Metody Statystyczne

wykład 3

5 XII 2020

kacper.topolnicki@uj.edu.pl

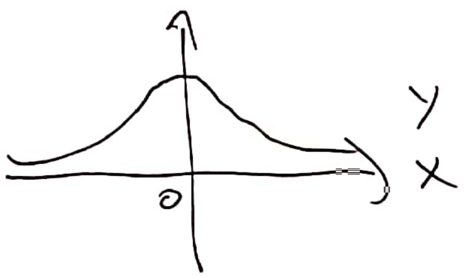
Plan:

- rozkład normalny
- estymacji punktowej
- estymacja przedziałowa
- procesy stochastyczne
  - łańcuchy Markova

[rozkład normalny

(x X)

założenia:

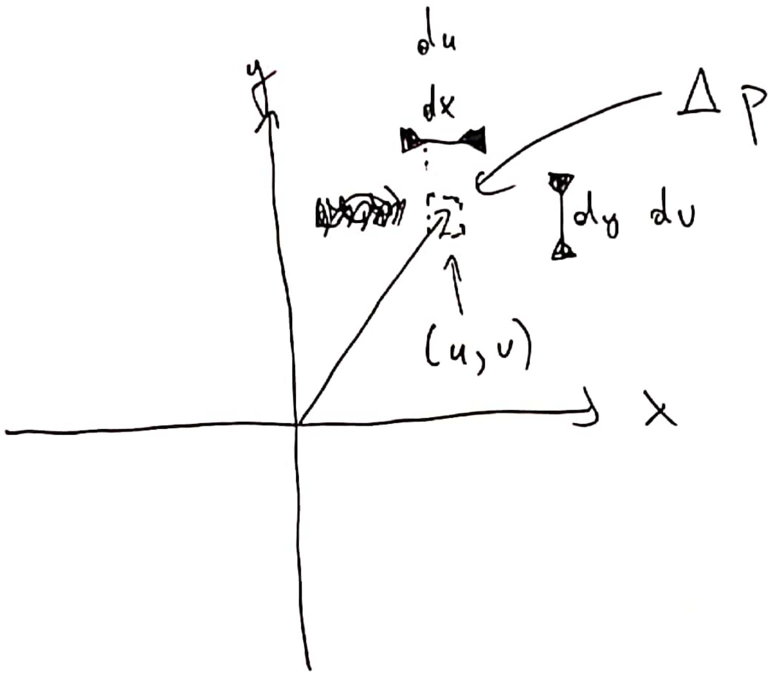


- X, Y – niezależne zmienne losowe, ~~taką~~ samą RGP

- FGP dla X, Y jest ciągła i symetryczna względem 0

- FGP dla X, Y → 0 „z dala" od $x=0$, $x \approx 0$

$f_X(-x) = f_X(x)$, $f_X(x) \to 0$ gdy $x \to \pm\infty$



$dx, dy >>$ мате?

$$\underbrace{f_{\cancel{X,Y}\vec{r}}(x, y)}_{\text{RGP dla X ono7 X}}$$

$$\Delta p = f_x(u)\,du \cdot f_y(v)\,dv = f_{\vec{r}}(u,v)\,\underbrace{du\,dv}$$

$$f_x = f_y = f$$

$$\underbrace{f(u)\,f(v)}\,du\,dv = \underbrace{f_{\vec{r}}(u,v)}\,du\,dv$$

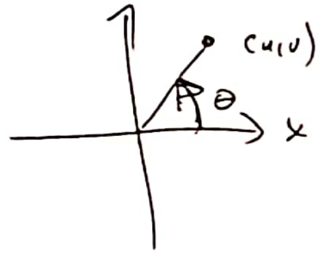
$$f_{\vec{r}}(u, v) = f(u) \cdot f(v)$$

$$f_{\vec{r}}(x, y) = g(x^2 + y^2)$$

u

$$f_{\vec{r}}(r, \theta)$$

$r = \sqrt{u^2 + v^2}$



$$\begin{cases} u = r \cos \theta \\ v = r \sin \theta \end{cases}$$

$$\frac{\partial f_{\vec{r}}(r, \theta)}{\partial \theta} = f(u) \frac{\partial f(v)}{\partial \theta} + \frac{\partial f(u)}{\partial \theta} f(v)$$

