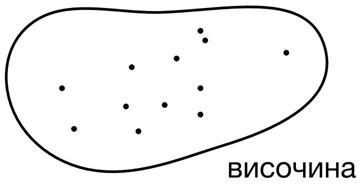
СЕМ, Л-13
(2021-01-07)

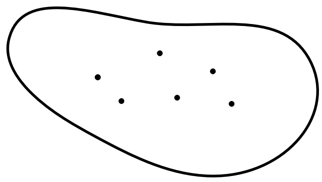
Статистика: точкови оценки

⊕ Всички българи

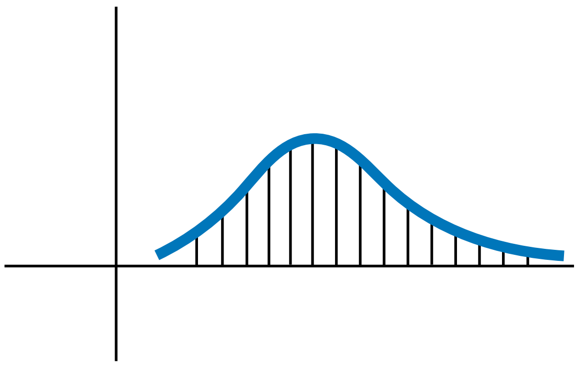
⊕ Гласоподаватели в САЩ



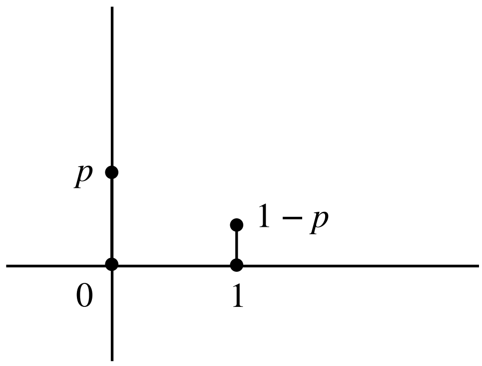
фиг. 1



фиг. 2



фиг. 3



фиг. 4

Основната идея на статистиката е, че разглеждаме извадка от дадена популация, за да направим изводи за цялата популация.

В илюстративния пример: ако знаехме всички индивидуални стойности (като височината на всеки българин), бихме могли директно да построим точната "синя" крива на разпределение (фиг. 1 и фиг. 3). Аналогично, ако знаехме политическите предпочитания на всеки гласоподавател (републикант или демократ), бихме могли директно да изчислим точната хистограма или пропорция (фиг. 2 и фиг. 4).

Проблемът в практиката е, че такова пълно изследване на популацията често е невъзможно, непрактично или непомерно скъпо. Особено когато обектите на изследване са сложни (например цели функционални зависимости или многокомпонентни конфигурации).

Следователно, целта на статистическия анализ е следната:
Наблюдавайки ограничена извадка от обекти, да получим надеждна информация за свойствата на цялата популация. Това включва:

- Формата на разпределението („синята крива“).
- Неговите числови характеристики (средна стойност, медиана, дисперсия).
- Пропорциите на различните категории (хистограмата).

Така от частични данни стигаме до обобщени знания с определена степен на достоверност.

Ключово условие за това е извадката да бъде представителна, тоест да се формира чрез напълно случаен избор. Това означава, че всеки обект в популацията трябва да има равен шанс да бъде включен в извадката. Ако работим само с обекти, притежаващи определен признак (например, анкетираме само хора в социални мрежи или измерваме височини само на баскетболисти), резултатите ще бъдат пристрастни (biased). Те тогава отразяват не цялата популация, а само тази специфична подгрупа, и следователно не са показателни за общите свойства, които се стремем да изследваме.

Едва когато извадката е случайна и представителна, направените от нея изводи за „синята крива“, средната стойност или пропорциите имат статистическа достоверност.

