# Лекция IX - Статистика. Точкови оценки.

# Лекция IX - Статистика. Точкови оценки.

- Метод на максимално правдоподобие
- Метод на моментите
- Неизместеност
- Състоятелност

# Генерална съвкупност и извадка

Задачата, която се решава в статистиката обикновено е следната. Разполагаме с голяма съвкупност от еднородни обекти. Това може да са всички жители на някоя държава, или цялата продукция на някой завод или всички ракови клетки в един организъм и т.н. Всички обекти наричаме "генерална съвкупност". Изследването на цялата генерална съвкупност най-често е невъзможно или прекалено скъпо. Затова се избират част от обектите, тази част се нарича "извадка", провеждат се изследвания върху нея, и по събраната информация се съди за генералната съвкупност. За да бъде проучването смислено е много важен начинът, по който е направена извадката. Един добър метод е извадката да е чисто случайна, т.е. обектите да не бъдат сортирани по никакъв признак. Това не винаги е лесно за съобразяване.

# Генерална съвкупност и извадка

Задачата, която се решава в статистиката обикновено е следната. Разполагаме с голяма съвкупност от еднородни обекти. Това може да са всички жители на някоя държава, или цялата продукция на някой завод или всички ракови клетки в един организъм и т.н. Всички обекти наричаме "генерална съвкупност". Изследването на цялата генерална съвкупност най-често е невъзможно или прекалено скъпо. Затова се избират част от обектите, тази част се нарича "извадка", провеждат се изследвания върху нея, и по събраната информация се съди за генералната съвкупност. За да бъде проучването смислено е много важен начинът, по който е направена извадката. Един добър метод е извадката да е чисто случайна, т.е. обектите да не бъдат сортирани по никакъв признак. Това не винаги е лесно за съобразяване.

Известно със своя неуспех е проучването направено от американско списание за президентските избори през 1936г. Списанието събрало адреси от телефонните указатели и разпратило въпросници на над 4 000 000 души. Обработило върнатите отговори и обявило победителя на президентските избори, който всъщност ги загубил. Основната грешка на списанието бил начинът, по който е направена извадката. По онова време телефонът бил относително рядка вещ и само по-заможните семейства можели да си го позволят. Така допитването било извършено само сред тях, а вотът на по бедните не бил отчетен.

# Генерална съвкупност и извадка

На същите избори социологът Дж.Галъп правилно предрича резултата използвайки само 4000 анкети. Неговия подход отразява другия начин за подбор на извадката. Той е отчел, че обществото се разпада на социални групи, който се сравнително еднородни в отношението си към кандидатите за президент. Затова в една социална група анкетите може да са малобройни, но е важно извадката да повтаря социалната структура на генералната съвкупност.

# Генерална съвкупност и извадка

На същите избори социологът Дж.Галъп правилно предрича резултата използвайки само 4000 анкети. Неговия подход отразява другия начин за подбор на извадката. Той е отчел, че обществото се разпада на социални групи, който се сравнително еднородни в отношението си към кандидатите за президент. Затова в една социална група анкетите може да са малобройни, но е важно извадката да повтаря социалната структура на генералната съвкупност.

Ще дадем още един пример за да поясним същността на проблема. Ако целта ни е да определим средния ръст на жителите на една страна, разбира се измерването на всички е непосилна задача. Ако измерваме само мъжете, естествено ще допуснем грешка, тъй като мъжете са по-високи. Би трябвало да използваме извадка, в която пропорцията на мъжете и жените е такава, каквато е сред цялото население. Но полът не е единствения фактор, който влияе на ръста, друг такъв е възрастта, а също расата и т.н. Отчитането на всички фактори е сложна задача, този проблем от само себе си се решава при използването на чиста случайна извадка.

# Генерална съвкупност и извадка

На същите избори социологът Дж.Галъп правилно предрича резултата използвайки само 4000 анкети. Неговия подход отразява другия начин за подбор на извадката. Той е отчел, че обществото се разпада на социални групи, който се сравнително еднородни в отношението си към кандидатите за президент. Затова в една социална група анкетите може да са малобройни, но е важно извадката да повтаря социалната структура на генералната съвкупност.

Ще дадем още един пример за да поясним същността на проблема. Ако целта ни е да определим средния ръст на жителите на една страна, разбира се измерването на всички е непосилна задача. Ако измерваме само мъжете, естествено ще допуснем грешка, тъй като мъжете са по-високи. Би трябвало да използваме извадка, в която пропорцията на мъжете и жените е такава, каквато е сред цялото население. Но полът не е единствения фактор, който влияе на ръста, друг такъв е възрастта, а също расата и т.н. Отчитането на всички фактори е сложна задача, този проблем от само себе си се решава при използването на чиста случайна извадка.

Ще запишем формално задачата на математическата статистика. Смятаме че има неизвестна сл.в. $X$ която желаем да измерим. Предполагаме, че генералната съвкупност е достатъчно голяма и можем да направим $n$ на брой независими наблюдения над $X$, който ще означим с $\overrightarrow{\mathrm{X}} = (X_1, \dots, X_n)$. Самите наблюдения сл.в. $X_k$ са копия на $X$, т.е. имат същото разпределение, очакване и т.н.

