Plan:

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

Repetisjon:

a) Simultant fordelte variabler X, Y:

Hvis X, Y er diskrete, kan simultan sannsynlighet

$$p(x,y) = p(X=x, Y=y)$$

gis som en tabell av sannsynligheter.

Eks:

$p(1,2)$ $=$ $p(X=1, Y=2)$

X\Y	1	2	3	
1	0.1	0.2	0.1	0.4
2	0.2	0.1	0.1	0.4
3	0.1	0.05	0.05	0.2
	0.4	0.35	0.25	

$p(Y=2)$

Marginale sannsynligheter:

$$p_X(x) = p(X=x) = \sum_y p(x,y)$$

$$p_Y(y) = p(Y=y) = \sum_x p(x,y)$$

finnes ved å summere rader/kolonner

Forventning:

$$E(X) = \sum_x x \cdot p_X(x) \qquad E(Y) = \sum_y y \cdot p_Y(y)$$

$$E(X^2) = \sum_x x^2 \cdot p_X(x) \qquad E(Y^2) = \sum_y y^2 \cdot p_Y(y)$$

kan finnes fra marginale sannsynligheter

$$E[g(X,Y)] = \sum_{x,y} g(x,y) \cdot p(x,y)$$

$$E(XY) = \sum_{x,y} xy \cdot p(x,y)$$

må bruke simultan sannsynlighet

Kan gjøre tilsvarende om X, Y er kontinuerlige men litt mer teknisk (ikke pensum)

Varians:

$$Var(X) = E[(X-\mu_X)^2] = E(X^2) - E(X)^2 \qquad \sigma_X = \sqrt{Var(X)}$$

$$Var(Y) = E[(Y-\mu_Y)^2] = E(Y^2) - E(Y)^2 \qquad \sigma_Y = \sqrt{Var(Y)}$$

Kovarians:

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)\cdot(Y-\mu_Y)]$$
$$= E(XY) - E(X)\cdot E(Y)$$

$E(XY) \neq E(X)\cdot E(Y)$

Korrelasjonskoeffisient:

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

Resultat:

a) $-1 \leq \rho \leq 1$

b) $\rho > 0$ betyr at X større eller sannsynligheten for at Y er større

$\rho = 1$: betyr $Y = aX + b$ med $a > 0$

c) $\rho < 0$ betyr at X større eller sannsynligheten for at Y er mindre

$\rho = -1$: betyr at $Y = aX + b$ med $a < 0$

Nyttige formler:

i) $E(X+Y) = E(X) + E(Y)$

ii) $E(aX+b) = a\cdot E(X) + b$

iii) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

iv) $\mathrm{Var}(aX+b) = a^2\cdot \mathrm{Var}(X)$

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

b) Uavhengighet:

X og Y uavhengige $\overset{}{\underset{\text{defn.}}{\Longleftrightarrow}}$ $p(x,y) = p_X(x) \cdot p_Y(y)$

$\boxed{\text{Resultat: Hvis } X \text{ og } Y \text{ er uavhengige, så er } Cov(X,Y)=0.}$ ← Det omvendte gjelder ikke.

Viktig formel:

$X_1, X_2, \dots, X_n$ uavhengige $\Rightarrow$ $\boxed{Var(a_1X_1 + \dots + a_nX_n) = a_1^2\, Var(X_1) + \dots + a_n^2 \cdot Var(X_n)}$

Spesialtilfelle: $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$
(X_1, X_2 uavh.)

Oppgaveark 7:

5) $\left.\begin{matrix} R_1 \\ R_2 \end{matrix}\right\}$ avkastning (gevinst) to aksjer

$E(R_1) = 1 \quad \sigma_{R_1} = 0.10$
$E(R_2) = 0.80 \quad \sigma_{R_2} = 0.12$
$\rho = -0.80$

$\rho = \dfrac{Cov(R_1,R_2)}{\sigma_{R_1} \cdot \sigma_{R_2}}$

a) Aksje I: størst forventet avkastn.
minst variasjon i avkastning

b) $R = 0.50R_1 + 0.50R_2$ total avkastning.

