# Šta su parametri raspodele?

Parametri raspodele su veličine koje zavise od funkcije raspodele.

X - sl. veličina sa f-jom raspodele F. Primeri parametara:
- EX, DX, F(t), p-ti kvantil, tj. broj t t.d. F(t) = p
- broj stepeni slobode kod $\chi^2$ raspodele, broj $\lambda$ kod $\mathcal{E}(\lambda)$ raspodele...

$EX_3 = EX, \quad DX_5 = DX \dots$

($X_1, X_2, \dots, X_n$) PSU sa istom raspodelom kao X(obeležje populacije). Neke od često korišćenih statistika za ocenu EX i DX su redom:

1. uzoračka sredina: $\overline{X}_n = \dfrac{X_1 + X_2 + \dots + X_n}{n}$
2. popravljena uzoračka disperzija: $\tilde{S}^2_n = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$

Ove ocene su nepristrasne ocene parametara EX i DX, redom!
Dokaz:

1. za vežbu. (Treba pokazati da je $E\overline{X}_n = EX$)

2.
$$
\begin{aligned}
E\tilde{S}_n^2 &= \frac{1}{n-1} E\sum_{i=1}^{n}\left(X_i^2 - 2\overline{X}_n X_i + (\overline{X}_n)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n \sum_{i=1}^{n} X_i + n\cdot(\overline{X}_n)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n \cdot n(\overline{X}_n) + n\cdot(\overline{X}_n)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n(\overline{X}_n)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n} \left(X_i^2 - (\overline{X}_n)^2\right)\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} E\left(X_i^2 - (\overline{X}_n)^2\right) = (*)
\end{aligned}
$$

$$DX = EX^2 - (EX)^2$$

---

$$E(\overline{X}_n)^2 = D\overline{X}_n + (E\overline{X}_n)^2 = D\left(\frac{X_1 + \dots + X_n}{n}\right) + (EX)^2 = \frac{1}{n^2}\cdot nDX + (EX)^2$$

X и Y нез ⇒
$D(X+Y) = DX + DY$

$$(*) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\underbrace{EX_i^2}_{=EX^2} - \frac{1}{n}DX - (EX)^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(DX - \frac{1}{n}DX\right) = \frac{1}{n-1}\cdot\frac{n-1}{n}\cdot nDX$$
$$= DX$$

# Метод максималне веродостојности

мотивација: изводимо експеримент са $Ber(p)$ расподелом

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}, \quad p \in [0,1]$$

Посматрајмо ПСУ $(X_1, \dots, X_n)$, $X_i \sim Ber(p)$

Нека је реализована вредност узорка: $(x_1, x_2, \dots, x_n)$, $x_i \in \{0,1\}$

Идеја је да оценимо $p$ тако да вероватноћа реализације датог узорка буде највећа.

Дакле, тражимо $p$ тако да је: $P\{X_1 = x_1, \dots, X_n = x_n\}$ највеће

Зато постављамо ф-ју од параметра $p$ коју зовемо ф-ја веродостојности и која гласи:

$$L(p) = P\{X_1 = x_1, \dots, X_n = x_n\}, \quad p \in [0,1]$$

Средимо мало ову једнакост:

$$\underbrace{L(p) = \prod_{i=1}^{n} P\{X_i = x_i\}}_{\text{независност } X_1, \dots, X_n}$$

Ми знамо да је $P\{X = x_i\} = \begin{cases} p, & x_i = 1 \\ 1-p, & x_i = 0 \end{cases}$, али згодније би нам било да то напишемо у једном реду. Предложено је да се то уради на следећи начин:

$$P\{X = x_i\} = p^{I\{x_i = 1\}} (1-p)^{I\{x_i = 0\}}$$

Како је увек испуњено следеће:

$$I\{x_i = 1\} + I\{x_i = 0\} = 1,$$

важи:

$$P\{X = x_i\} = p^{I\{x_i = 1\}} \cdot (1-p)^{1 - I\{x_i = 1\}}$$

Сада $L$ изгледа овако: $L(p) = \prod_{i=1}^{n} p^{I\{x_i = 1\}} (1-p)^{1 - I\{x_i = 1\}}$

$$L(p) = \prod_{i=1}^{n} p^{I\{x_i=1\}} (1-p)^{1-I\{x_i=1\}} = p^{\sum_{i=1}^{n} I\{x_i=1\}} \cdot (1-p)^{n-\sum_{i=1}^{n} I\{x_i=1\}}$$

Како је компликовано тражити max ове ф-је, често ћемо прелазити на ф-ју $l(p) := \ln L(p)$. $l$ и $L$ имају исту тачку максимума, јер је $\ln(\cdot)$ растућа ф-ја.