# Точкова оценка

За начало ще предполагаме, че разпределението на сл.в. $X$ е известно, но то зависи от неизвестен за нас параметър $\theta$. Например, знаем че $X$ е нормално разпределена, но не знаем с какво очакване и дисперсия. Нека $F_X(x, \theta)$ е функцията на разпеделение на $X$. Ние търсим точната стойност на $\theta$, като неизвестния параметър $\theta$ може да бъде и вектор, т.е. да имаме за оценяване едновременно няколко константи.

Нека $\overrightarrow{\mathrm{X}} = X_1, \ldots, X_n$ са независими наблюдения над $X$. Построяваме функция $\hat{\theta} = \hat{\theta}(\overrightarrow{\mathrm{X}})$ и нейната стойност приемаме за стойност на неизвестния параметър $\theta$. Наричаме $\hat{\theta}$ точкова оценка (или статистика) за параметъра. Наблюденията $X_1, \ldots, X_n$ са случайни величини и за теоритични цели ги разглеждаме като такива, но ако извършим измерванията, т.е. в конткретен случай те са числа, и стойността на $\hat{\theta}$ е число.

# Точкова оценка

За начало ще предполагаме, че разпределението на сл.в. $X$ е известно, но то зависи от неизвестен за нас параметър $\theta$. Например, знаем че $X$ е нормално разпределена, но не знаем с какво очакване и дисперсия. Нека $F_X(x, \theta)$ е функцията на разпеделение на $X$. Ние търсим точната стойност на $\theta$, като неизвестния параметър $\theta$ може да бъде и вектор, т.е. да имаме за оценяване едновременно няколко константи.

Нека $\overrightarrow{\mathrm{X}} = X_1, \ldots, X_n$ са независими наблюдения над $X$. Построяваме функция $\hat{\theta} = \hat{\theta}(\overrightarrow{\mathrm{X}})$ и нейната стойност приемаме за стойност на неизвестния параметър $\theta$. Наричаме $\hat{\theta}$ точкова оценка (или статистика) за параметъра. Наблюденията $X_1, \ldots, X_n$ са случайни величини и за теоритични цели ги разглеждаме като такива, но ако извършим измерванията, т.е. в конткретен случай те са числа, и стойността на $\hat{\theta}$ е число.

Нашата задача е намирането на точковата оценка и ще опишем някой методи за целта, но преди това ще въведем следното означение.

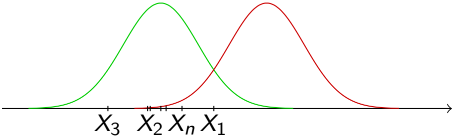
Нека $f_{X_k}(x_k, \theta)$ е плътността на $k$-тото наблюдение. Съвместната плътност на наблюденията наричаме функция на правдоподобие

$$L(\overrightarrow{\mathrm{X}}, \theta) = \prod_{k=1}^{n} f_{X_k}(x_k, \theta)$$

Тя отразява вероятността на наблюденията.

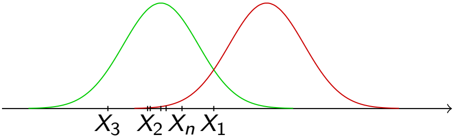
# Метод на максимално правдоподобие

Нека например се опитваме да оценим очакването $\mu$ на нормално разпределение, т.е. търсим къде е центрирана кривата. Естествено е да предпочетем $\mu$, при което наблюденията са с голяма вероятност (зелената крива), пред такава стойност, при която наблюденията са малко вероятни (червената крива).



# Метод на максимално правдоподобие

Нека например се опитваме да оценим очакването $\mu$ на нормално разпределение, т.е. търсим къде е центрирана кривата. Естествено е да предпочетем $\mu$, при което наблюденията са с голяма вероятност (зелената крива), пред такава стойност, при която наблюденията са малко вероятни (червената крива).




В това се състои идеята на метода на максимално правдоподобие. Неизвестният параметър се избира по такъв начин, че направените наблюдения да се окажат с възможно най-голяма вероятност. В някакъв смисъл напасваме теорията към действителността. Този метод е бил познат още на Гаус, а в началото на 20 век е доразвит от Фишер. Тъй като познаваме функцията на правдоподобие, методът се свежда до намиране на максимума на тази функция по неизвестния параметър.

## Дефиниция - Максимално правдоподобна оценка (м.п.о.)

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. $X - F(x, \theta)$.

Максимално правдоподобна оценка (м.п.о.) за неизвестен параметър $\boldsymbol{\theta}$ е тази стойност $\hat{\boldsymbol{\theta}}$, за която функцията на правдоподобие достига максимум.