В статистическото моделиране често правим **параметрични допускания** за основното разпределение на популацията. Това значи, че приемаме, че неизвестната случайна величина X (която може да представлява височина, мнение, време на живот и т.н.) **принадлежи към предварително дефиниран клас разпределения**, описан от краен брой параметри θ. Формално, предполагаме, че $X \sim F_\theta$, където F_θ е функцията на разпределение, а f_θ — съответната плътност (или вероятностна маса).

Пример 1: Гласуване в САЩ.
В този случай резултатът е категоричен. Ако $X = 1$ означава глас за „Партия 1“, а $X = 0$ — за „Партия 2“, то X следва бернулиево разпределение с параметър p (вероятността за глас за Партия 1). Тук класът на разпределението е фиксиран (бернулиево), а целта е да оценим единствения параметър $p \in (0,1)$.

Пример 2: Непрекъснати величини.
За други изследвания може да се предполага, че X принадлежи към по-богати параметрични семейства:

- Нормално разпределение: $X \sim \mathcal{N}(\mu, \sigma^2)$, с параметри $\theta = (\mu, \sigma^2)$.
- Гама разпределение: $X \sim \text{Gamma}(\alpha, \beta)$, с параметри форма α и мащаб β.
- Експоненциално разпределение: $X \sim \text{Exp}(\lambda)$ — специален случай на Гама разпределение с $\alpha = 1$.

При тези допускания нашата задача се свежда до оценка на неизвестния параметър θ (или на вектор от параметри) въз основа на наблюдаваната извадка. По този начин, вместо да търсим произволна крива на разпределението (непараметричен подход), ние се фокусираме върху намиране на най-доброто приближение на θ в рамките на избрания клас, което опростява анализа и увеличава ефикасността, стига допускането да е правдоподобно.

Дефиниции.

- **Статистика:** Всяка функция на извадковите данни $T = g(X_1, X_2, \ldots, X_n)$, която не зависи от неизвестни параметри.

- **Оценител (estimator):** Правилото или формулата (статистиката), използвана за изчисляване на оценката. Той е *случайна величина* преди наблюдението на данните.
$$\hat{\theta} = T(X_1, \ldots, X_n)$$
- **Оценка (estimate):** Конкретната числена стойност на оценителя, получена за дадена реализация на извадката.
- **Точкова оценка:** Оценката на параметъра θ, представена като единична точка (число) от параметричното пространство.

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

Въпросите, които стоят пред нас сега са следните:

- Какви са желаните свойства (характеристики) на точковите оценители
- Какви са основните методи за изграждане на точкови оценки (point estimators) в статистиката и кои от тях се считат за най-ефективни (безпристрастни, с малка дисперсия)

Желани свойства (характеристики) на точковите оценители.

За да бъде оценителят $\hat{\theta}$ „добър“, той трябва да притежава определени свойства.

- **Несместеност (Безпристрастност):** Очакваната стойност на оценителя е равна на истинския параметър.
$$\mathbb{E}[\hat{\theta}] = \theta$$
Ако $\mathbb{E}[\hat{\theta}] \neq \theta$, оценителят е **сместен**, а величината $\mathbb{E}[\hat{\theta}] - \theta$ се нарича смещение (bias).

- **Ефективност:** Сред оценителите за даден параметър, предпочитаме точи с по-малка дисперсия. Ако за два несместени оценителя $\hat{\theta}_1$ и $\hat{\theta}_2$ е изпълнено $\mathrm{Var}[\hat{\theta}_1] < \mathrm{Var}[\hat{\theta}_2]$, то $\hat{\theta}_1$ е по-ефективен.

- **Състоятелност (Сходимост):** При безкрайно нарастване на обема на извадката ($n \to \infty$), оценителят **схожда по вероятност** към истинския параметър. Това означава, че с повече данни нашата оценка става все по-точна.
$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta, \text{ при } n \to \theta$$
Несместеността и намаляващата дисперсия до нула водят до състоятелност.

