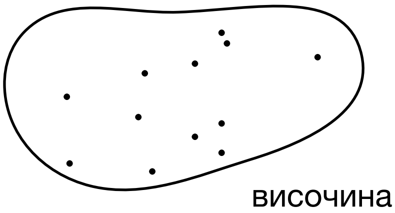
СЕМ, лекция 13

(2021-01-07)

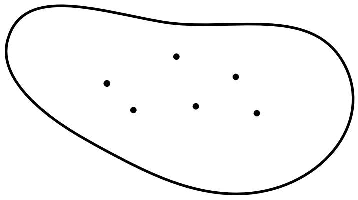
Статистика – точкови оценки

⊕ Всички българи

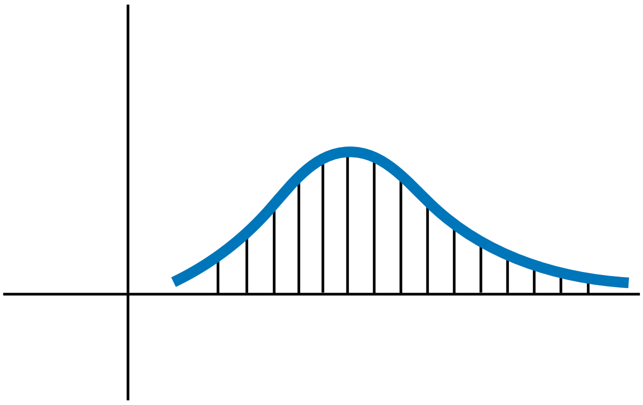


фиг. 1

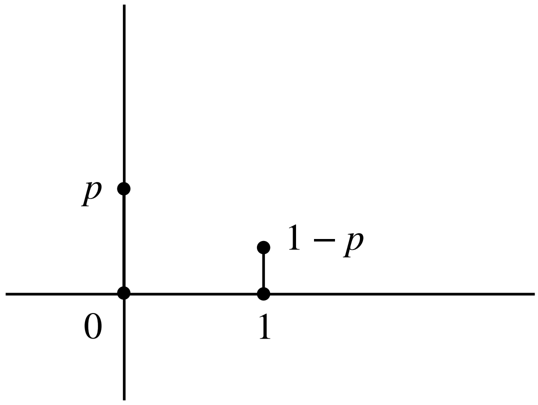
⊕ Гласоподаватели в САЩ



фиг. 2



фиг. 3



фиг. 4

Ако знаехме предварително височината на всеки от българите, щяхме да можем да си построим такава крива на разпределението, като от фиг. 1 и аналогично за гласоподавателите от фиг. 2.
Проблема е, че или е невъзможно или е твърде скъпо (по някой път обектите са цели функции или някакви сложни конфигурации). Целта е, наблюдавайки някаква подизвадка от обекти от цялата популация, да разберем нещо за синята крива на разпределението, за хистограмата, за средното или за пропорцията $\dfrac{p}{1-p}$ от фиг. 2 и т.н.

Най-ефективния начин за намирането на такава информация е като си направим една подизвадка от n човека, на които ще направим желаната характеристика и на база на тази информация ще се опитаме да извлечем нещо (дисперсия, средно и т.н.) за цялата популация.

Проблемът, който възниква е свързан с това да се направи избора напълно случайно. Т.е. всеки един обект от популацията да има равни шансове за избор спрямо останалите.

Ако работим само върху извадка от обекти с определен признак, то има риск да имаме пристрастие към нашите резултати и те няма да са показателни.

Постановка. Имаме случайна величина X с функция на разпределение F_X или плътност f_X. Искаме да разберем възможно най-много за някоя от характеристиките на X (μ, σ, F_X, f_X).

$\overrightarrow{X} = (X_1, \ldots, X_n)$, $(X_i)_{i=1}^n$ са независими случайни величини, които се реализрат чрез избор от генералната съвкупност (от всички изучавани обекти).

Цел. Някаква информация за X и най-вече F_X и f_X.