Подсећање особина логаритма: $\ln(a \cdot b) = \ln a + \ln b$

$\ln\left(\frac{a}{b}\right) = \ln a - \ln b$

$\ln a^b = b \ln a$

Запишимо сада $l(p)$:

$$l(p) = \sum_{i=1}^{n} I\{x_i=1\} \cdot \ln p + \left(n - \sum_{i=1}^{n} I\{x_i=1\}\right) \ln(1-p)$$

Како је $l$ диференцијабилна ф-ја на $(0,1)$, њен max знамо како можемо наћи:

$$l'(p) = \sum_{i=1}^{n} I\{x_i=1\} \cdot \frac{1}{p} + \left(n - \sum_{i=1}^{n} I\{x_i=1\}\right) \cdot \frac{(-1)}{1-p}$$

$$= \frac{1}{p} \cdot \sum_{i=1}^{n} I\{x_i=1\} - \frac{n}{1-p} + \sum_{i=1}^{n} I\{x_i=1\} \cdot \frac{1}{1-p}$$

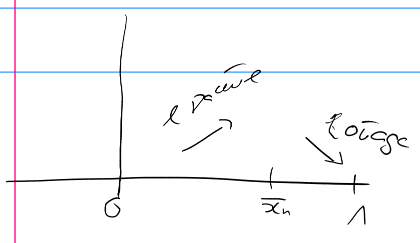
$p(1-p) \cdot /$ $\quad l'(p) = 0 \iff (1-p)\sum_{i=1}^{n} I\{x_i=1\} - np + p\sum_{i=1}^{n} I\{x_i=1\} = 0$

$$\iff p = \frac{\sum_{i=1}^{n} I\{x_i=1\}}{n}$$

Како је: $l'(p) > 0 \iff \dots \; p < \frac{\sum_{i=1}^{n} I\{x_i=1\}}{n}$ и $l'(p) < 0 \iff p > \frac{\sum_{i=1}^{n} I\{x_i=1\}}{n}$

следи да је $p = \frac{\sum_{i=1}^{n} I\{x_i=1\}}{n}$ заиста тачка максимума ф-је $l$, а самим тим и $L$.

У нашем случају важи следеће: $I\{x_i=1\} = x_i$ (јер $x_i \in \{0,1\}$), па је згодније написати да је $p = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}_n$



Дакле, вредност параметра $p$ за коју је

$P\{X_1=x_1, \dots, X_n=x_n\}$ највеће гласи

$$p = \overline{x}_n = \frac{\sum_{i=1}^{n} x_i}{n}$$

Како су оцене параметара статистике, преостаје нам да
ф-је од $X_1, \dots, X_n$
оцену запишемо као статистику:

$$\hat{p}_{mmv} = \overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

оцена добијена методом максималне веродостојности

Овде није крај!

Потребно је да се уверимо да је заиста $\hat{p}_{mmv}$ увек у $[0,1]$.

Ово јесте испуњено јер је:

$$X_{(1)} = \frac{n X_{(1)}}{n} \leq \frac{\sum_{i=1}^{n} X_i}{n} \leq \frac{n \cdot X_{(n)}}{n} = X_{(n)},$$

а статистике $X_{(1)}$ и $X_{(n)}$ узимају вредности из $\{0,1\}$, јер $X_i \in \{0,1\}$. Тј. $X_{(n)}$ је увек $\leq 1$ и $X_{(1)}$ је увек $\geq 0$