# Метод на максимално правдоподобие

Намирането на максимума обикновено се извършва по стандартния начин с намиране и нулиране на производната, т.е. $\hat{\theta}$ е решение на уравнението

$$\frac{\partial L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

Намирането на максимума обикновено се извършва по стандартния начин с намиране и нулиране на производната, т.е. $\hat{\theta}$ е решение на уравнението

$$\frac{\partial L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

Много често при решаването на тази задача се работи с логаритъм. Идеята е, че логаритъмът е монотонна функция и затова $L$ и $\ln L$ достигат максимум в една и съща точка, а от друга страна тъй като $L$ е дефинирана като произведение от плътности, то $\ln L$ е сума и има по-прост вид, съответно намирането на максимума на $\ln L$ е по-лесна изчислителна задача. Затова спрямо $\theta$ се решава следното уравнение

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

# Метод на максимално правдоподобие

Намирането на максимума обикновено се извършва по стандартния начин с намиране и нулиране на производната, т.е. $\hat{\theta}$ е решение на уравнението

$$\frac{\partial L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

Много често при решаването на тази задача се работи с логаритъм. Идеята е, че логаритъмът е монотонна функция и затова $L$ и $\ln L$ достигат максимум в една и съща точка, а от друга страна тъй като $L$ е дефинирана като произведение от плътности, то $\ln L$ е сума и има по-прост вид, съответно намирането на максимума на $\ln L$ е по-лесна изчислителна задача. Затова спрямо $\theta$ се решава следното уравнение

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}}, \theta)}{\partial \theta} = 0$$

В дефиницията на метода на максимално правдоподобие параметъра $\theta$ може да бъде и многомерен, т.е. да се търси максимум едновременно по няколко променливи. Тогава се решава система от съответните частни производни.

### Пример

Нека сл.в. $X$ попада по случаен начин в интервала $[0, \theta]$, т.е. $X \in U(0, \theta)$. Ще намерим м.п.о. оценка за $\theta$, ако разполагаме с вектора на наблюденията.

# Метод на максимално правдоподобие

## Пример

Нека сл.в. $X$ попада по случаен начин в интервала $[0,\theta]$, т.е. $X \in U(0,\theta)$. Ще намерим м.п.о. оценка за $\theta$, ако разполагаме с вектора на наблюденията.

Да припомним, вероятността $X$ да е навсякъде в интервала е една и съща, което значи, че плътността е константа, в случая $f_X(x) = \frac{1}{\theta}$ за $x \in [0,\theta]$. Така за функцията на правдоподобие получаваме

$$L(\overrightarrow{\mathrm{X}},\theta) = \prod_{k=1}^{n} f_{X_k}(x_k) = \begin{cases} \dfrac{1}{\theta^n} & ,\forall x_k \in [0,\theta] \\ 0 & ,\exists x_k \notin [0,\theta] \end{cases}$$

# Метод на максимално правдоподобие

## Пример

Нека сл.в. $X$ попада по случаен начин в интервала $[0, \theta]$, т.е. $X \in U(0, \theta)$. Ще намерим м.п.о. оценка за $\theta$, ако разполагаме с вектора на наблюденията.

Да припомним, вероятността $X$ да е навсякъде в интервала е една и съща, което значи, че плътността е константа, в случая $f_X(x) = \frac{1}{\theta}$ за $x \in [0, \theta]$. Така за функцията на правдоподобие получаваме

$$L(\overrightarrow{\mathrm{X}}, \theta) = \prod_{k=1}^{n} f_{X_k}(x_k) = \begin{cases} \frac{1}{\theta^n} & , \forall x_k \in [0, \theta] \\ 0 & , \exists x_k \notin [0, \theta] \end{cases}$$

По метода на максимално правдоподобие, трябва да намерим стойността на $\theta$, за която функцията $L(\overrightarrow{\mathrm{X}}, \theta)$ достига максимум. Ясно е, че тази функция е намаляваща по $\theta$, тогава колкото по-малки стойности взима $\theta$ толкова по голяма е $L(\overrightarrow{\mathrm{X}}, \theta)$. От друга страна, ако $\theta$ стане по-малка от някое наблюдение $X_k$, то ще има наблюдение извън интервала $[0, \theta]$ и тогава $L(\overrightarrow{\mathrm{X}}, \theta) = 0$.

# Метод на максимално правдоподобие

## Пример

Нека сл.в. $X$ попада по случаен начин в интервала $[0,\theta]$, т.е. $X \in U(0,\theta)$. Ще намерим м.п.о. оценка за $\theta$, ако разполагаме с вектора на наблюденията.

Да припомним, вероятността $X$ да е навсякъде в интервала е една и съща, което значи, че плътността е константа, в случая $f_X(x) = \frac{1}{\theta}$ за $x \in [0,\theta]$. Така за функцията на правдоподобие получаваме

$$L(\overrightarrow{\mathrm{X}},\theta) = \prod_{k=1}^{n} f_{X_k}(x_k) = \begin{cases} \frac{1}{\theta^n} & , \forall x_k \in [0,\theta] \\ 0 & , \exists x_k \notin [0,\theta] \end{cases}$$

По метода на максимално правдоподобие, трябва да намерим стойността на $\theta$, за която функцията $L(\overrightarrow{\mathrm{X}},\theta)$ достига максимум. Ясно е, че тази функция е намаляваща по $\theta$, тогава колкото по-малки стойности взима $\theta$ толкова по голяма е $L(\overrightarrow{\mathrm{X}},\theta)$. От друга страна, ако $\theta$ стане по-малка от някое наблюдение $X_k$, то ще има наблюдение извън интервала $[0,\theta]$ и тогава $L(\overrightarrow{\mathrm{X}},\theta) = 0$.