- **Адекватност (Достатъчност):** Оценителят използва *цялата информация* от извадката, свързана с параметъра θ. Формално, статистиката T е достатъчна за θ, ако условното разпределение на извадката при дадено T не зависи от θ.

Често използвани точкови оценители.

- **Извадкова средна:** Оценка за популационната среда μ.

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  Свойства: Несместен, състоятелен и ефективен при нормално разпределение.

- **Извадкова дисперсия:** Обикновено се използва коригираната извадкова дисперсия за оценка на популационната дисперсия σ^2.

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2$$

  Свойства: Несместена оценка за σ^2. (Обърнете внимание, че $\frac{1}{n}\Sigma(X_i - \overline{X})^2$ е сместена).

- **Извадкова пропорция:** Оценка за вероятността p при бернулиево разпределение.

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ където } X_i \in \{0,1\}$$

  Свойства: Несместен и състоятелен.

Изборът на точков оценител е критичен етап в статистическия анализ. Идеалният оценител е **несместен, ефективен, състоятелен и достатъчен**, но на практика често се търси баланс между тези свойства.

(A) Метод на максималното правдоподобие (ММП)

ММП е метод за намиране на най-вероятните стойности на параметрите на даден статистически модел, базиран на наблюдаваните данни. Той избира стойностите на параметрите, които максимизират вероятността (правдоподобието) да наблюдаваме точно тези данни.

Основна идея (Интуиция)

Представете си, че имате данни и модел, който ги описва. Моделът има някакви неизвестни параметри (напр. средна стойност μ и дисперсия σ^2 за нормално разпределение).

- **Правдоподобие (Likelihood)** не е вероятността параметрите да са верни. Това е вероятността **да наблюдаваме нашите конкретни данни**, при дадени стойности на параметрите.
- **Целта на ММП е:** "При какви стойности на параметрите нашите данни са били най-вероятни да се случат?"

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 $\overrightarrow{X} = (X_1, \dots, X_n)$ - n случайни обекта, за които сме извадили някакви стойности, съответно $(x_1, \dots, x_n)$. 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 $X_1, \dots, X_n$ и това ще бъде:

$$f_{\overrightarrow{X}}(x_1, \dots, x_n, \theta) \overset{\text{независими, еднакво разпр.}}{\underset{\text{i.i.d.}}{=}} \underbrace{\prod_{i=1}^{n} f_X(x_i, \theta)}_{\text{съвместното разпределение}} \xrightarrow{\text{ще се опитаме да максимизираме по } \theta}$$

$\longrightarrow$ функция на максималното правдоподобие

Как работи? Стъпка по стъпка

1. **Имаме:**
   - Данни: $X = (x_1, x_2, \dots, x_n)$ (независими наблюдения)
   - Предполагаем модел (разпределение) с неизвестни параметри θ (например, $\theta = (\mu, \sigma)$ за нормално разпределение).

2. **Построяваме функция на правдоподобието (Likelihood Function) $L(\theta | X)$:**
   - Това е съвместната вероятност (или плътност) да наблюдаваме всички наши данни, при паранметрите θ.
   - За независими наблюдения, това произведение на индивидуалните вероятности е:

$$L(\theta | X) = \mathbb{P}(x_1 | \theta) \cdot \mathbb{P}(x_2 | \theta) \cdot \ldots \cdot \mathbb{P}(x_n | \theta) = \prod_{i=1}^{n} \mathbb{P}(x_i | \theta)$$

   - За удобство често се работи с логаритмично правдоподобие (Log-Likelihood) $\ell(\theta | X)$:

$$\ell(\theta | X) = \ln\big(L(\theta | X)\big) = \sum_{i=1}^{n} \ln\big(\mathbb{P}(x_i | \theta)\big)$$

   Сумирането е по-лесно от умножението и е адитивно.