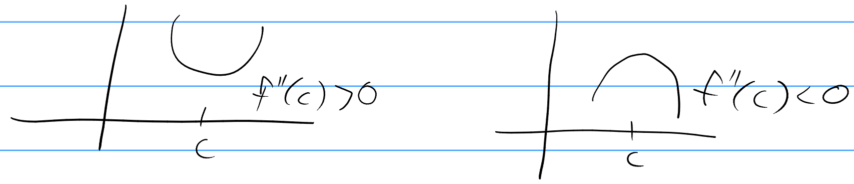
$X_i \sim \mathcal{P}(a),\ a>0$ $\qquad P\{X=x\} = e^{-a}\cdot \frac{a^x}{x!},\ x\in\mathbb{N}_0$

16. Позивамо се на тврђење:

Ако су $f$ и $f'$ на $(a,b)$ непр. и $f''$ постоји у $c\in(a,b)$ и $f'(c)=0$

Тада: $f''(c)>0 \Rightarrow c$ је локални минимум

$f''(c)<0 \Rightarrow c$ је локални максимум



Прим.: 2 пута диференцијабилне ф-је ($\exists f''$ у свакој тачки) задовољавају услов из тврђења

Ово тврђење смо искористили да покажемо да је $a=\overline{x}_n$ заиста тачка max ф-је $\ln L(a)$, $a>0$ јер је

$$\frac{\partial^2 \ln L(a)}{\partial a^2} = -\frac{1}{a^2}\sum_{i=1}^{n} x_i < 0 \quad \forall a>0,\ \text{па и за } a=\overline{x}_n$$

Други начин би био да покажемо да $f'(a)<0$ за $a>\overline{x}_n$ и $f'(a)>0$ за $a<\overline{x}_n$

Још је потребно проверити да је $\hat{a}=\overline{X}_n$ ненегативна. То важи ако бар једна од $X_1,\dots,X_n$ узме вредност из $\mathbb{N}$, док ако све $X_1,\dots,X_n$ узму вредност $0$, следи да је $\overline{x}_n=0$.

Међутим, параметар Пуасонове расподеле мора бити позитиван, па ова оцена нема смисла ако су реализоване вредности узорка: $(x_1,\dots,x_n)=(0,\dots,0)$

Но, ово је крајње редак случај, те у будуће нећемо обраћати превише пажње на ове детаље...

16. б) $X \sim G(p)$, $0 < p \leq 1$ $\quad P\{X = k\} = (1-p)^{k-1} p$, $k \in \mathbb{N}$

$$L(p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{\sum_{i=1}^{n} x_i - n} p^n$$

$$l(p) = \ln L(p) = \left(\sum_{i=1}^{n} x_i - n\right) \ln(1-p) + n \ln p$$

$$l'(p) = \frac{n - \sum_{i=1}^{n} x_i}{1-p} + \frac{n}{p} \Rightarrow l'(p) = 0 \iff pn - p\sum_{i=1}^{n} x_i + n - np = 0$$

$$\iff p = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$\left.\begin{array}{l} l'(p) > 0 \iff p < \dfrac{n}{\sum_{i=1}^{n} x_i} \\ l'(p) < 0 \iff p > \dfrac{n}{\sum_{i=1}^{n} x_i} \end{array}\right\} \Rightarrow p = \frac{n}{\sum_{i=1}^{n} x_i} \text{ је тачка max ф-је } l$$

$$\boxed{\hat{p}_{mmv} = \frac{1}{\overline{X}_n}}$$

Да ли је $\hat{p}_{mmv}$ увек у $(0, 1]$?

1° $\overline{X}_n$ је увек позитивно, па је и $\hat{p}_{mmv}$ увек позитивно

2° $\overline{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n} \geq \dfrac{n \cdot X_{(1)}}{n} = X_{(1)}$

$X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$ и сви $X_i$ су најмање 1, па и вредност статистике $X_{(1)}$ је најмање 1.