Следователно максимум ще се достига при най-малкото $\theta$, по-голямо или равно на всички $X_k$, т.е оценката е

$$\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

Ще намерим оценка за всяка от двете константи при условие, че другата е известна или неизвестна, т.е. ще разгледаме четири случая. В началото ще изведем функцията на правдоподобие, която е обща за четирите случая

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над $X \in N(\mu, \sigma^2)$, тогава $f_{X_k}(x_k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}$ и за функцията на правдоподобие получаваме

$$L(\overrightarrow{\mathrm{X}}, \mu, \sigma) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_k-\mu)^2}{2\sigma^2}} = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\sum_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^2}}$$

Ще пресметнем логаритъм на функцията на правдоподобие,

$$\ln L(\overrightarrow{\mathrm{X}}, \mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2$$

а също и производните по всеки от двата параметъра

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}}, \mu, \sigma)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu)$$

$$\frac{\partial \ln L(\overrightarrow{\mathrm{X}}, \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{k=1}^{n} (x_k - \mu)^2$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е известна.

В този случай трябва да намерим максимум на функцията на правдоподобие $L$ по $\mu$, т.е. трябва да решим уравнението $\frac{\partial \ln L}{\partial \mu} = 0$ спрямо $\mu$

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu) = 0$$

$$\sum_{k=1}^{n} x_k - n\mu = 0$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е известна.

В този случай трябва да намерим максимум на функцията на правдоподобие $L$ по $\mu$, т.е. трябва да решим уравнението $\frac{\partial \ln L}{\partial \mu} = 0$ спрямо $\mu$

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu) = 0$$

$$\sum_{k=1}^{n} x_k - n\mu = 0$$

От тук получаваме оценката $\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k = \overline{X_n}$
( $\overline{X_n}$ е стандартно означение за средноаритметично).

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е известна.

В този случай трябва да намерим максимум на функцията на правдоподобие $L$ по $\mu$, т.е. трябва да решим уравнението $\frac{\partial \ln L}{\partial \mu} = 0$ спрямо $\mu$

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu) = 0$$

$$\sum_{k=1}^{n} x_k - n\mu = 0$$

От тук получаваме оценката $\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k = \overline{X_n}$ ( $\overline{X_n}$ е стандартно означение за средноаритметично).

- Оценка за дисперсията $\sigma^2$, ако очакването $\mu$ е известно.

Сега трябва да намерим максимум на $L$ по $\sigma$, т.е. $\frac{\partial \ln L}{\partial \sigma} = 0$

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{k=1}^{n} (x_k - \mu)^2 = 0$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е известна.

В този случай трябва да намерим максимум на функцията на правдоподобие $L$ по $\mu$, т.е. трябва да решим уравнението $\frac{\partial \ln L}{\partial \mu} = 0$ спрямо $\mu$

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k - \mu) = 0$$

$$\sum_{k=1}^{n} x_k - n\mu = 0$$

От тук получаваме оценката $\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{X_n}$
( $\overline{X_n}$ е стандартно означение за средноаритметично).

- Оценка за дисперсията $\sigma^2$, ако очакването $\mu$ е известно.

Сега трябва да намерим максимум на $L$ по $\sigma$, т.е. $\frac{\partial \ln L}{\partial \sigma} = 0$

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{k=1}^{n}(x_k - \mu)^2 = 0$$

Следователно търсената оценка е $\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \mu)^2$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е неизвестна.

И двата параметъра са неизвестни, което означава че трябва да намерим максимум на функцията на правдоподобие по две променливи, т.е. трябва да изразим $\mu$ от системата

$$\left| \begin{array}{l} \dfrac{\partial \ln L}{\partial \mu} = 0 \\[2ex] \dfrac{\partial \ln L}{\partial \sigma} = 0 \end{array} \right. \quad \Leftrightarrow \quad \left| \begin{array}{l} \dfrac{1}{\sigma^2} \displaystyle\sum_{k=1}^{n} (x_k - \mu) = 0 \\[2ex] -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3} \displaystyle\sum_{k=1}^{n} (x_k - \mu)^2 = 0 \end{array} \right.$$

Първото уравнение има решение идентично с това от първия случай, а второто не влияе на стойността на $\mu$. Така получената оценка съвпада с първи случай $\hat{\mu} = \overline{X_n}$, т.е. средното аритметично е оценка за очакването на нормално разпределена сл.в. независимо дали знаем или не дисперсията.

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

- Оценка за очакването $\mu$, ако дисперсията $\sigma^2$ е неизвестна.

И двата параметъра са неизвестни, което означава че трябва да намерим максимум на функцията на правдоподобие по две променливи, т.е. трябва да изразим $\mu$ от системата

$$\left| \begin{array}{l} \dfrac{\partial \ln L}{\partial \mu} = 0 \\[2ex] \dfrac{\partial \ln L}{\partial \sigma} = 0 \end{array} \right. \quad \Leftrightarrow \quad \left| \begin{array}{l} \dfrac{1}{\sigma^2} \displaystyle\sum_{k=1}^{n} (x_k - \mu) = 0 \\[2ex] -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3} \displaystyle\sum_{k=1}^{n} (x_k - \mu)^2 = 0 \end{array} \right.$$

Първото уравнение има решение идентично с това от първия случай, а второто не влияе на стойността на $\mu$. Така получената оценка съвпада с първи случай $\hat{\mu} = \overline{X_n}$, т.е. средното аритметично е оценка за очакването на нормално разпределена сл.в. независимо дали знаем или не дисперсията.