3. Максимизираме правдоподобието:
- Намираме стойностите на параметрите θ, които правят $L(\theta|X)$ (или $\ell(\theta|X)$ най-големи).
- Това става като вземем производните на $\ell(\theta|X)$ спрямо всеки параметър, приравним ги на нула и решим системата уравнения.
- Получените стойности са оценките по метода на максималното правдоподобие (ОММП): $\hat{\theta}_{MLE}$.

Прост пример: Оценка на вероятност за успех (p) при хвърляне на монета

Представете си, че хвърляте монета 10 пъти и получавате 7 „ези“ (успех) и 3 „тура“ (неуспех). Монетата вероятно не е честна. Каква е оценката за p (вероятност за успех)?

- Моде: Всяко хвърляне следва разпределение на Бернули с параметър p.
- Функция на правдоподобието: Вероятността да видим точно 7 успеха от 10 опити (биномно разпределение):

$$L(p : \text{данни}) = \binom{10}{7} \cdot p^7 \cdot (1-p)^3$$

- Логаритмично правдоподобие: (пропускаме константата $\binom{10}{7}$):

$$\ell(p : \text{данни}) = 7 \cdot \ln(p) + 3 \cdot \ln(1-p)$$

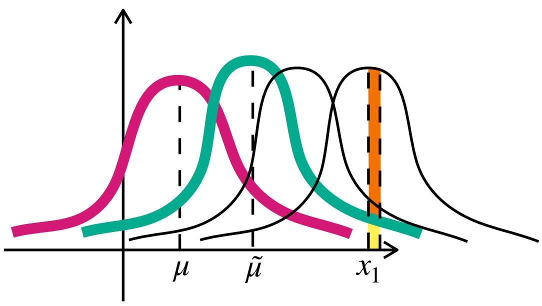
- Вземаме производна и приравняваме на нула:

$$\mathrm{d}\ell/\mathrm{d}p = 7/p - 3/(1-p) = 0$$
$$7/p = 3/(1-p) \Rightarrow p = 0.7$$

- Оценка по ММП: $\hat{p}_{MLE} = 7/10 = 0.7$

Заключение: Най-вероятната монета, която е произвела нашите данни, е такава с $p = 0.7$.

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



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

$$\operatorname*{Sup}_{\theta} \mathbb{P}\left(X_1 \in (x_1 - \varepsilon, x_1 + \varepsilon)\right) = 2\varepsilon f_X(x_1, x_1)$$

Тоест, 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:

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

$$L(\overrightarrow{X}, \theta) = \prod_{i=1}^{n} f_X(\underbrace{x_i}_{\text{или цялото } X_i}, \theta) = f_{\overrightarrow{X}}(\overrightarrow{X}, \theta)$$

$\oplus$ $\overrightarrow{X} = (x_1, \dots, x_n) = L(\overrightarrow{X}, \theta) = \prod_{i=1}^{n} f_X(x_i, \theta)$

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

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

$$\begin{cases} \dfrac{\partial L}{\partial \mu} = 0 \\ \dfrac{\partial L}{\partial \sigma^2} = 0 \end{cases} \Rightarrow (\mu, \sigma^2)$$

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

$\oplus\, X \sim \text{Uni}(0, \theta),\ \overrightarrow{X} = (X - 1, \ldots, X_n) = (x_1, \ldots, x_n).$

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

$$L(\overrightarrow{X}, \theta) = \prod_{i=1}^{n} f_X(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \cdot \mathbf{1}_{[0,\,\theta]}(x_i) = \frac{1}{\theta^n} \cdot \mathbf{1}_{[0,\,\theta]}(x^*) = \begin{cases} 0, & \theta < x^* \\ \theta^{-n}, & x^* \le \theta \end{cases},$$

но последната функция не е диференцируема.

Но пък от друга страна, много лесно се максимизира:

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

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

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 $\hat{\mu} = \overline{X} = \dfrac{1}{n}\Sigma_{i=1}^{n} X_i$, МПО (макс. правд. оценка) на неизвестния параметър μ.

