estimator: stokastisk variabel som brukes til å estimere en parameter

Eks:

$$\hat{\mu} = \bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

stokastisk variabel

estimator

$$\bar{x} = \frac{23 + 78 + 34 + 144 + 12}{5} = \underline{\underline{58.2}}$$

estimat
(punktestimat)

Krav til estimatorer: $\hat{\theta}$ estimator for parameter θ
(θ = theta, gresk bokstav)

θ = en parameter vi ønsker å estimere
(μ, σ^2, p)

i) Forventningsrett: $E(\hat{\theta}) = \theta$
ii) $Var(\hat{\theta})$ minst mulig
iii) $Var(\hat{\theta}) \to 0$ når $n \to \infty$

a) Estimator for μ:

Anta $X_1, X_2, \ldots, X_n$ er uavhengige med $\begin{cases} E(X_i) = \mu. \\ Var(X_i) = \sigma^2 \end{cases}$

$$\hat{\mu} = \underline{\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)} \quad \text{estimator.}$$

i) $E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{\cancel{n}}(\cancel{n} \cdot \mu) = \mu$

$\bar{X}$ er forventningsrett.

ii) $Var(\bar{X}) = Var\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{n^{\cancel{2}}}(\cancel{n} \cdot Var(X_i))$

$$= \frac{Var(X_i)}{n} = \frac{\sigma^2}{n}$$

$$SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\sigma^2/n} \quad \leftarrow \text{standard feil til estimatoren.}$$

$$= \underline{\underline{\sigma/\sqrt{n}}}$$

iii) $SE(\bar{X}) = \sigma/\sqrt{n} \to 0$ når $n \to \infty$

b) Estimator for σ^2:

$$\hat{\sigma^2} = S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

Eks:

1	23
2	78
3	34
4	144
5	12

$\bar{x} = 58.2$

$$s^2 = \frac{(23-58.2)^2}{n-1} + \frac{(78-58.2)^2}{n-1} + \ldots$$

estimat $= \underline{2928.2}$

Formel for S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$= \frac{1}{n-1}\left(\sum X_i^2 - 2\bar{X}\sum X_i + \bar{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum X_i^2 - 2\bar{X}\cdot n\bar{X} + n\bar{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

S^2 er forventningsrett: Anta $X_1,\dots,X_n$ uavhengige med $E(X_i)=\mu$, $Var(X_i)=\sigma^2$

$$E(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i^2) - n\,E(\bar{X}^2)\right)$$

$$= \frac{1}{n-1}\left(E(X_1^2)+\dots+E(X_n^2) - n\cdot E(\bar{X}^2)\right)$$

$$= \frac{n\cdot(\mu^2+\sigma^2) - n(\mu^2+\sigma^2/n)}{n-1}$$

$$= \frac{\cancel{n\mu^2} + n\sigma^2 - \cancel{n\mu^2} - \sigma^2}{n-1}$$

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

$E(X_i^2)-E(X_i)^2 = Var(X_i)$

$\Downarrow$

$E(X_i^2) = \mu^2+\sigma^2$

$E(\bar{X}^2)-E(\bar{X})^2 = Var(\bar{X}) = \sigma^2/n$

$\Downarrow$

$E(\bar{X}^2) = \mu^2+\sigma^2/n$

c) Estimator for p:

$$\hat{p} = \frac{X}{n}$$

X binomisk med parametre (n, p)

$$E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{\cancel{n}} \cdot \cancel{n}p = = p$$

$\hat{p}$ er forventningsrett.