- Оценка за дисперсията $\sigma^2$, ако очакването $\mu$ е неизвестно.

Отново трябва да е в сила горната система, но в този случай изразяваме $\sigma^2$. От второто уравнение получаваме търсената оценка, като $\mu$ заместваме с решението на първо уравнение

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \mu)^2 = \frac{1}{n} \sum_{k=1}^{n} \left( x_k - \overline{X_n} \right)^2$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

Ще обобщим получените резултати.

## Твърдение

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. $N(\mu, \sigma^2)$. Тогава максимално правдоподобните оценки са:

- за математическото очакване

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} X_k = \overline{X_n}$$

- за дисперсията

при $\mu$ известно

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2$$

при $\mu$ неизвестно

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2$$

# Максимално правдоподобни оценки за $N(\mu, \sigma^2)$

Ще обобщим получените резултати.

## Твърдение

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. $N(\mu, \sigma^2)$. Тогава максимално правдоподобните оценки са:

- за математическото очакване

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} X_k = \overline{X_n}$$

- за дисперсията

при $\mu$ известно

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \mu)^2$$

при $\mu$ неизвестно

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} \left(X_k - \overline{X_n}\right)^2$$

Тези статистики се използват не само за точково оценяване на параметрите на нормално разпределение, но са отправна точка при построяването на доверителни интервали или за проверката на хипотези свързани с нормалното разпределение. Ние подробно ще изследваме техните свойства, но преди това разгледаме още един начин за получаване на точкови оценки.

# Метод на моментите

Нека сл.в. $X$ има функция на разпределение $F_X(x, \theta)$ която ни е известна като вид, но не знаем стойността на параметъра $\theta$. Ние можем да пресметнем математическото очакване на $X$, което също ще зависи от $\theta$, т.е. $\mathsf{E}X = \mu(\theta)$ е функция на $\theta$.

Нека $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ е векторът на наблюденията, тези сл.в. са независими, еднакво разпределени и следователно за тях е изпълнен законът за големите числа.

$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \underset{n\to\infty}{\overset{P}{\longrightarrow}} \quad \mathsf{E}X = \mu(\theta)$$

# Метод на моментите

Нека сл.в. $X$ има функция на разпределение $F_X(x, \theta)$ която ни е известна като вид, но не знаем стойността на параметъра $\theta$. Ние можем да пресметнем математическото очакване на $X$, което също ще зависи от $\theta$, т.е. $EX = \mu(\theta)$ е функция на $\theta$.

Нека $\overrightarrow{X} = (X_1, \ldots, X_n)$ е векторът на наблюденията, тези сл.в. са независими, еднакво разпределени и следователно за тях е изпълнен законът за големите числа.

$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \xrightarrow[n\to\infty]{P} \quad EX = \mu(\theta)$$

Тогава, съвсем естествено е да изберем тази стойност на $\theta$, за която ще имаме равенство, т.е. теоретичният момент $\mu(\theta)$ ще съвпадне със средното аритметично на наблюденията $\overline{X_n}$ (прието е $\overline{X_n}$ да се нарича емперичен момент, т.е. момент пресметнат от данните).

## Дефиниция - оценка по метода на моментите

Нека $\overrightarrow{X} = (X_1, \ldots, X_n)$ са независими наблюдения над сл.в. $X - F(x, \theta)$, и нека съществува математическото очакване $EX = \mu(\theta)$.

Оценка по метода на моментите (м.м.) за неизвестния параметър $\theta$ наричаме решението $\hat{\theta}$, на уравнението $\quad \mu(\theta) = \overline{X_n}$

По метода на моментите също могат да бъдат намирани оценки на повече от един параметър.

# Метод на моментите

Ако $\theta = (\theta_1, \theta_2, \dots, \theta_s)$, т.е. ако искаме да оценим едновременно $s$ параметъра, тогава ще работим с първите $s$ момента на сл.в. $X$ и $\hat{\theta}$ е решение на следната система

$$\left|\begin{array}{l} \mu^1(\theta) = \overline{X_n^1} \\ \mu^2(\theta) = \overline{X_n^2} \\ \dots \\ \mu^s(\theta) = \overline{X_n^s} \end{array}\right.$$

където $\mu^i(\theta) = \mathsf{E}X^i$ са теоритичните моменти, а $\overline{X_n^i} = \frac{1}{n}\sum_{k=1}^{n}(X_k)^i$ емперичните. За да работи методът предполагаме, че теоритичните моменти до ред $s$ съществуват, освен това системата трябва да има решение. Обикновено намирането на оценка по метода на моментите е по-лесно технически от откриването на м.п.о.

# Метод на моментите

Ако $\theta = (\theta_1, \theta_2, \dots, \theta_s)$, т.е. ако искаме да оценим едновременно $s$ параметъра, тогава ще работим с първите $s$ момента на сл.в. $X$ и $\hat{\theta}$ е решение на следната система

$$\left| \begin{array}{l} \mu^1(\theta) = \overline{X_n^1} \\ \mu^2(\theta) = \overline{X_n^2} \\ \dots \\ \mu^s(\theta) = \overline{X_n^s} \end{array} \right.$$

където $\mu^i(\theta) = \mathrm{E}X^i$ са теоритичните моменти, а $\overline{X_n^i} = \frac{1}{n}\sum_{k=1}^{n}(X_k)^i$ емпеличните. За да работи методът предполагаме, че теоритичните моменти до ред $s$ съществуват, освен това системата трябва да има решение. Обикновено намирането на оценка по метода на моментите е по-лесно технически от откриването на м.п.о.