$\|$

0

$$0 = f(u) \cdot \frac{\partial f(u)}{\partial \theta} + \frac{\partial f(u)}{\partial \theta} f(v)$$

$r \sin(\theta)$ $\quad$ $r \cos(\theta)$

$$0 = f(u) \cdot f'(v) \underbrace{r \cos(\theta)}_{u} + \underbrace{(-r \sin \theta)}_{v} f'(u) f(v)$$

$$v f'(u) f(v) = u f(u) f'(v)$$

$$\frac{f'(u)}{u f(u)} = \frac{f'(v)}{v f(v)}$$

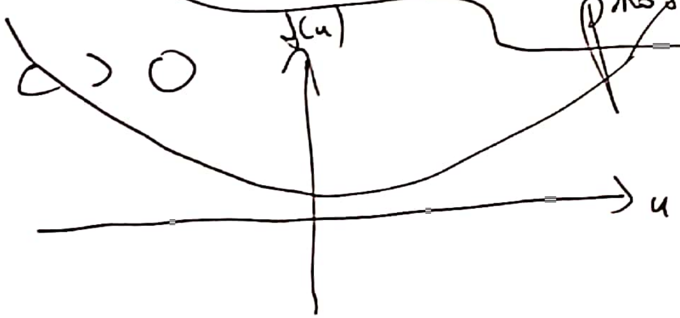
spetnuone allo
kaidego. u, v

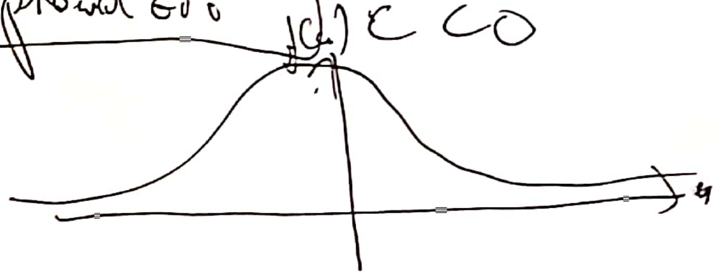
$$\frac{f'(u)}{u\, f(u)} = C \qquad\qquad \frac{f'(v)}{v\, f(v)} = C$$

$$f'(u) = C\, u \cdot f(u) \qquad\qquad f'(v) = C\, v\, f(v)$$

$$f(u) = A \cdot e^{\frac{C}{2} u^2} \qquad\qquad f(v) = A \cdot e^{\frac{C}{2} v^2}$$

proszę to sprawdzić





$c < 0$ ons $\int_{-\infty}^{+\infty} du \, f(u) = 1$

$e^x = \exp(x)$

$$\begin{cases} f(u) \overset{\mu=0}{=} \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}\dfrac{(u-\cancel{\mu})^2}{\sigma^2}\right) \\ f(u) \overset{\mu=0}{=} \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}\dfrac{(u-\cancel{\mu})^2}{\sigma^2}\right) \end{cases}$$