$E(R) = E(0.5R_1 + 0.5R_2) = 0.5 \cdot E(R_1) + 0.5\, E(R_2)$
$= 0.5 \cdot 1 + 0.5 \cdot 0.80 = \underline{0.90}$

$Var(R) = Var(0.5R_1 + 0.5R_2) = 0.5^2 \cdot Var(R_1) + 0.5^2 \cdot Var(R_2)$
$+ 2 \cdot 0.5 \cdot 0.5 \cdot Cov(R_1,R_2)$
$= 0.25 \cdot 0.10^2 + 0.25 \cdot 0.12^2 + 0.5 \cdot (-0.80) \cdot 0.10 \cdot 0.12 = 0{,}0013$

$\sigma_R = \sqrt{0{,}0013} \simeq \underline{\underline{0{,}036}}$

7.) c) $X = \frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \frac{1}{2}X_4$

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

$$E(X) = \frac{1}{8}E(X_1) + \frac{1}{4}E(X_2) + \frac{1}{8}E(X_3) + \frac{1}{2}E(X_4)$$
$$= \frac{1}{8}\mu + \frac{1}{4}\mu + \frac{1}{8}\mu + \frac{1}{2}\mu = \underline{\underline{\mu}}$$

$$Var(X) = \left(\frac{1}{8}\right)^2 \cdot \sigma^2 + \left(\frac{1}{4}\right)^2 \cdot \sigma^2 + \left(\frac{1}{8}\right)^2 \cdot \sigma^2 + \left(\frac{1}{2}\right)^2 \cdot \sigma^2$$
$$= \left(\frac{1}{64} + \frac{1}{16} + \frac{1}{64} + \frac{1}{4}\right)\sigma^2 = \frac{1+4+1+16}{64}\sigma^2$$
$$= \frac{22}{64}\sigma^2 = \underline{\underline{\frac{11}{32}\sigma^2}}$$

$$\sigma_X = \sqrt{\frac{11}{32}\sigma^2} = \underline{\sqrt{\frac{11}{32}} \cdot \sigma} = \frac{\sqrt{22}}{\sqrt{64}}\sigma = \underline{\underline{\frac{\sqrt{22}}{8}\sigma}}$$

① Store talls lov

Anta $x_1, x_2, \dots, x_n$ er uavhengige og at

$E(x_i) = \mu$ og $Var(x_i) = \sigma^2$ for alle i.

Vi ser på: $X = \sum_{i=1}^{n} x_i = x_1 + x_2 + \dots + x_n$

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

Da har vi:

i) $E(X) = n\mu$

$Var(X) = n \cdot \sigma^2$

$$E(X) = E(x_1) + \dots + E(x_n)$$
$$= \mu + \mu + \dots + \mu$$
$$Var(X) = Var(x_1) + Var(x_2) + \dots + Var(x_n)$$
$$= \sigma^2 + \sigma^2 + \dots + \sigma^2$$

$$\boxed{\begin{array}{l} \mu_X = n\mu \\ \sigma_X = \sqrt{n} \cdot \sigma \end{array}}$$

ii) $E(\bar{X}) = \mu$

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

$$E(\bar{X}) = E\left[\frac{1}{n}(x_1 + \dots + x_n)\right]$$
$$= \frac{1}{n} \cdot n\mu = \mu$$
$$Var(\bar{X}) = Var\left(\frac{1}{n}(x_1 + \dots + x_n)\right)$$
$$= Var\left(\frac{1}{n}x_1 + \frac{1}{n}X_2 + \dots\right)$$
$$= \frac{1}{n^2} \cdot \sigma^2 + \frac{1}{n^2} \cdot \sigma^2 + \dots + \frac{1}{n^2}\sigma^2$$
$$= \frac{1}{n^2}(\cancel{n}\sigma^2) = \frac{1}{n} \cdot \sigma^2$$

$$\boxed{\begin{array}{l} \mu_{\bar{X}} = \mu \\ \sigma_{\bar{X}} = \sigma / \sqrt{n} \end{array}}$$