Допускания. Когато например изследваме гласоподавателите в САЩ (избира се между двама) ние имаме естествена рестрикция и знаем, че гласоподавател гласува за „партия 1“ или „партия 2“, което е конкретен клас разпределение (бернулиево – или за едната или за другата, ако е гласувал).

✤ X е някакъв клас разпределение, който се характеризира с някакъв вектор от параметри θ. Т.е. $F_X(x, \theta)$ и $f_X(x, \theta)$ (функцията на разпределение и плътност на X зависят от θ)

⊕ Ако допуснем, че $X \in \mathcal{N}(\mu, \sigma^2)$, където $\theta = (\mu, \sigma^2)$,
$f_X(x, \theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$; В някой ситуации може да допускаме и, че $\theta = \mu$ при $\sigma^2 = 1$ $\big(\mathcal{N}(\mu, 1)\big)$

⊕ (при гласоподавателите, както споменахме, нещата са още по прости) $X \in \mathrm{Ber}(p)$, т.е. $\theta = p = \mathbb{P}(X = 1)$. Трябва да приближим възможно най-добре този параметър p;

⊕ Класа на Гама разпределение: $\theta = (\alpha, \beta)$;

⊕ Класа на експоненциално разпределение: $\theta = (\lambda)$ и т.н.

Ще се опитваме да търсим информация при някакви допускания за разпределение на неизвестната случайна величина (било то гласоподавател, клетки, функции или каквото и да е)

Цел. $\overrightarrow{X} = \underbrace{(X_1, \ldots, X_m)}_{\text{априори}}$ и искаме да оценим θ на базата на априорните слчайни величини. Означаваме с $\hat{\theta} = \hat{\theta}(\overrightarrow{X})$, което е случайна величина.
Ако имаме конкретна реализация $X_i = x_i$, то $\hat{\theta} = \hat{\theta}(x_1 \ldots, x_n)$ е число.

Дефиниция. Оценката $\hat{\theta} = \hat{\theta}(\overrightarrow{X})$ на неизвестното θ се нарича ТОЧКОВА ОЦЕНКА на θ или статистика на θ.

Въпроса сега е: Как може да направим добри точкови оценки?

A. МЕТОД НА МАКСИМАЛНОТО ПРАВДОПОДОБИЕ (ММП)

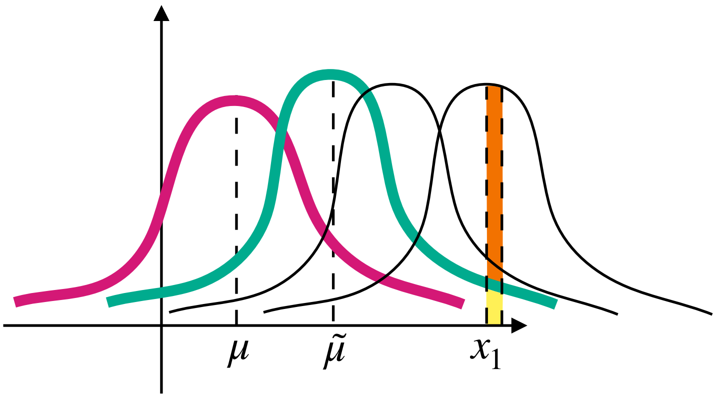
Този метод изхожда от следната схема:
Търсим информация за X с плътност $f_X(x, \theta)$ (допускаме, че я има), която зависи от някакви параметри θ. Наблюдаваме $\overrightarrow{X} = (X_1, \ldots, X_n)$ - n случайни обекта, за които сме извадили някакв бстойности, съответно $(x_1, \ldots, x_n)$. На базата на тези стойности трябва да конструираме по някакъв оптимизационен (смислен/облягащ се на някакви закономерности) начин - оценка за параметъра θ.

Първо ще погледнем какво е съвместното разпределение на $X_1, \ldots, X_n$ и това ще

$$\text{бъде } f_{\overrightarrow{X}}(x_1, \ldots, x_n, \theta) \overset{\text{независими, еднакво разпр.}}{=} \underbrace{\prod_{j=1}^{n} f_X(x_j, \theta)}_{\text{съвместното разпределение}} \xrightarrow{\text{ще се опитаме максимизираме по } \theta}$$

$\longrightarrow$ функция на максималното правдоподобие.