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

а) $\sigma^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2$, ако μ се допусне, че е известно,

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

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

Доказателство:

$$L(\overrightarrow{X}, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}, \qquad \theta = (\mu, \sigma^2)$$

$$\overrightarrow{L}(\overrightarrow{X}, \theta) = \ln\big(L(\overrightarrow{X}, \theta)\big) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$0 = \frac{\partial \tilde{L}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \Rightarrow \hat{\mu} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n}_{\text{не зависи от } \sigma} \rightarrow \text{МПО}$$

$$0 = \frac{\partial \tilde{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2$$

$\xrightarrow{\mu}$ $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2$, μ - известно

$\xrightarrow{\hat{\mu}}$ $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$, μ - неизвестно

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.

Методът на моментите оценява неизвестните параметри на разпределение, като приравнява теоретичните моменти на разпределението (изразени като функции на параметрите) към съответните извадкови моменти, изчислени от данните. След това се решава получената система уравнения.

Основна идея (Интуиция)

Параметрите на едно разпределение определят неговите моменти (средна стойност, дисперсия, асиметрия и т.н.). Следователно, ако изчислим тези моменти от нашата извадка (данни), можем да „обърнем“ процеса и да намерим параметрите, които водят до моменти, близки до извадковите.

Как работи? Стъпка по стъпка

1. **Имаме:**
   - Данни: $X = (x_1, x_2, \ldots, x_n)$
   - Предполагаемо разпределение с k неизвестни параметри $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ напр. За нормално: $\theta_1 = \mu$, $\theta_2 = \sigma^2$).

2. **Дефинираме моментите:**

- Теоретичен (момент от популацията) момент от ред j: $\mu_j(\theta) = \mathbb{E}[X^j]$ (Очакване на X на степен j). Този момент зависи от неизвестните параметри θ.
- Извадков момент от ред j: $m_j = \dfrac{1}{n} \cdot \Sigma x_i^j$ (Средно аритметично на X на степен j). Този момент се изчислява директно от данните.

3. **Приравняване и решаване:**
   - Създаваме система от k уравнения, която приравняваме на първите k теоретизни моменти на първите k извадкови моменти:

$$\begin{cases} \mu_1(\theta) = m_1 \\ \mu_2(\theta) = m_2 \\ \dots \\ \mu_k(\theta) = m_k \end{cases}$$

   - Решаваме тази система спрямо θ. Получените решения са оценките по метода на моментите (ОММ): $\hat{\theta}_{MM}$.

Прост пример 1: Оценка на параметрите на нормално разпределение

Искаме да оценим μ и σ^2 за нормално разпределение $\mathcal{N}(\mu, \sigma^2)$.

- Броят на параметрите $k = 2$. Нужни са ни първите 2 момента.
- **Теоритични моменти:**
  1. $\mu_1 = \mathbb{E}[X] = \mu$
  2. $\mu_2 = \mathbb{E}[X^2] = \text{Var}[X] + (\mathbb{E}[X])^2 + \sigma^2 + \mu^2$
- **Извадкови моменти:**
  1. $m_1 = \dfrac{1}{n} \cdot \Sigma x_i$ (извадкова средна)
  2. $m_2 = \dfrac{1}{n} \cdot \Sigma x_i^2$ (извадков среден квадрат)
- **Приравняваме и решаваме:**
  Ур-ние 1: $\mu = m_1$
  Ур-ние 2: $\sigma^2 + \mu^2 = m_2$
  Решение: заместваме μ от ур-ние 1 в ур-ние 2: $\sigma^2 = m_2 - m_1^2$. Но $m_2 - m_1^2 = \dfrac{1}{n} \cdot \Sigma x_i^2 - \left[\dfrac{1}{n} \cdot \Sigma x_i^2\right]$, което е извадковата дисперсия (с максимума на правдоподобието, некоригирана).
- **Оценка по ММ:**
  $\hat{\mu}_{MM} = \dfrac{1}{n} \cdot \Sigma x_i = \tilde{x}$ (извадкова средна)
  $\tilde{\sigma}^2_{MM} = \dfrac{1}{n} \cdot \Sigma (x_i - \tilde{x})^2$ (извадкова дисперсия)

Прост пример 2: Оценка на параметрите на равномерно разпределение $\mathrm{Uni}(a, b)$.