Дакле, вредност $\dfrac{1}{X_{(1)}}$ је највише 1

$$\Rightarrow \hat{p}_{mmv} = \frac{1}{\overline{X}_n} \leq \frac{1}{X_{(1)}} \leq 1$$

$X \sim Bin(5,p), \quad p \in [0,1]$

17.

$$L(p) = \prod_{i=1}^{n} P\{X_i = \underset{\in \{0,1,\dots,5\}}{x_i}\} = \prod_{i=1}^{n} \binom{5}{x_i} p^{x_i} (1-p)^{5-x_i}$$

$$= \prod_{i=1}^{n} \binom{5}{x_i} \cdot p^{\sum_{i=1}^{n} x_i} (1-p)^{5n - \sum_{i=1}^{n} x_i}$$

$$l(p) = \ln(L(p)) = \underbrace{\ln\left(\prod_{i=1}^{n} \binom{5}{x_i}\right)}_{\text{const}} + \sum_{i=1}^{n} x_i \cdot \ln p + \left(5n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

диференц. на $(0,1)$

const
не зависи од $p$!

$$l'(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{5n - \sum_{i=1}^{n} x_i}{1-p}$$

$$p(1-p) / \quad l'(p) = 0 \iff (1-p)\sum_{i=1}^{n} x_i - p \cdot 5n + p\sum_{i=1}^{n} x_i = 0$$

$$\iff \sum_{i=1}^{n} x_i = p \cdot 5n \iff p = \frac{\sum_{i=1}^{n} x_i}{5n} = \frac{\overline{x}_n}{5}$$

$l'(p) > 0$ за $p < \frac{\overline{x}_n}{5}$ и $l'(p) < 0$ за $p > \frac{\overline{x}_n}{5}$

Дакле, $p = \frac{\overline{x}_n}{5}$ је тачка max ф-је $l$.

$$\hat{p}_{\text{ммв}} = \frac{\overline{X}_n}{5}$$

Провера да је $\hat{p}_{\text{ммв}}$ увек из $[0,1]$..

На основу реализованог узорка

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $n_i$ | 6 | 10 | 14 | 13 | 6 | 1 |

број појављивања $x_i$ у узорку

$\Rightarrow n = 6+10+14+13+6+1 = 50$

оцена $\hat{p}_{\text{ммв}}$ је: $\frac{0\cdot 6 + 1\cdot 10 + \dots 5 \cdot 1}{50} \cdot \frac{1}{5}$

$$= \frac{10+28+39+24+5}{250}$$

$$= 0{,}424$$

# Метод максималне веродостојности у апс-непр. случају

Ако ПСУ $(X_1,\dots,X_n)$ долази из апсолутно непрекидне расподеле са густином $f$, како можемо применити метод макс. верод.?

Поново, нека је $(x_1,\dots,x_n)$ реализована вредност узорка.
Како сада нема смисла рачунати $P\{X_1=x_1,\dots,X_n=x_n\}$ јер је она једнака нули, послужићемо се познатим триком и рачунати:

$$(*)\quad P\{x_1<X_1<x_1+h,\dots,x_n<X_n<x_n+h\},\ \text{за } h \text{ мало}$$

Због независности $X_1,\dots,X_n$ важи:

$$(*) = P\{x_1<X_1<x_1+h\}\cdot\ldots\cdot P\{x_n<X_n<x_n+h\}$$

Како је $P\{x<X<x+h\} = \int_x^{x+h} f(t)\,dt \approx f(x)\cdot h$,

$$(*) \approx f(x_1)\cdot\ldots\cdot f(x_n)\cdot h^n$$

Сада је идеја да тражимо max ф-је:

$$L(p) = f(x_1)\cdot\ldots\cdot f(x_n)$$

↳ овако изгледа ф-ја веродостојности у апс-непр. случају

18. а) $X\sim\mathcal{E}(a)$, $a>0$

$$f(x)=\begin{cases} ae^{-ax}, & x>0\\ 0, & x\le 0\end{cases} = ae^{-ax}\, I\{x>0\}$$

$$L(a) = \prod_{i=1}^{n} ae^{-ax_i}\cdot \underbrace{I\{x_i>0\}} = a^n e^{-a\sum_{i=1}^{n} x_i}\prod_{i=1}^{n} I\{x_i>0\} = a^n e^{-a\sum_{i=1}^{n} x_i}\cdot I\{x_{(1)}>0\}$$