Eks: Vi kaster en mynt 10.000 ganger.

$$x_i = \begin{cases} 1, & \text{kron på kast } i \\ 0, & \text{mynt} \text{ —"— } \end{cases}$$

$x_1, x_2, \ldots, x_{10.000}$: uavhengige

x_i	$p(x_i)$
0	1/2
1	1/2

$$E(x_i) = 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = \underline{1/2}$$

$$\begin{aligned} Var(x_i) &= E(x_i^2) - E(x_i)^2 \\ &= \left(0^2 \cdot \tfrac{1}{2} + 1^2 \cdot \tfrac{1}{2}\right) - (1/2)^2 \\ &= 1/2 - 1/4 = \underline{1/4} \end{aligned}$$

$$\boxed{\begin{aligned} \mu &= 1/2 \\ \sigma &= 1/2 \end{aligned}} \leftarrow$$ forventning og std. avvik for hver x_i

i) $X = x_1 + x_2 + \ldots + x_{10.000}$ = totalt antall kron.

$$E(X) = 10.000 \cdot \tfrac{1}{2} = \underline{5.000}$$

$$\sigma_X = \sqrt{10.000} \cdot \tfrac{1}{2} = \underline{50}$$

$$\boxed{\mu_X = n\mu}$$

$$\boxed{\sigma_X = \sqrt{n} \cdot \sigma}$$

ii) $\bar{X} = \dfrac{x_1 + x_2 + \ldots + x_{10000}}{10.000}$ = andel kast som gir kron

$$\mu_{\bar{X}} = \tfrac{1}{2} = 0.50$$

$$\sigma_{\bar{X}} = (1/2) / \sqrt{10.000} = 0{,}005$$

$$\boxed{\begin{aligned} \mu_{\bar{X}} &= \mu \\ \sigma_{\bar{X}} &= \sigma/\sqrt{n} \end{aligned}}$$

Empirisk: $\underline{0{,}5067}$ (Kerrich kastet mynt/kron 10.000 ganger og fikk kron 5.067 ganger)

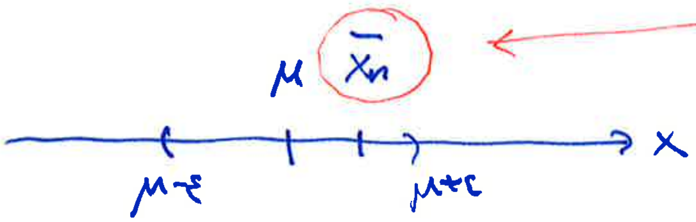
Store talls lov: "n stor $\Rightarrow \bar{X} \approx \mu$"

Hvis $X_1, X_2, \ldots$ er uavhengige stokastiske variable med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$ for alle i, og $\varepsilon > 0$ vilkårlig feilmargin, så har vi:

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

når $\bar{X}_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

$$\Leftrightarrow \boxed{P(|\bar{X}_n - \mu| \leq \varepsilon) \to 1}$$



Sannsynligheten for at $\bar{X}_n$ ligger innenfor feilmarginen, dvs i $(\mu - \varepsilon, \mu + \varepsilon)$ går mot 1 når $n \to \infty$.

Essensen: Når $n \to \infty$ så vil $\sigma_{\bar{X}_n} = \dfrac{\sigma}{\sqrt{n}} \to 0$.