## Пример

Отново ще намерим оценка за горната граница $\theta$ на равномерното разпределение, $X \in U(0, \theta)$.

# Метод на моментите

Ако $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$, т.е. ако искаме да оценим едновременно $s$ параметъра, тогава ще работим с първите $s$ момента на сл.в. $X$ и $\hat{\theta}$ е решение на следната система

$$\left| \begin{array}{l} \mu^1(\theta) = \overline{X_n^1} \\ \mu^2(\theta) = \overline{X_n^2} \\ \ldots \\ \mu^s(\theta) = \overline{X_n^s} \end{array} \right.$$

където $\mu^i(\theta) = \mathsf{E}X^i$ са теоритичните моменти, а $\overline{X_n^i} = \frac{1}{n}\sum_{k=1}^n (X_k)^i$ емперичните. За да работи методът предполагаме, че теоритичните моменти до ред $s$ съществуват, освен това системата трябва да има решение. Обикновено намирането на оценка по метода на моментите е по-лесно технически от откриването на м.п.о.

## Пример

Отново ще намерим оценка за горната граница $\theta$ на равномерното разпределение, $X \in U(0, \theta)$.Знаем че очакването е точно в средата на интервала, т.е. $\mathsf{E}X = \frac{\theta}{2}$ уравнението има вида

$$\frac{\theta}{2} = \overline{X_n}$$

Търсената оценка е $\hat{\theta} = 2\,\overline{X_n}$. Тази оценка явно е различна от м.п.о. която получихме преди (стр.7).

# Оценки за $N(\mu, \sigma^2)$ по метода на моментите

Нека $X_1, \ldots, X_n$ са независими наблюдения над $X \in N(\mu, \sigma^2)$. Ще приемем че и двата параметъра са неизвестни и ще намерим сатистика за тях по метода на моментите. Ще трябва да разгледаме системата.

$$\left| \begin{array}{l} \mathsf{E}X = \overline{X_n} \\ \mathsf{E}X^2 = \overline{X_n^2} \end{array} \right.$$

Знаем, че при нормално разпределение $\mathsf{E}X = \mu$, така че първото уравнение директно води до решение $\hat{\mu} = \overline{X_n}$.

Също така знаем, че $\mathsf{D}X = \sigma^2$, но освен това $\mathsf{D}X = \mathsf{E}X^2 - (\mathsf{E}X)^2$, следователно $\mathsf{E}X^2 = \mathsf{D}X + (\mathsf{E}X)^2$ и можем да запишем второто уравнение на система във вида $\mathsf{E}X^2 = \sigma^2 + \mu^2 = \overline{X_n^2}$. Решението спрямо $\sigma^2$ е

$$\hat{\sigma}^2 = \overline{X_n^2} - \mu^2 = \overline{X_n^2} - \left(\overline{X_n}\right)^2$$

Оценките, който получихме по метода на моментите напълно съвпадат с максимално правдоподобните оценки. За $\hat{\mu}$ това е очевидно, а за $\hat{\sigma}^2$ следва от следното представяне на м.п.о.

$$\hat{\sigma}^2(\text{м.п.о.}) = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k^2 - 2X_k\overline{X_n} + \left(\overline{X_n}\right)^2\right) =$$

$$= \frac{1}{n}\sum_{k=1}^{n}X_k^2 - 2\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)\overline{X_n} + \frac{1}{n}n\left(\overline{X_n}\right)^2 = \overline{X_n^2} - 2\,\overline{X_n}\,.\,\overline{X_n} + \left(\overline{X_n}\right)^2 = \hat{\sigma}^2(\text{м.м.})$$

# Неизместеност

Оценките получени по двата метода могат да са еднакви, но могат и да са различни. Както видяхме от примерите, ако имаме наблюдения попадащи по случаен начин в интервал $[0, \theta]$, разполагаме с два начина да оценим параметъра. Единия е да вземем $\theta$ равно на най-голямото наблюдение, другият да приемем $\theta$ равно на два пъти средно аритметично на наблюденията. Въпросът, който възниква е кой от двата начина е по-добър. Имаме нужда от критерии, с които да оценяваме точковите оценки. Ще въведем някой важни характеристики на точковите оценки.

# Неизместеност

Оценките получени по двата метода могат да са еднакви, но могат и да са различни. Както видяхме от примерите, ако имаме наблюдения попадащи по случаен начин в интервал $[0, \theta]$, разполагаме с два начина да оценим параметъра. Единия е да вземем $\theta$ равно на най-голямото наблюдение, другият да приемем $\theta$ равно на два пъти средно аритметично на наблюденията. Въпросът, който възниква е кой от двата начина е по-добър. Имаме нужда от критерии, с които да оценяваме точковите оценки. Ще въведем някой важни характеристики на точковите оценки.

