Plan

1. Store talls lov
2. Sentralgrenseteoremet
3. Normaltilnærmingen til binomisk fordeling

Formelsamling Forelesn. 1-7
delt ut + på nettsiden

Prøve-eksamen Forelesn. 1-7
kommer snart

Repetisjon:

Simultant fordelte stokastiske variabler

Diskret tilfelle: $p(x,y) = p(X=x, Y=y)$
simultan sannsynlighetsfordeling

Marginale fordelinger:

$p_X(x) = p(x,y_1) + p(x,y_2) + \ldots \leftarrow p(X=x)$

$p_Y(y) = p(x_1,y) + p(x_2,y) + \ldots \leftarrow p(Y=y)$

Forventning:

$E(X) = x_1 \cdot p_X(x_1) + x_2 \cdot p_X(x_2) + \ldots$

$E(Y) = y_1 \cdot p_Y(y_1) + y_2 \cdot p_Y(y_2) + \ldots$

Eks:

$Y \backslash X$	1	2	
1	0.3	0.1	0.4
2	0.2	0.4	0.6
	0.5	0.5	1

Varians:

$Var(X) = E(X^2) - E(X)^2$, $E(X^2) = x_1^2 \cdot p_X(x_1) + x_2^2 \cdot p_X(x_2) + \ldots$

$Var(Y) = E(Y^2) - E(Y)^2$, $E(Y^2) = y_1^2 \cdot p_Y(y_1) + y_2^2 \cdot p_Y(y_2) + \ldots$

Kovarians:

$Cov(X,Y) = E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right] = E(XY) - E(X)\cdot E(Y)$

$E(XY) = x_1 y_1 \cdot p(x_1,y_1) + x_1 y_2 \cdot p(x_1,y_2) + \ldots$

Korrelasjonskoeffisient: $\rho = \dfrac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$

$$= E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\cdot\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$

Resultat:

a) $-1 \leq \rho \leq 1$ for alle X, Y

b) Tolkning:

$\rho > 0$ positiv sammenheng mellom X og Y

$\rho = 1$ $Y = aX + b$ for konstanter a, b med $a > 0$

$\rho < 0$ negativ sammenheng mellom X og Y

$\rho = -1$ $Y = aX + b$ for konstanter a, b med $a < 0$

c) X og Y uavhengige $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$, $\rho = 0$

$$\boxed{\text{Defn: } p(x,y) = p_X(x)\cdot p_Y(y)}$$

Regneregler:

i) $E(aX + bY + c) = aE(X) + bE(Y) + c$

ii) $\mathrm{Var}(aX + bY + c) = a^2 \cdot \mathrm{Var}(X) + b^2 \cdot \mathrm{Var}(Y) + 2ab \cdot \mathrm{Cov}(X,Y)$

iii) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ $\quad \mathrm{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$

iv) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

v) $\mathrm{Cov}(aX + bY + c, Z) = a \cdot \mathrm{Cov}(X,Z) + b \cdot \mathrm{Cov}(Y,Z)$

Eks: $Var(X+Y) = Cov(X+Y, X+Y)$

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

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

① Store talls lov

$X_1, X_2, X_3, \ldots, X_n$: uavhengige stokastiske variable

$E(X_1) = E(X_2) = \ldots = E(X_n) = \mu$

$Var(X_1) = Var(X_2) = \ldots = Var(X_n) = \sigma^2$

$\Downarrow$

i) $X = X_1 + X_2 + \ldots + X_n$ ← summen

ii) $\overline{X} = \frac{1}{n} \cdot (X_1 + X_2 + \ldots + X_n)$ ← gjennomsnitt

$= \dfrac{X_1 + X_2 + \ldots + X_n}{n}$

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

ii) $\boxed{E(\overline{X}) = \mu \quad Var(\overline{X}) = \sigma^2/n}$ $\mu_{\overline{X}} = \mu$, $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

$E(X) = E(X_1) + E(X_2) + \ldots + E(X_n)$

$= \mu + \mu + \ldots + \mu = n\mu$

$Var(X) = Var(X_1 + X_2 + \ldots + X_n)$

$= Var(X_1) + Var(X_2) + \ldots$

$= \sigma^2 + \sigma^2 + \ldots + \sigma^2 = n\sigma^2$

$E(\overline{X}) = E\left[\frac{1}{n}(X_1 + \ldots + X_n)\right] = \frac{1}{n} \cdot n\mu = \mu$

$Var(\overline{X}) = Var\left(\frac{1}{n}(X_1 + \ldots + X_n)\right)$

$= \left(\frac{1}{n}\right)^2 \cdot Var(X_1 + \ldots + X_n) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$

Eks: Vi kaster mynt / kron $n = 10.000$ ganger

$$X_1 = \begin{cases} 1, & \text{kron i første kast} \\ 0, & \text{mynt i} \quad -\text{"}- \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{kron i andre kast} \\ 0, & \text{mynt} \quad -\text{"}- \end{cases}$$

$\vdots$

$$X_n = \begin{cases} 1, & \text{kron i n'te kast} \\ 0, & \text{mynt} \quad -\text{"}- \end{cases}$$

$E(X_1) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = 1/2$

$Var(X_1)$:

$E(X_1^2) = 1^2 \cdot \frac{1}{2} + 0^2 \cdot \frac{1}{2} = 1/2$

$Var(X_1) = \frac{1}{2} - (1/2)^2 = 1/4$

$\boxed{\mu = 1/2 \quad \sigma^2 = 1/4}$

i) $X = X_1 + X_2 + \ldots + X_n$

= antall kron på n kast.

ii) $\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$

= andel av kastene gir kron

$E(\bar{X}) = \mu = \frac{1}{2}$

$Var(\bar{X}) = \sigma^2/n = \frac{1}{4n}$

når $n \to \infty$ så vil

$\sigma_{\bar{X}} = \frac{1}{2\sqrt{n}} \to 0$

Store talls lov:

μ: Forventningen til X_i $\simeq$ gjennomsnittet $\bar{X}$ når n er stor.

$$\boxed{P(|\bar{X} - \mu| > \varepsilon) \to 0 \quad \text{når} \quad n \to \infty}$$

for enhver $\varepsilon > 0$

ε = feilmargin

② Sentralgrenseteoremet

Anta: $X_1, X_2, X_3, \ldots, X_n$: uavhengige stokastiske variabler som er identisk fordelt.

$\Downarrow$

$E(X_1) = \cdots = E(X_n) = \mu$

$Var(X_1) = \cdots = Var(X_n) = \sigma^2$

$\Downarrow$

$X = X_1 + \cdots + X_n$

$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$

$E(\bar{X}) = \mu \quad Var(\bar{X}) = \frac{\sigma^2}{n}$

Eks: $X \sim Binom(n=100, p=1/4)$

$E(X) = np = 100 \cdot \frac{1}{4} = \underline{25}$

$Var(X) = np \cdot (1-p) = 100 \cdot \frac{1}{4} \cdot \frac{3}{4}$

$= \frac{75}{4} = \underline{18{,}75}$

$\mu = 25, \quad \sigma^2 = 18{,}75$

$Y \sim N(\mu = 25, \sigma^2 = 18{,}75)$

Sentralgrenseteoremet

Når $n \to \infty$, så blir $\bar{X}$ normalfordelt.

Sentral grenseteoremet

Anta $X_1, X_2, \ldots, X_n$ er uavhengige og identiske fordelte stokastiske variable med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$. Da har vi:

$$X = X_1 + X_2 + \ldots + X_n \longrightarrow N(\mu_X = n\mu,\ \sigma_X^2 = n\sigma^2)$$

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) \longrightarrow N(\mu_{\overline{X}} = \mu,\ \sigma_{\overline{X}}^2 = \sigma^2/n)$$

når $n \to \infty$.

Eks 1: $X_1, X_2, \ldots, X_n$ er normalfordelte $\sim N(\mu, \sigma^2)$. Da er $X, \overline{X}$ er normalfordelte for alle n.

$$Z = \frac{X - \mu}{\sigma} \quad : \quad X \sim N(\mu, \sigma^2) \implies Z \sim N(0,1)$$

normalfordelt — standard normalf.

Mer generelt: X, Y normalfordelte og uavh. $\Rightarrow$ $aX + bY$ normalfordelt

$$Z = \frac{1}{\sigma}(X - \mu) = \frac{1}{\sigma} \cdot X \div \frac{\mu}{\sigma}$$

Eivind Eriksen, Kontor B4-032, eivind.eriksen@bi.no
http://www.dr-eriksen.no/teaching/MET1190/

Eks:

$X_1 \sim \text{Binom}(n=1, p=1/2)$ ← myntkast.

$X_2 \sim \text{Binom}(n=1, p=1/2)$ uavhengige

$\vdots$

$X_n \sim \text{Binom}(n=1, p=1/2)$

$X = X_1 + X_2 + \dots + X_n$ ← antall kron i n kast

$X \sim \underline{\text{Binom}(n=n, p=1/2)}$

eksakt riktig

$E(X) = n \cdot \frac{1}{2} = \frac{n}{2}$

$Var(X) = n \cdot \frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{n}{4}$

Sentralgrenseteoremet:

$X \sim N(\frac{n}{2}, \frac{n}{4})$ når n blir stor

Tilnærmingen blir bedre:

- jo større n er
- jo mer symmetrisk den opprinnelige fordelingen har

$n \geq 30$: generell "regel"

3 Normaltilnærming til binomisk fordeling

$X \sim \text{Binom}(n, p)$

binomisk fordelt variabel

$X_1 \sim \text{Binom}(n=1, p)$

$X_2 \sim \text{Binom}(n=1, p)$

$\vdots$

$X_n \sim \text{Binom}(n=1, p)$

uavhengige

$X = X_1 + X_2 + \dots + X_n$

$X \sim \text{Binom}(n, p) \leftarrow$

$$\begin{cases} E(X) = np \\ Var(X) = np(1-p) \end{cases}$$

$Y \sim N(\mu = np, \sigma^2 = np(1-p))$

Sentralgrenseteoremet: $X \simeq Y$ når n er stor

$$P(a \le X \le b) = P(X=a) + P(X=a+1) + \dots + P(X=b)$$

$$\simeq P(a \le Y \le b) \leftarrow$$ ganske god tilnærming

Heltallskorreksjon: $\boxed{P(a \le X \le b) = P(a-0.5 \le Y \le b+0.5)}$