У овом примеру сви индикатори $I\{x_i>0\}$ ће бити једнаки 1, јер реализоване вредности из $\mathcal{E}(a)$ јесу позитивне, па их не морамо писати.
Међутим, у наредним примерима ће бити важно њихово наглашавање, па ћемо их оставити ради предострожности.

$$\ln L(a) = n\ln a - a\sum_{i=1}^{n} x_i + \ln I\{x_{(1)}>0\}$$

$$\ln L(a) = n \ln a - a \sum_{i=1}^{n} x_i + \ln \underbrace{I\{x_{(1)} > 0\}}_{\text{ово је } 1}$$

па је цео $\ln = 0$

Но, важно је само да је у питању const која не зависи од $a$, па кад диференцирамо $\ln L(a)$ по $a$, извод $\frac{\partial \ln I\{X_{(1)} > 0\}}{\partial a}$ ће бити $0$

$$\frac{\partial}{\partial a} \ln L(a) = \frac{n}{a} - \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial a} \ln L(a) = 0 \iff a = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}_n}$$

$$\frac{\partial}{\partial a} \ln L(a) > 0 \iff a < \frac{1}{\bar{x}_n},$$

$$\frac{\partial}{\partial a} \ln L(a) < 0 \iff a > \frac{1}{\bar{x}_n}$$

$\Rightarrow a = \frac{1}{\bar{x}_n}$ је тачка max ф-је $\ln L(a)$

$$\hat{a}_{\text{ммв}} = \frac{1}{\bar{X}_n}.$$

Заиста, $\hat{a}_{\text{ммв}}$ ће увек бити нека позитивна вредност па ова оцена има смисла. ($a$ је параметар $\mathcal{E}(a)$ расподеле, па је $a > 0$ по дефиницији)

19. $f(x) = \theta x^{\theta-1}, \quad x \in (0,1) \quad \sim \mathcal{B}(\theta, 1)$

$$f_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathcal{B}(\alpha,\beta)} \underset{\alpha=\theta,\beta=1}{=} \frac{x^{\theta-1}\cdot 1}{\mathcal{B}(\theta,1)} = \theta x^{\theta-1}$$

$$\mathcal{B}(\theta,1) = \frac{\Gamma(\theta)\Gamma(1)}{\Gamma(\theta+1)} = \frac{\cancel{\Gamma(\theta)}\cdot 1}{\theta\cancel{\Gamma(\theta)}} = \frac{1}{\theta}$$

20. а)

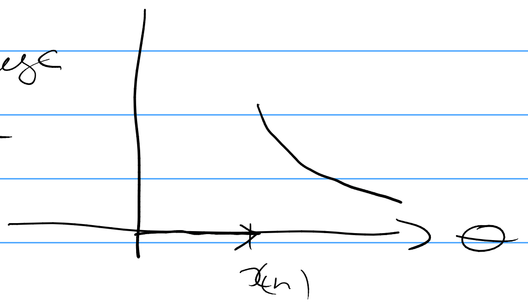
$X \sim U[0,\theta], \quad \theta > 0$

$$f(x) = \begin{cases} \frac{1}{\theta}, & x \in [0,\theta] \\ 0, & \text{иначе} \end{cases} = \frac{1}{\theta} I\{0 \le x \le \theta\}$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\{0 \le x_i \le \theta\} = \frac{1}{\theta^n} \prod_{i=1}^{n} I\{x_i \le \theta\}$$

$$= \frac{1}{\theta^n} I\{x_{(n)} \le \theta\} = \begin{cases} 0, & 0 < \theta < x_{(n)} \\ \frac{1}{\theta^n}, & x_{(n)} \le \theta \end{cases}$$

скице од $L$



дакле за $\theta = x_{(n)}$ се достиже max $L$

$\hat{\theta}_{\text{ммв}} = X_{(n)}$

б) $X \sim U[-\theta, \theta], \quad \theta > 0$

$$f(x) = \frac{1}{2\theta} I\{-\theta \le x \le \theta\}$$

$$L(\theta) = \frac{1}{(2\theta)^n} \prod_{i=1}^{n} I\{-\theta \le x_i \le \theta\} = \frac{1}{(2\theta)^n} I\{x_{(n)} \le \theta\} \cdot I\{-\theta \le x_{(1)}\}$$