## Неизместеност

Нека $\overrightarrow{\mathrm{X}}$ са независими наблюдения над $X - F(x, \theta)$. Казваме, че $\hat{\theta}$ е неизместена оценка за $\theta$, ако

$$\mathsf{E}\, \hat{\theta}(\overrightarrow{\mathrm{X}}) = \theta$$

Наблюденията $\overrightarrow{\mathrm{X}} = (X_1, \ldots, X_n)$ са случайни величини, оценката е функция от сл.в. и като такава има своето математическо очакване. Логично е да поискаме това очакване да съвпадне с оценявания параметър. Ако последното не е изпълнено, то при използването на оценката ще се натрупва грешка. Прието е разликата $\mathsf{E}\, \hat{\theta} - \theta$ да се нарича "систематична грешка".

# Неизместеност

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.11).

- Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathrm{E}\,\hat{\mu} = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}X_k = \frac{1}{n}\sum_{k=1}^{n} \mu = \mu$$

# Неизместеност

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.11).

- Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathrm{E}\,\hat{\mu} = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}X_k = \frac{1}{n}\sum_{k=1}^{n}\mu = \mu$$

- Оценката за дисперсията $\sigma^2$ при известно очакване $\mu$, е неизместена

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(X_k-\mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathrm{D}X_k = \sigma^2$$

# Неизместеност

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.11).

- Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathrm{E}\,\hat{\mu} = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}X_k = \frac{1}{n}\sum_{k=1}^{n}\mu = \mu$$

- Оценката за дисперсията $\sigma^2$ при известно очакване $\mu$, е неизместена

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(X_k-\mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathrm{D}X_k = \sigma^2$$

- Оценката за дисперсията $\hat{\sigma}^2$ при неизвестно очакване е изместена. Ще използваме вида на $\hat{\sigma}^2$ изведен на стр.14.

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\overline{X_n^2}\right) - \mathrm{E}\left(\overline{X_n}\right)^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}X_k^2\right) - \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)^2 =$$

# Неизместеност

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.11).

- Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathrm{E}\,\hat{\mu} = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}X_k = \frac{1}{n}\sum_{k=1}^{n}\mu = \mu$$

- Оценката за дисперсията $\sigma^2$ при известно очакване $\mu$, е неизместена

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(X_k-\mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathrm{D}X_k = \sigma^2$$

- Оценката за дисперсията $\hat{\sigma}^2$ при неизвестно очакване е изместена. Ще използваме вида на $\hat{\sigma}^2$ изведен на стр.14.

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\overline{X_n^2}\right) - \mathrm{E}\left(\overline{X_n}\right)^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}X_k^2\right) - \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)^2 =$$

Втората сума ще разделим на две части - сума от квадратите и сума от смесените произведения.

$$= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}X_k^2 - \frac{1}{n^2}\mathrm{E}\left(\sum_{k=1}^{n}X_k^2 + \sum_{i=1}^{n}\sum_{j\neq i}X_iX_j\right)$$

# Неизместеност

Ще проверим за неизместеност получените оценки на параметрите на нормалното разпределение (стр.11).

- Оценката за очакването $\hat{\mu}$ е неизместена, защото

$$\mathrm{E}\,\hat{\mu} = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}X_k = \frac{1}{n}\sum_{k=1}^{n} \mu = \mu$$

- Оценката за дисперсията $\sigma^2$ при известно очакване $\mu$, е неизместена

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(X_k-\mu)^2 = \frac{1}{n}\sum_{k=1}^{n}\mathrm{D}X_k = \sigma^2$$

- Оценката за дисперсията $\hat{\sigma}^2$ при неизвестно очакване е изместена. Ще използваме вида на $\hat{\sigma}^2$ изведен на стр.14.

$$\mathrm{E}\,\hat{\sigma}^2 = \mathrm{E}\left(\overline{X_n^2}\right) - \mathrm{E}\left(\overline{X_n}\right)^2 = \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2\right) - \mathrm{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2 =$$

Втората сума ще разделим на две части - сума от квадратите и сума от смесените произведения.

$$= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}X_k^2 - \frac{1}{n^2}\mathrm{E}\left(\sum_{k=1}^{n}X_k^2 + \sum_{i=1}^{n}\sum_{j\neq i}X_iX_j\right) = \frac{n-1}{n^2}\sum_{k=1}^{n}\mathrm{E}X_k^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\mathrm{E}(X_iX_j) =$$

# Неизместеност

От независимостта на наблюденията следва $\mathrm{E}(X_i X_j) = \mathrm{E}X_i \, \mathrm{E}X_j$

$$= \frac{n-1}{n^2} \sum_{k=1}^{n} \mathrm{E}X^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \mathrm{E}X_i \, \mathrm{E}X_j =$$

Наблюденията са копия на сл.в. $X$, следователно $\mathrm{E}X_i = \mathrm{E}X_j = \mathrm{E}X$. Така всички събираеми в двойната сума са еднакви, а броят на събираемите е $n(n-1)$.

$$= \frac{n-1}{n} \mathrm{E}X^2 - \frac{n-1}{n} (\mathrm{E}X)^2 = \frac{n-1}{n} \, \mathrm{D}X = \frac{n-1}{n} \, \sigma^2 \neq \sigma^2$$

Оказа се, че оценката за дисперсията при неизвестно очакване

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} \left( X_k - \overline{X_n} \right)^2$$

е изместена с множител $\frac{n-1}{n}$ .