mozhted Normaly

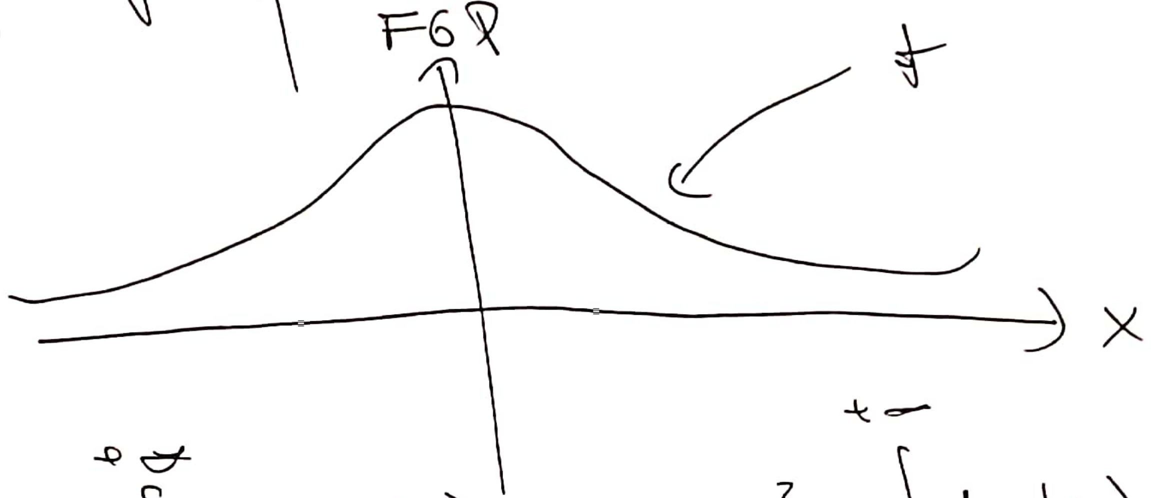
σ^2 — wariancja

μ — wartość oczekiwana

$\underbrace{X, Y}_{\text{rozkład normalny}}$ $\qquad$ $\underbrace{Z = X + Y}_{\text{jaki jest w FGP}}$

[zadanie domowe

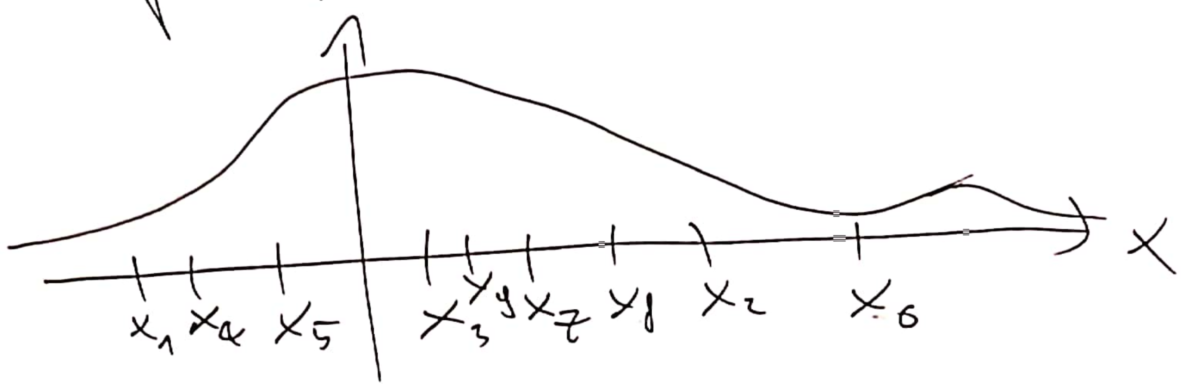
— w materiałach dodatkowych — centralne twierdzenie graniczne [estymacja punktowa



$$E(x) = \int_{-\infty}^{+\infty} dx\, f(x) \cdot x \qquad \sigma^2 = \int_{-\infty}^{+\infty} dx\, f(x)\,(x - E(x))^2$$

$\rho_{x,y}$ – współczynnik korelacji dla 2 zmiennych losowych.

II w praktyce



eksperyment: zamiast FGP

dane: $\{x_1, x_2, x_3, \ldots, x_N\}$

jak oszacować $E(X)$, $\sigma(X)$, $\rho_{X,Y}$?

[?] jakich własności oczekujemy od estymatorów?

$T_n(\theta)$ — jest zmienną losową

θ — wielkość estymowana

— estymator nieobciążony (unbiased estimator)

$$E(T_n(\theta)) = \theta_0$$

każde x_i jest losowane z $f(x)$

θ_0 — wartość θ dla $f(x)$

E — „średnia” po $\{x_1, x_2, x_3, \ldots, x_n\}$

– estymator zgodny
(consistent estimator)