Пример. Нека например имаме $X \in \mathcal{N}(\mu, 1)$, $X_1 = x_1$; оценка: $\hat{\theta} = x_1$ $(\theta = \mu)$



$$\mathbb{P}\left(X_1 \in (x_1 - \varepsilon, x_1 + \varepsilon)\right) = \frac{1}{\sqrt{2\pi}} \int_{x_1-\varepsilon}^{x_1+\varepsilon} e^{-\frac{(x-\theta)^2}{2}} \, \mathrm{d}\, x \approx \frac{2\varepsilon}{\sqrt{2\pi}} e^{-\frac{-(x_1-\theta)^2}{2}} = 2\varepsilon f_X(x_1, \theta)$$

$\operatorname*{Sup}_{\theta} \mathbb{P}\left(X_1 \in (x_1 - \varepsilon, x_1 + \varepsilon)\right) = 2\varepsilon f_X(x_1, x_1)$. Т.е. взимаме там (околността) където плътността достига своя максимум.

Дефиниция. $\overrightarrow{X}$ е вектор от n независими, еднакво разпределени (копия) на X. Нека X има плътност $f_X(x, \theta)$ (допускаме, че се параметризира от някакъв параметър θ), където $\theta \in \Theta$ (допустимо множество). Тогава МПП (максимално правдоподобие/ максимално правдоподобно приближение) $\hat{\theta}$ на θ е този вектор/стойност, който/ която удовлетворява:

$$\mathrm{L}(X, \hat{\theta}) = \operatorname*{Sup}_{\theta \in \Theta} \underbrace{\mathrm{L}(\overrightarrow{X}, \theta)}_{\text{функция на макс. правдоподобие}} \overset{def.}{=} \operatorname*{Sup}_{\theta \in \Theta} \prod_{j=1}^{n} f_X(x_j, \theta)$$

$$\mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(\underbrace{x_j}_{\text{или цялото } X_j}, \theta) = f_{\overrightarrow{X}}(\overrightarrow{X}, \theta)$$

⊕ $\overrightarrow{X} = (x_1, \dots, x_n) = \mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(x_j, \theta)$

$$\operatorname*{Sup}_{\theta \in \Theta} \prod_{j=1}^{n} f_X(x_j, \theta) \qquad \frac{\partial \mathrm{L}}{\partial \theta} = 0 \longmapsto \frac{\partial \ln \mathrm{L}}{\partial \theta} = 0$$

⊕ Ако $\underbrace{X \in \mathcal{N}(\mu, \sigma^2)}_{X \text{ е от класа}}$ и $n \to$ наблюдения, то трябва да решим системата:

$\begin{cases} \frac{\partial \mathrm{L}}{\partial \mu} = 0 \\ \frac{\partial \mathrm{L}}{\partial \sigma^2} = 0 \end{cases} \Rightarrow (\mu, \sigma^2)$ - намираме тези μ и σ^2, които максимизират функцията на максималното правдоподобие, но е много по удобно да го правим за $\begin{cases} \frac{\partial \ln \mathrm{L}}{\partial \mu} = 0 \\ \frac{\partial \ln \mathrm{L}}{\partial \sigma^2} = 0 \end{cases}$, като $\ln$ също е нарастваща функция.

$\Theta = \mathbb{R} \times (0, \infty) \leftarrow$ максимизираме системата върху тази област.

Но не винаги ще може да имаме функция на правдоподобие, която да е диференцируема.

⊕ $X \in \mathcal{U}nif(0, \theta)$, $\overrightarrow{X} = (X-1, \dots, X_n) = (x_1, \dots, x_n)$.

Означаваме $X^* = \max\limits_{1 \le j \le n} x_j$ $\quad$ (ДС (допустими стойности): $\Theta = (0, \infty)$)

$$\mathrm{L}(\overrightarrow{X}, \theta) = \prod_{j=1}^{n} f_X(x_j, \theta) = \prod_{j=1}^{n} \frac{1}{\theta} 1_{[0,\theta]}(x_j) = \frac{1}{\theta^n} 1_{[0,\theta]}(x^*) = \begin{cases} 0, & \theta < x^* \\ \theta^{-n}, & x^* \le \theta \end{cases}, \text{ но}$$

последната функция не е диференцируема.