Tolkning: Forventning μ ← fra teori

$\approx$

gjennomsnitt ved mange gjentagelser ← empiri

($\bar{X}_n$ når n er stor)

② Sentralgrenseteoremet

"normalfordelingen er normal"

Eks: $Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma}\cdot X - \frac{\mu}{\sigma}$

Viktige egenskaper ved normalfordelingen:

i) X normalfordelt $\Rightarrow$ $aX+b$ normalfordelt $(a \neq 0)$

ii) X, Y normalfordelt $\Rightarrow$ $X+Y$ —"—

Det betyr:

$X_1, X_2, \ldots, X_n$ normalfordelt $\Rightarrow$ i) $X = \sum x_i = x_1 + \ldots + x_n$, ii) $\bar{X} = \frac{1}{n}\sum x_i = \frac{1}{n}(x_1 + \ldots + x_n)$ normalfordelte

Sentralgrenseteoremet

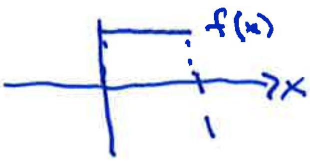
Anta: $X_1, X_2, \ldots, X_n, \ldots$ uavhengige, identisk fordelte (iid) stokastiske variabler, med forventning $E(x_i) = \mu$ og varians $Var(x_i) = \sigma^2$

Da har vi: $\bar{X}_n = \frac{x_1 + \ldots + x_n}{n}$ $E(\bar{X}_n) = \mu$, $Var(\bar{X}_n) = \sigma^2/n$

Når $n \to \infty$, så vil $\bar{X}_n \to N(\mu, \sigma/\sqrt{n})$

Hva betyr $\bar{X}_n \to N(\mu, \sigma/\sqrt{n})$?

Eks:

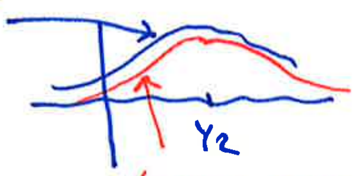


X_i uniform $[0,1]$

$E(x_i) = \frac{1}{2}$ $Var(x_i) = \frac{1}{12}$

$\bar{X}_n \to N(\mu, \sigma/\sqrt{n})$ betyr at tetthetsfn. til $\bar{X}_n$ ~~nærmer~~ seg tetthetsfn. til normalfordelingen (Gauss-kurve)

tetthets-fn. for $\bar{X}_n$.



Gauss-kurve med $\mu = 1/2$ $\sigma = 1/\sqrt{12}\cdot\frac{1}{\sqrt{n}}$

Eks: Vi kaster en mynt n ganger

$$X_i = \begin{cases} 1, & \text{kron på } i\text{'te kast} = \text{kast nr. } i \\ 0, & \text{mynt} - \text{"} - \end{cases}$$

$X = X_1 + \ldots + X_n$ ← totalt ant kron

Forutsetninger:
- uavhengighet : ok
- identisk distribuert : ok

X_i:

x_i	$p(x_i)$
0	1/2
1	1/2

Binomisk: $n = 1$, $p = 1/2$

Sentral grense teoremet:

$\bar{X}_n \to$ normalfordeling når $n \to \infty$.

$$\boxed{X: \text{ binomisk } n,\ p = 1/2}$$

$X = X_1 + \ldots + X_n$ normalfordelt

$\Updownarrow$

$\bar{X}_n = \dfrac{X_1 + \ldots + X_n}{n}$ normalfordelt.

Dette betyr at den binomiske fordelingen til X er tilnærmet normal fordelt

Eks: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = \dfrac{1/2}{1 - \frac{1}{2}} = 1$ ← konvergerer raskt

$n = 5$: $\frac{31}{32} \approx 0.97$

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{b\to\infty} \left[-\frac{1}{x}\right]_1^b = \lim_{b\to\infty} \left(-\frac{1}{b} + 1\right) = 1$$

↑ konvergerer sakte

$b = 5$: 0.80

$b = 30$: $1 - \frac{1}{30} \approx 0.97$

Tommelfingerregel: $\boxed{\text{Hvis } n \geq 30 \text{ og (uavh, identisk data) så er } \bar{X}_n \text{ tilnærmet normalfordelt.}}$