$$\operatorname{plim}_{N \to \infty} T_N(\theta) = \theta_0$$

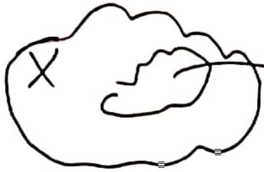
$\underbrace{N \to \infty}_{}$

zwiększamy liczbę pomiarów

Dokładnie formalnie

dla każdego ϵ: $\lim_{N \to \infty} P(|T_N(\theta) - \theta_0| > \epsilon) = 0$

- przykład



$\{x_1, x_2, x_3, \ldots x_N\}$

$$T_N(E(X)) = \frac{1}{N-1}\sum_{i=1}^{N} x_i$$

$$E\left(T_N(E(X))\right) = E\left(\frac{1}{N-1}\sum_{i=1}^{N} x_i\right) =$$

$$= \frac{1}{N-1} E\left(\sum_{i=1}^{N} x_i\right) = \frac{1}{N-1}\sum_{i=1}^{N} E(x_i) =$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} E(X) = \frac{N}{N-1} E(X)$$

losowanie

~~z takiego samego~~

z takiej samej PGP

$$\underbrace{E(T_N(R(x))) = \frac{N}{N-1} R(x)}_{\text{estymator obciążony (biased estimator)}}$$

$$\underbrace{\frac{N}{N-1} R(x) \xrightarrow[N \to \infty]{} R(x)}_{\text{estymator zgodny}}$$

- asymptotycznie normalny

$$FGP\Big(\underbrace{\overline{T}_N(\theta)}_{\text{zmienna losowa}}\Big) \xrightarrow{N\to\infty} \text{rozkładu normalnego}$$

[estymacja przedziałowa

$$\overline{T}_N(\theta)$$

$$\downarrow$$

$$\underbrace{\left[\overline{T}_N^{L}(\theta),\ \overline{T}_N^{R}(\theta)\right]}_{\text{przedział ufności}}$$

$$P\left(T_N^L(\theta) \leq \theta \leq T_N^R(\theta)\right) =$$

$$= 1 - \alpha = \gamma$$

[illegible]

poziom ufności

≈ 0.8

I. I. wartość oczekiwana

I. I. I. założenie: $\sigma(X)$ znane.

$$Z = \frac{\overline{X} - E(X)}{\sigma(\cancel{\overline{X}})} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{(\overline{X} - E(X))\sqrt{n}}{\sigma(X)}$$