Но пък от друга страна, много лесно се максимизира:

$$\operatorname*{Sup}_{\theta>0} \mathrm{L}(\overrightarrow{X}, \theta) = \mathrm{L}(\overrightarrow{X}, x^*) = \frac{1}{(x^*)^n} \Rightarrow \hat{\theta} = \max(X_1, \ldots, X_n)$$

$\Big(\hat{\theta}$ се нарича точкова оценка по метода на максималното правдоподобие, ако $\mathrm{L}(\overrightarrow{X}, \hat{\theta}) = \operatorname*{Sup}_{\theta\in\Theta} \mathrm{L}(\overrightarrow{X}, \theta)\Big)$

Ще изведем оценките по максималното правдоподобие, в случая, в който $X \in \mathcal{N}(\mu, \sigma^2)$.

Твърдение. X е случайна величина от класа на нормално разпределените случайни величини. $\overrightarrow{X} = (X_1, \ldots, X_n)$ са независими еднакво разпределени наблюдения. Тогава $\hat{\mu} = \overline{X} = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} X_j$, МПО (макс. правд. оценка) на неизвестния параметър μ.

МПО за σ^2 е изразът (макс. правдоподобна оценка)

$a)$ $\quad \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (x_j - \mu)^2$, ако μ се допусне, че е известно;

б) $\quad \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} (x_j - \hat{x}_n)^2$, ако μ не е известно.

Разликата е чувствителна, тъй като в $a)$ използваме истинското μ, докато в б) използваме оценка.

Доказателство. $\mathrm{L}(\overrightarrow{X}, \theta) = \displaystyle\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}}$, $\theta = (\mu, \sigma^2)$

$$\overrightarrow{\mathrm{L}}(\overrightarrow{X}, \theta) = \ln \mathrm{L}(\overrightarrow{X}, \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \sum_{j=1}^{n} \frac{(x_j-\mu)^2}{2\sigma^2}$$

$$0 = \frac{\partial \tilde{\mathrm{L}}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu) \Rightarrow \hat{\mu} = \underbrace{\frac{1}{n}\sum_{j=1}^{n} X_j =: \overline{X}_n}_{\text{не зависи от } \sigma} \to \text{МПО}$$

$$0 = \frac{\partial \tilde{\mathrm{L}}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{j=1}^{n}(X_j - \mu)^2 \begin{cases} \xrightarrow{\ \mu\ } \hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}(X_j - \mu)^2, \ \mu \text{ - известно} \\[2ex] \xrightarrow{\ \hat{\mu}\ } \hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{j=1}^{n}(X_j - \overline{X}_n)^2, \ \mu \text{ - неизвестно} \end{cases}$$

Има две възможни оценки за $\hat{\sigma}^2$ (максимална правдоподобна оценка) според това дали знаем средното или не.

Б. МЕТОД НА МОМЕНТИТЕ

Тук вече няма нужда да допускаме съществуването на нищо друго освен съществуването на моментите.

$\oplus$ X е случайна величина с параметър $\theta \in \mathbb{R}$. Знаем, че средното $\mathbb{E}X = \mu(\theta)$.

За $\overrightarrow{X}$ от ЗГЧ имаме $\overline{X}_n^{(1)} = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\mathbb{P}(\text{п.с.})} \mathbb{E}X = \mu(\theta)$

Ако може да решим $\theta = \mu^{-1}\left(\overline{X}_n^{(1)}\right)$, където взимаме (като приближение) $\overline{X}_n = \mu(\theta)$, то θ ще е оценена по метода на моментите. Т.е. тук използваме ЗГЧ, за да оценим θ.

Дефиниция. Нека X е случайна величина, $F_X(x, \theta)$, $\theta = (\theta_1, \ldots, \theta_s)$. Нека имаме $\overrightarrow{X} = (X_1, \ldots, X_n)$ случайни величини, независими и разпределени като X (прототипи н случайна величина X).