# Неизместеност

От независимостта на наблюденията следва $\mathrm{E}(X_i X_j) = \mathrm{E}X_i \, \mathrm{E}X_j$

$$= \frac{n-1}{n^2} \sum_{k=1}^{n} \mathrm{E}X^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \mathrm{E}X_i \, \mathrm{E}X_j =$$

Наблюденията са копия на сл.в. $X$, следователно $\mathrm{E}X_i = \mathrm{E}X_j = \mathrm{E}X$. Така всички събираеми в двойната сума са еднакви, а броят на събираемите е $n(n-1)$.

$$= \frac{n-1}{n} \mathrm{E}X^2 - \frac{n-1}{n} (\mathrm{E}X)^2 = \frac{n-1}{n} \, \mathrm{D}X = \frac{n-1}{n} \, \sigma^2 \neq \sigma^2$$

Оказа се, че оценката за дисперсията при неизвестно очакване

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} \left( X_k - \overline{X_n} \right)^2$$

е изместена с множител $\frac{n-1}{n}$ .

Лесно можем да получим неизместена оценка за дисперсията, достатъчно е да умножим $\hat{\sigma}^2$ с реципрочния множител $\frac{n}{n-1}$. По този начин се получава известната оценка $S^2$.

$$S^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left( X_k - \overline{X_n} \right)^2$$

# Неизместеност

Дотук определяхме неизместеността като използвахме свойствата на математическото очакване, понякога при по-нетипични точкови оценки този подход е невъзможен.

## Пример

Ще проверим за неизместеност статистиката получена в първия пример $\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$, т.е. интересува ни дали $E\hat{\theta} = \theta$.

# Неизместеност

Дотук определяхме неизместеността като използвахме свойствата на математическото очакване, понякога при по-нетипични точкови оценки този подход е невъзможен.

## Пример

Ще проверим за неизместеност статистиката получена в първия пример $\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$, т.е. интересува ни дали $\mathrm{E}\hat{\theta} = \theta$. За целта първо ще намерим функцията на разпределение на $\hat{\theta}$. Нека $y \in [0, \theta]$

$$F_{\hat{\theta}}(y) = \mathrm{P}(\hat{\theta} < y) = \mathrm{P}(\max(X_1, X_2, \ldots, X_n) < y) =$$

$$= \mathrm{P}(X_1 < y, X_2 < y, \ldots, X_n < y) = \mathrm{P}(X_1 < y)\,\mathrm{P}(X_2 < y), \ldots, \mathrm{P}(X_n < y)$$

Наблюдението $X_k$ е копие на $X$, тогава $\forall k$: $\mathrm{P}(X_k < y) = \int_0^y \frac{1}{\theta} dx = \frac{y}{\theta}$. Така получаваме функцията на разпределение, а от нея и плътността

$$F_{\hat{\theta}}(y) = \left(\frac{y}{\theta}\right)^n, \qquad f_{\hat{\theta}}(y) = \frac{\partial F_{\hat{\theta}}(y)}{\partial y} = \frac{n\, y^{n-1}}{\theta^n}, \qquad 0 \le y \le \theta$$

Тогава очакването на $\hat{\theta}$ е следното

$$\mathrm{E}\hat{\theta} = \int_0^\theta y\, \frac{n\, y^{n-1}}{\theta^n} dy = \frac{n}{\theta^n}\, \frac{y^{n+1}}{n+1}\bigg|_{y=0}^{\theta} = \frac{n}{n+1}\,\theta \ne \theta$$

Оценката явно е изместена, но не е проблем да се превърне в неизместена, $\frac{n+1}{n}\max(X_1, X_2, \ldots, X_n)$ е такава.

# Състоятелност

Друго съвсем естествено изискване към точковите оценки е - когато броя на наблюденията расте оценките да стават по-точни, т.е. вероятността за грешка да намалява.

## Състоятелност

Нека $X_1, X_2, \dots X_n$ са независими наблюдения над $X$ с функция на разпределение $F(x, \theta)$. Казваме, че $\hat{\theta}$ е състоятелна оценка за $\theta$, ако

$$\hat{\theta} \xrightarrow[n\to\infty]{\mathsf{P}} \theta$$

Ще припомним, сходимостта по вероятност от дефиницията означава, че за всяко $\varepsilon$ е изпълнено

$$\mathsf{P}(|\hat{\theta} - \theta| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Съгласно закона за големи числа

$$\overline{X_n} \xrightarrow[n\to\infty]{P} \mathsf{E}X = \mu(\theta)$$

По метода на моментите, ние избираме $\hat{\theta}$ така, че горната сходимост се превръща в равенство. Затова оценките получени по метода на моментите винаги са състоятелни. Следователно и оценките за нормално разпределение, който ние изведохме са състоятелни.

# Състоятелност

Ще оставим любознателния читател сам да докаже състоятелността на статистиката $\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$ от първия пример.

# Състоятелност

Ще оставим любознателния читател сам да докаже състоятелността на статистиката $\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$ от първия пример.

17.5.2023 ЕК