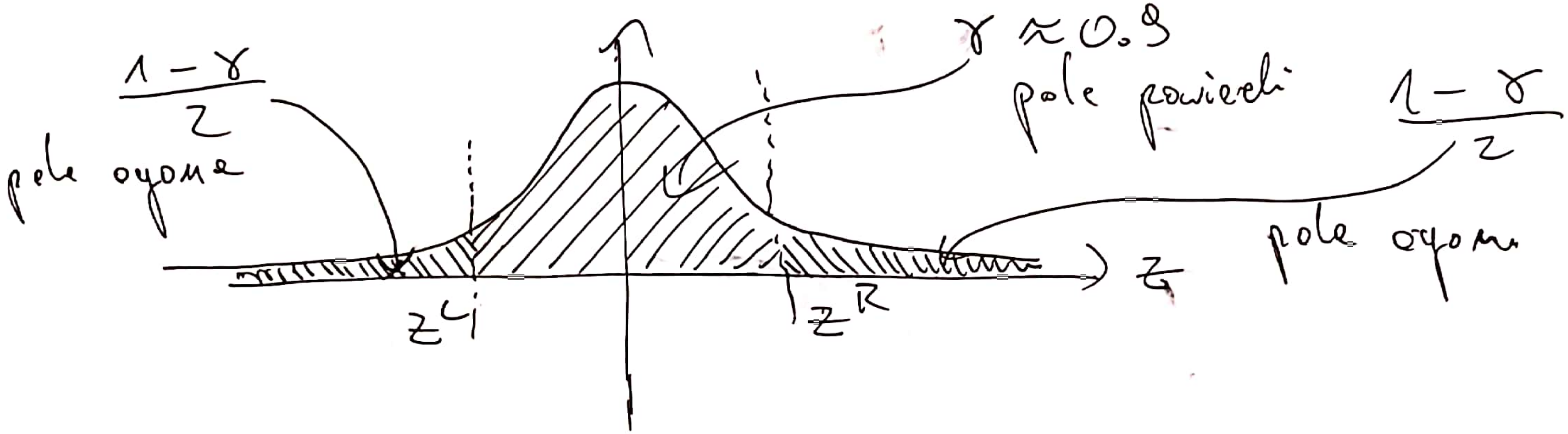
FGP(Z) – rozkład normalny

$N(0,1)$

$$P(z^L \leq z \leq z^R) = \gamma_{z_{0.9}}$$

kwantyle rozkładu normalnego

FGP



$$z^L = \underbrace{z_{\frac{1-\gamma}{2}}} = -z_{\frac{\gamma+1}{2}}$$

$$z^P = z_{\frac{1-\gamma}{2}+\gamma} = z_{\frac{\gamma+1}{2}}$$

$$P\left(z^L \leq z \leq z^R\right) = \gamma =$$

$$= P\left(z^L \leq \frac{(\overline{x} - E(x))\sqrt{N}}{\sigma(x)} \leq z^R\right) =$$

$$= P\left(\frac{z^L \sigma(x)}{\sqrt{N}} \leq \overline{x} - E(x) \leq \frac{z^R \sigma(x)}{\sqrt{N}}\right) =$$

$$= P\left(\frac{z^L \sigma(x)}{\sqrt{N}} - \overline{x} \leq - E(x) \leq \frac{z^R \sigma(x)}{\sqrt{N}} - \overline{x}\right) =$$

$$= P\left(\overline{x} - \frac{z^L \sigma(x)}{\sqrt{N}} \geq E(x) \geq \overline{x} - \frac{z^R \sigma(x)}{\sqrt{N}}\right) = \ldots$$

$$\ldots = P\left(\underbrace{\bar{X} - \frac{z_{\frac{1+\alpha}{2}}\,\sigma(x)}{\sqrt{N}}}_{T_N^L(E(x))} \le E(x) \le \underbrace{\bar{X} + \frac{z_{\frac{1+\alpha}{2}}\,\sigma(x)}{\sqrt{N}}}_{T_N^R(E(x))}\right)$$

$$\underbrace{\left[\,T_N^L(E(x)) \qquad T_N^R(E(x))\,\right]}_{\text{przedział ufności}}$$

II. nie znamy $\sigma(x)$

N – pomiarów

$$t = \frac{\left(\overline{X} - E(x)\right)\sqrt{N}}{s(x)}$$

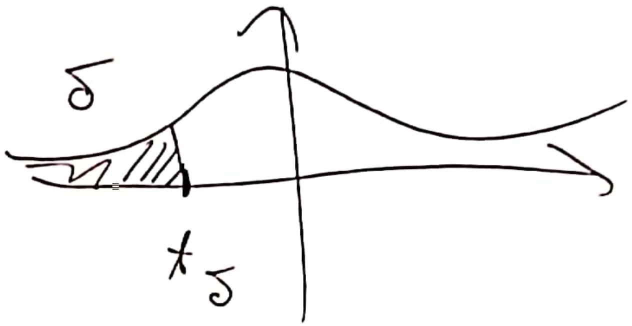
$$s(x) = \frac{1}{N-1}\sum_{i=1}^{N}\left(x - \overline{x}\right)^2$$

FGP(t)

rozkład studenta o $N-1$ stopniach swobody

$$P\left(\overline{X} - \frac{1}{\sqrt{N}} S\, t_{\frac{1+\gamma}{2}} \le E(x) \le \overline{X} + \frac{1}{\sqrt{N}} S\, t_{\frac{1+\gamma}{2}}\right) = \gamma$$

kwantyl rozkład student



$$\left[T_N^L(E(x)),\ T_N^R(E(x))\right]$$

PU wariancja przedział ufności

~~odchylenie standardowe~~

odchylenie standardowe

$$\chi = \frac{(N-1)\, s^2}{\sigma^2(x)} \qquad s = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

ma F.GP χ^2_{N-1} N-1 stopniach swobody

$$P\left(\underbrace{\sqrt{\frac{(N-1)\, s^2}{\chi^2_{N-1}\left(\frac{1+\gamma}{2}\right)}}}_{\text{CN}} \le \sigma(x) \le \underbrace{\sqrt{\frac{(N-1)\, s^2}{\chi^2_{N-1}\left(\frac{1-\gamma}{2}\right)}}}_{\text{TP}}\right) = \gamma$$