Искаме да оценим a и b от данни, разпределени равномерно в интервала $[a, b]$.

- Броят на параметрите $k = 2$.
- **Теоритични моменти за $\mathrm{Uni}(a, b)$:**
  1. $\mu_1 = \mathbb{E}[X] = (a + b)/2$
  2. $\mu_2 = \mathbb{E}[X^2] = \mathrm{Var}[X] + (\mathbb{E}[X])^2 = (b - a)^2/12 + ((a + b)/2)^2$
- **Извадкови моменти:** m_1 и m_2 (както в предния пример).
- **Приравняваме и решаваме:**
  От ур-ние 1: $(a + b)/2 = m_1 \Rightarrow a + b = 2m_1$
  От ур-ние 2, след заместване и алгебрични преобразувания, можем да получим, че: $(b - a)^2 = 12(m_2 - m_1^2)$
  Решаваме системата:
  $$\begin{cases} a + b = 2m_1 \\ b - a = \sqrt{12(m_2 - m_1^2)} \end{cases}$$
  Решението е:
  $$\begin{cases} \hat{a}_{MM} = m_1 - \sqrt{3(m_2 - m_1^2)} \\ \hat{b}_{MM} = m_1 + \sqrt{3(m_2 - m_1^2)} \end{cases}$$
  Забележете, че $\sqrt{m_2 - m_1^2}$ е озвадково стандартно отклонение (некоригирано). Така оценките се изразяват чрез извадковата средна и стандартното отклонение.

$\oplus$ X е случайна величина с параметър $\theta \in \mathbb{R}$. Знаем, че средното $\mathbb{E}[X] = \mu(\theta)$.
За $\overrightarrow{X}$ от ЗГЧ имаме $\overline{X}_n^{(1)} = \dfrac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{п.с.}} \mathbb{E}[X] = \mu(\theta)$

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. Тоест 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 на случайна величина X).

Означаваме: $\overline{X}_n^{(k)} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$ и $\mu^k(\theta) = \mathbb{E}[X^k]$. Тогава решението на системата

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

$\oplus\ X \sim \mathrm{Uni}(0, \theta)$ и $\overrightarrow{X}$, $\overline{X}_n^{(1)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

$$\begin{cases} \mathbb{E}[X] = \frac{\theta}{2} \Rightarrow \overline{X}_n^{(1)} = \frac{\theta}{2} \Rightarrow \hat{\theta} = 2\overline{X}_n^{(1)} & \text{ММО} \\ \theta = \max(X_1, \ldots, X_n) & \text{ММП} \end{cases}$$

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

$\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 = \overline{X}_n^{(2)} = \mathbb{E}[X^2] = \sigma^2 + \mu^2 \Rightarrow \sigma^2 = \overline{X}_n^{(1)} = \left(X_m^{(1)}\right)^2$, ММО=ММП

Трябва да проверим, че $X_n^{(2)} - \left(X_n^{(1)}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n^{(1)}\right)^2$. 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 $\mathbb{E}[\hat{\theta}_j(\overrightarrow{X})] = \theta_j,\ \forall\ 1 \leq j \leq s$.