Означаваме: $\overline{X}_n^{(k)} = \frac{1}{n}\sum_{j=1}^{n} X_j^k$ и $\mu^k(\theta) = \mathbb{E}X^k$. Тогава решението на системата

$\overline{X}^{(k)} = \mu^k(\theta)$, $1 \le k \le s$ за θ, се нарича оценка по метода на моментите.

$\oplus$ $X \in \mathcal{Unif}(0, \theta)$ и $\overrightarrow{X}$, $\overline{X}_n^{(1)} = \frac{1}{n}\sum_{j=1}^{n} X_j$

$$\mathbb{E}X = \frac{\theta}{2} \Rightarrow \overline{X}_n^{(1)} = \frac{\theta}{2} \Rightarrow \hat{\theta} = 2\overline{X}_n^{(1)} \qquad \text{ММО}$$

$$\theta = \max(X_1, \ldots, X_n) \qquad \text{ММП}$$

$\oplus$ $X \in \mathcal{N}(\mu, \sigma^2)$, $\overline{X}_n^{(1)} = \mu = \mathbb{E}X$, ММО=ММП

$$\frac{1}{n}\sum_{j=1}^{n} X_j^2 = \overline{X}_n^{(2)} = \mathbb{E}X^2 = \sigma^2 + \mu^2 \Rightarrow \sigma^2 = \overline{X}_n^{(1)} = \left(X_m^{(1)}\right)^2 \qquad \text{ММО=ММП}$$

Трябва да проверим, че $X_n^{(2)} - \left(X_n^{(1)}\right) = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}_n^{(1)}\right)^2$. Трябва да намерим начин, който да ни показва колко добра е всяка една от оценките.

Свойства на точковите оценки/статистики

а) Неизместеност.

Дефиниция. Казваме, че $\hat{\theta}$ е неизместена оценка на θ, ако $\mathbb{E}\hat{\theta}(\overrightarrow{X}) = \theta$. Когато $\theta = (\theta_1, \ldots, \theta_s)$, равенството се разбира като $\mathbb{E}\hat{\theta}_j(\overrightarrow{X}) = \theta_j$, $\forall\ 1 \le j \le s$.

Иначе, ако това не е изпълнено, тогава $\theta - \mathbb{E}\hat{\theta}(\overrightarrow{X})$ Се нарича систематична грешка.

$\oplus$ $\quad X \in \mathcal{N}(\mu, \sigma^2)$ $\hat{\mu} = X_m^{(1)}$, $\hat{\sigma}^2 = \overline{X}^{(2)} - \left(X_n^{(1)}\right)^2$, $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$

$$\mathbb{E}\hat{\mu} = \mathbb{E}\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}X_j = \frac{n\mu}{n} = \mu \Rightarrow \hat{\mu} \text{ е неизвестно. } \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2$$

Нека μ е известно, тогава оценката е $\displaystyle\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_j - \mu)^2 = \frac{1}{n}n\sigma^2 = \sigma^2$.

Ако μ е известно, то

$$\mathbb{E}\hat{\sigma}^2 = \mathbb{E}\frac{1}{n}\sum_{j=1}^{n}X_j^2 = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 = \mathbb{E}X_1^2 - \mathbb{E}\left(\frac{1}{n^2}\sum_{j=1}^{n}X_j^2 + \frac{1}{n^2}\sum_{i\neq j}X_iX_j\right) =$$

$$= \mathbb{E}X_1^2 - \frac{1}{n}\mathbb{E}X_1^2 + \frac{1}{n^2}\sum_{i\neq j}\underbrace{\mathbb{E}X_iX_j}_{\mu^2 \text{ произведение на две независими очаквания}} =$$

$$= \left(1 - \frac{1}{n}\right)\mathbb{E}X_1^2 + \frac{\mu^2}{n^{\cancel{2}}}\cancel{n}(n-1) =$$

$$= \left(1 - \frac{1}{n}\right)\left(\sigma^2 + \mu^2\right) - \frac{(n-1)\mu^2}{n} = \frac{n-1}{n}\sigma^2,$$

$$\hat{\sigma}^2 \neq \sigma^2.$$

$$\underbrace{S^2}_{\text{оценка}} = \frac{n}{n-1}\hat{\sigma}^2 \Rightarrow \mathbb{E}S^2 = \sigma^2, \text{ т.е. } \hat{\mu} = \overline{X}_n^{(1)}, S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2.$$