[] metody znajdowania estymatorów

— w materiałach dodatkowych

[procesy stochastyczne

X – zmienna losowa

$$X : \underset{\text{zbiór zdarzeń elementarnych}}{\Omega} \to \underset{\text{liczba}}{\mathbb{R}}$$

X(⚁) = 2

$$\underbrace{X(\underset{\text{czas}}{t}, \underset{\text{zd. losowe}}{\omega})}_{\text{proces stochastyczny}}$$

$X(t, \omega)$

- ustalony czas → $X_t(\omega)$
  - dyskretny → proces o stanach dyskretnych
    - dyskretny czas
    - ciągły czas
  - ciągłe → proces o stanach ciągłych
    - czas dyskretny
    - czas ciągły
- ustalone ω → $X_\omega(t)$

śledzenie
procesu
szkodliwego

— zużycie procesora

— mierzymy temperaturę co godzinę

zadanie domowe

II procesy Markova

$$P\left(X(t)\in A \mid X(t_1)\in A_1 \wedge X(t_2)\in A_2 \wedge \ldots \wedge X(t_N)\in A_N\right)=$$

$$= P\left(X(t)\in A \mid X(t_N)\in A_N\right)$$

$$t_1 < t_2 < \ldots < t_N < t$$

P(przyszłość | teraźniejszość, przeszłość) =

= P(przyszłość | teraźniejszość)

- Projektowanie balistyczne pocisku.

- Projektowanie [illegible] z krótką pamięcią

Czy to są procesy Markova?

[[Łańcuch Markova

- dyskretny czas $t = 0, 1, 2, \ldots$
- dyskretne stany $x = 0, 1, 2, \ldots$

$$P(x(t+1) = j \mid x(t) = i) =$$

$$= P(x(t+1) = j \mid x(0) = a \wedge x(1) = b \wedge \ldots \wedge x(t) = i)$$

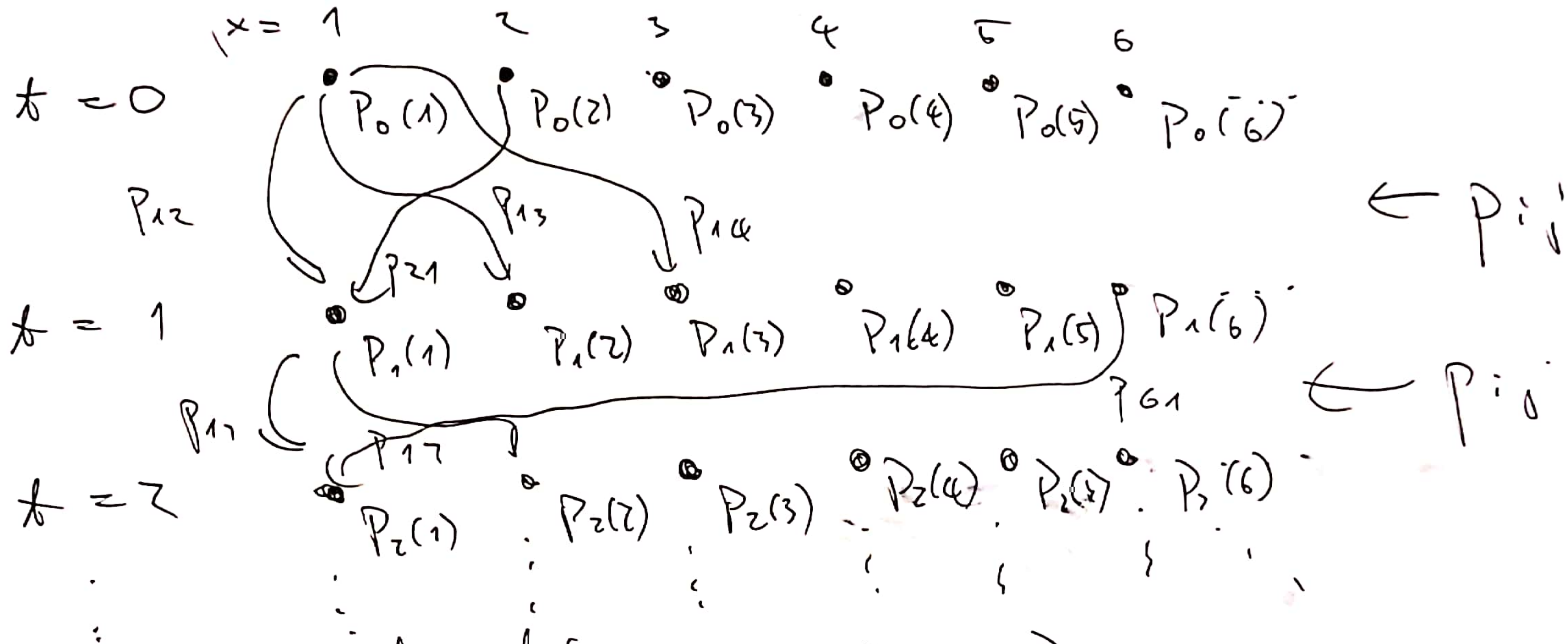
$$P_{ij}(t) \equiv P(X(t+1) = j \mid x(t) = i)$$