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

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

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{n\mu}{n} = \mu \Rightarrow \hat{\mu} \text{ е неизвестно.}$$

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

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

Ако μ е известно, то

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

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

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

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

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

$\oplus$ $X \in \mathrm{Uni}(0, \theta)$, $\theta = Y = \max\limits_{1\leq i\leq n}(X_i)$ е ММП. $\mathbb{E}[Y] = ?$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(\max_{i\leq n}(X_i) \leq y\right) = \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \leq y\}\right) = \prod_{i=1}^{n}\mathbb{P}(X_i \leq y) = \frac{y^n}{\theta^n}$$

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

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

Оценката на максималното правдоподобие е изместена оценка на θ.

$$\mathbb{E}[\hat{\theta}] = \frac{n}{n+1}\theta,\ \tilde{\theta} = \frac{n+1}{n}\hat{\theta}.\ \hat{\theta} = 2\overline{X}_n^{(1)} \text{ е неизместена за } \theta.$$

б) Състоятелност на статистика (точкова оценка).

Дефиниция. Казваме, че $\hat{\theta}$ е състоятелна оценка на $\theta = (\theta_1, \ldots, \theta_s)$, ако

$$\lim_{n\to\infty} \hat{\theta}_i(\overrightarrow{X}) \stackrel{\mathbb{P}}{=} \theta_i(\overline{\theta}_i \xrightarrow{\mathbb{P}} \theta_i),\ 1 \leq i \leq s.$$

$\oplus\ X \sim \mathcal{N}(\mu, \sigma^2)$, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mu$ е състоятелна.

μ - известно; $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[Y_1] = \sigma^2$ е състоятелна.

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

$\overline{X}_n^{(k)} = \frac{1}{n}\sum_{j=1}^{n} X_j^k$, 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 $\underbrace{\overline{X}_n^{(k)} = \mu^k(\theta)}_{\substack{\text{това в граница} \\ \text{е тавтология}}}$ и решаването му ни

гарантира състоятелна оценка по принцип.

Сравнителна таблица: ММП vs ММ (ММО)

Критерии / Характеристика	Метод на максималното правдоподобие (ММП)	Метод на моментите (ММ)
Основен принцип	Максимизира вероятността (правдоподобието) за наблюдение на данните: $L(\theta \mid X) \to \max$	Приравнява теоретични и извадкови моменти: $m_j(\theta) = m_j$
Философия	„Кои параметри правят нашите данни най-вероятни?“	„Кои параметри водят до моменти, равни на наблюдаваните?“
Ефикасност	Асимптотично ефикасен (достига долна граница на Рао-Крамер - минимална възможна дисперсия при големи n)	Неефикасен (по-голяма дисперсия от ММП при големи извадки)
Състоятелност	Състоятелен (при стандартни условия)	Състоятелен (обикновено)
Смесеност (пристрастие)	Може да е пристрастен за малки извадки (пример: σ^2 в нормално), но често може да се коригира	Често пристрастен, но не винаги (зависи от случая)
Инвариантност	Да: Ако $\hat{\theta}$ е ОММП (Оценка по Метода на Максималното Правдоподобие) за θ, то $g(\hat{\theta})$ е ОММП за $g(\theta)$	Не: Ако $\hat{\theta}$ е ОММ за θ, то $g(\hat{\theta})$ НЕ е задължително ОММ за $g(\theta)$
Сила на използване на информацията	Използва цялата информация за формата на разпределението (чрез функцията на плътност)	Използва само информация за моментите (обикновено първите няколко)
Сложност на изчисление	По-сложно: често изисква числени методи за оптимизация	По-просто: обикновено се свежда до решаване на система уравнения
Чувствителност към outliers	По-чувствителен (стреми се да ги обясни)	По-малко чувствителен (ако не се включат високи моменти)
Ситуации с проблеми	Може да има проблеми при многомодални функции или сингулярности	Може да дава параметри извън допустимия диапазон (напр. Отрицателна дисперсия)
Устойчивост	По-малко устойчив	По-устойчив в някои случаи
Избор пти многобройни решения	Дава конкретно решение (максимум на правдоподобието)	Може да има няколко набора от моменти за приравняване, водещи до различни оценки
Асимптотично разпределение	Нормално с минимална дисперсия	Нормално, но с по-голяма дисперсия