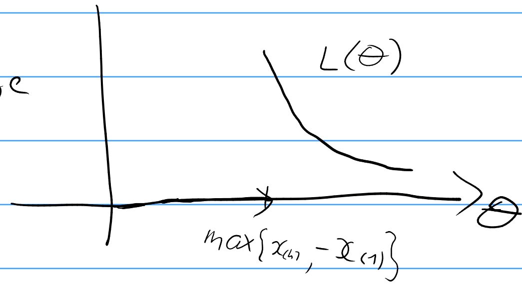
20. a) $X \sim U[-\theta, \theta]$, $\theta > 0$

$$L(\theta) = \frac{1}{(2\theta)^n} I\{x_{(n)} \le \theta\} \cdot I\{x_{(1)} \ge -\theta\}$$

$$= \frac{1}{(2\theta)^n} I\{\theta \ge x_{(n)}, \theta \ge -x_{(1)}\}$$

$$= \left(\frac{1}{2\theta}\right)^n \cdot I\{\theta \ge \max\{x_{(n)}, -x_{(1)}\}\}$$

опада




$\theta > 0$

да ли је $\max\{x_{(n)}, -x_{(1)}\} > 0$ ?

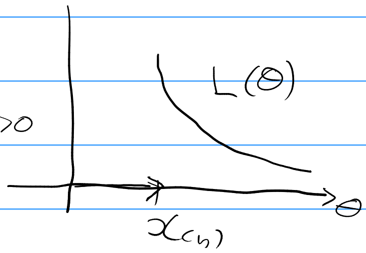
ако је $x_{(n)} > 0$ ✓

ако је $x_{(n)} < 0 \Rightarrow x_{(1)} < 0 \Rightarrow -x_{(1)} > 0$ ✓

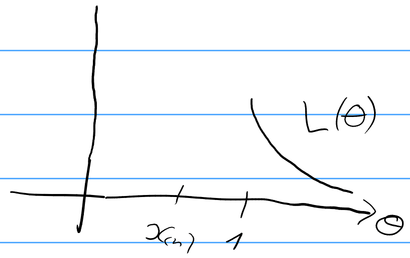
$$\hat{\theta}_{MMV} = \max\{X_{(n)}, -X_{(1)}\}$$

b) $X \sim U[0, \theta]$, $\theta \ge 1$

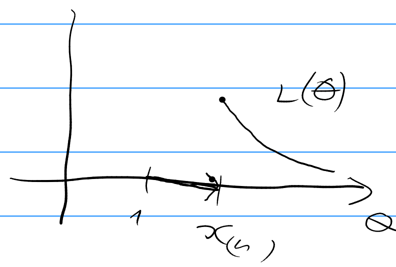
a) домен: $\theta > 0$




1°




2°

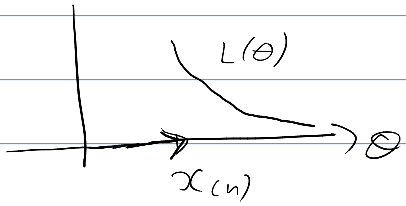



Дакле, $\hat{\theta}_{MMV} = \max\{1, X_{(n)}\}$

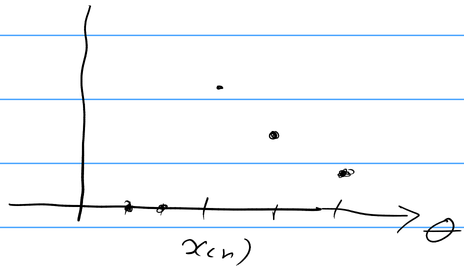
i) $X \sim U[0, \theta]$, $\theta \in \mathbb{N}$

$L(\theta)$ сада цртамо само у тачкама из $\mathbb{N}$

a)




1° $x_{(n)} \in \mathbb{N}$




$$\hat{\theta}_{MMV} = \begin{cases} X_{(n)}, & X_{(n)} \in \mathbb{N} \\ [X_{(n)}]+1, & X_{(n)} \notin \mathbb{N} \end{cases}$$

2° $x_{(n)} \notin \mathbb{N}$