$p_{ij}(t) = p_{ij}$

jednorodny łańcuch Markowa

– nie zależy od czasu

III jak określić proces Markowa



— prawdopodobieństwo $P_0(x)$ — dla wszystkich x

— prawdopodobieństwo przejść p_{ij}

stan $\rightarrow$ $\left(P(1), P(2), \dots \right)$
v danym czasie

$$p_{ij} \rightarrow \begin{pmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\left(P_0(1), P_0(2), \dots \right) \begin{pmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} = \left(P_1(1), P_1(2), P_1(3) \dots \right)$$

III co dzieje się później

$$p_{ij} \longrightarrow p_{ij}^{(n)}$$

$P_n(x) \longleftrightarrow$ prawdopodobieństwo, że po czasie n zmienna losowa przyjmie wartość x

$$\lim_{n \to \infty} p_{ij}^{(n)} = ? \qquad \lim_{n \to \infty} P_n(x) = ?$$

t – :

$$p_{ij}^{(h-2)}(t_1, t_3) = \sum_k p_{ik}(t_1, t_2)\, p_{kj}(t_2, t_3)$$

$$t_1 < t_2 < t_3$$

dw :

$$p_{ij}(t_1, t_3) = P(X(t_3) = j \mid X(t_1) = i) =$$

$$= \sum_k P(X(t_3) = j \cap X(t_2) = k \mid X(t_1) = i) \underset{=\,\ldots}{\overset{P(A|B) = \frac{P(A\cap B)}{P(B)}}{=}}$$

$$= \sum_k \frac{P(X(t_3) = j \cap X(t_2) = k \cap X(t_1) = i)}{P(X(t_1) = i)} \overset{P(A\cap B) = P(A|B)\,P(B)}{=}$$

$$= \sum_k \frac{P(X(t_3)=j \mid X(t_2)=k \wedge X(t_1)=i) \cdot P(X(t_2)=k \wedge X(t_1)=i)}{P(X(t_1)=i)} =$$

$$= \sum_k \underbrace{P(X(t_3)=j \mid X(t_2)=k \wedge X(t_1)=i)} \cdot P(X(t_2)=k \mid X(t_1)=i) =$$

$$= \sum_k P(X(t_3)=j \mid X(t_2)=k) \cdot P(X(t_2)=k \mid X(t_1)=i) =$$

$$= \sum_k p_{kj}(t_2,t_3) \cdot p_{ik}(t_1,t_2) =$$

$$= \sum_k p_{ik}(t_1,t_2)\, p_{kj}(t_2,t_3)$$