$\oplus$ $\quad X \in \mathcal{U}nif(0, \theta)$, $\theta = Y = \max\limits_{1\le j\le n}(X_j)$ е ММП

$\mathbb{E}Y = ?$

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}\left(\max_{j \le n}(X_j) \le y\right) = \mathbb{P}\left(\bigcap_{j=1}^{n}\{X_j \le y\}\right) = \prod_{j=1}^{n}\mathbb{P}(X_j \le y) = \frac{y^n}{\theta^n}$$

$$F_Y(y) = \begin{cases} \frac{y^n}{\theta^n}, & y \in [0, \theta] \\ \text{за останалите стойности не се интересуваме} & \end{cases} \Rightarrow \begin{cases} n\frac{y^{n-1}}{\theta^n}, & y \in (0, \theta) \\ 0, & \text{иначе} \end{cases}$$

$$\Rightarrow \mathbb{E}Y = \frac{n}{\theta^n}\int_0^\theta y^{n-1} y \, \mathrm{d}y = \frac{n}{n+1}\theta.$$

Оценката на максималното правдоподобие е изместена оценка на θ.

$\mathbb{E}\hat{\theta} = \dfrac{n}{n+1}\theta$; $\tilde{\theta} = \dfrac{n+1}{n}\hat{\theta}$. $\quad \hat{\theta} = 2\overline{X}_n^{(1)}$ е неизместена за θ.

б) Състоятелност на статистика (точкова оценка).

Дефиниция. Казваме, че $\hat{\theta}$ е състоятелна оценка на $\theta = (\theta_1, \ldots, \theta_s)$, ако $\lim\limits_{n\to\infty} \hat{\theta}_j(\overrightarrow{X}) \overset{\mathbb{P}}{=} \theta_j(\overline{\theta}_j \xrightarrow{\mathbb{P}} \theta_j)$, $1 \le j \le s$.

$\oplus$ $\quad X \in \mathcal{N}(\mu, \sigma^2)$, $\hat{\mu} = \dfrac{1}{n}\sum\limits_{j=1}^{n} X_j \xrightarrow[n\to\infty]{\mathbb{P}} \mu$ е състоятелна.

μ - известно; $\quad \hat{\sigma}^2 = \dfrac{1}{n}\sum\limits_{j=1}^{n}(X_j - \mu)^2 = \dfrac{1}{n}\sum\limits_{j=1}^{n} Y_j \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}Y_1 = \sigma^2$ е състоятелна.

$\hat{\mu}$ и $\hat{\sigma}^2$ Са състоятелни оценки, когато μ е известно ($\hat{\mu}$ по принцип винаги е състоятелна)

$$\overline{X}_n^{(k)} = \frac{1}{n}\sum_{j=1}^{n} X_j^k, \text{ то с полагането } Y_j = X_j^k \Rightarrow \overline{X}_n^{(k)} = \frac{1}{n}\sum_{j=1}^{n} \underbrace{Y_j}_{\substack{\text{еднакво разпр.} \\ \text{независими, т.к.} \\ X_j \text{ са такива и само} \\ \text{сме ги вдигнали} \\ \text{на степен } k}}.$$

Тоест $\overline{X}_n^{(k)}$ е състоятелна оценка за k-тия момент. Така, че тези оценки, които се намират по метода на моментите по принцип са състоятелни. Това е така, защото имаме ЗГЧ и той ни казва, че средното аритметично на наблюденията на степен k се схожда по вероятност (по траекторно) до k-тия момент на случайната величина, която изучаваме.

$\overline{X}_n^{(k)} \underbrace{\approx}_{\text{е мн. близо}} \mu^{(k)}(\theta) = \mathbb{E}X^k$, решаваме $\underbrace{\overline{X}_n^{(k)} = \mu^k(\theta)}_{\text{това в граница е тавтология}}$ и решаването му ни гарантира състоятелна оценка по принцип.