Merk: - skjeve fordelinger konvergerer saktere enn symmetriske fordelinger

- fordelinger med lange haler konvergerer saktere

③ Normaltilnærming til binomiske variable

Hvis X er binomisk fordelt med (n, p), og n er stor og p er nær $1/2$, så har vi

$\binom{n}{a} p^a q^{n-a} + \binom{n}{a+1} p^{a+1} q^{n-a-1} + \ldots \longrightarrow$

$$p(a \leq X \leq b) \simeq p(a - 0.5 \leq X^{norm} \leq b + 0.5)$$

± 0.5: heltallskorreksjon gjør tilnærmingen mer nøyaktig

X^{norm}: normalfordelt variabel med

$\mu = np$

$\sigma = \sqrt{npq}$

Merk: Heltallskorreksjon brukes når X er diskret og tilnærmes med normalfordelt variabel (kontinuerlig), og gir større nøyaktighet enn

$$p(a \leq X \leq b) \simeq p(a - 0.5 \leq X^{norm.} \leq b + 0.5)$$

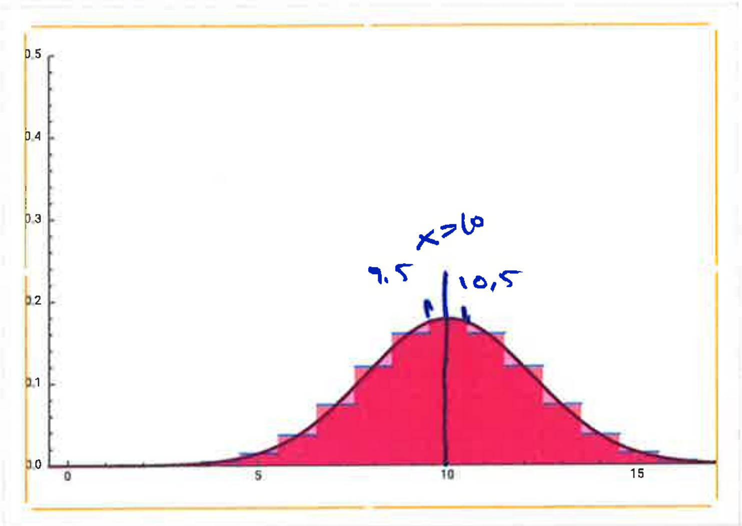
heltalls-korreksjon

$$p(a \leq X \leq b) \simeq p(a \leq X^{norm} \leq b)$$

tilnærming uten heltallskorreksjon

Normal Approximation to a Binomial Random Variable

parameter n 20
parameter p 0.5
compute agreement



Download to Desktop | Copy to Clipboard | Source

A binomial random variable with parameters n and p can be thought of as a sum of independent Bernoulli random variables, each with parameter p. The central limit theorem implies that for large values of n a binomial random variable can be well approximated by a normal random variable with the same mean and variance. A measure of agreement between the two is obtained by computing the purple area; 100% represents complete agreement between the two distributions. The computation of the purple area is slow, so use the sliders with the "compute agreement" box unchecked.

Contributed by: Chris Boucher (March 2011)

Share Demonstration

Related Demonstrations More by Author

Binomial Approximation to a Poisson Random Variable
Chris Boucher

Binomial Approximation to a Hypergeometric Random Variable
Chris Boucher

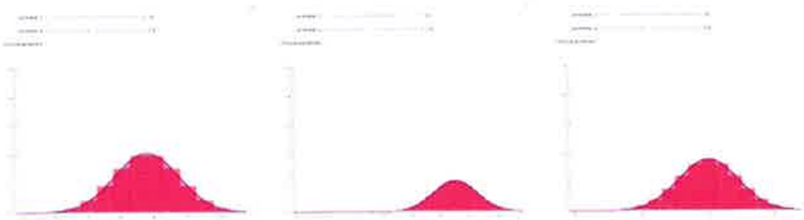

Normal Approximation to a Poisson Random Variable
Chris Boucher


The Arcsine Transformation of a Binomial Random Variable

Related Topics

College Mathematics »
Probability »
Statistics »
Browse all topics »

SNAPSHOTS



RELATED LINKS

Binomial Distribution (Wolfram *MathWorld*)
Normal Distribution (Wolfram *MathWorld*)