Основы статистики

Александр Сахнов
linkedin.com/in/amsakhnov

Staff MLE at Alibaba Group

2 сентября 2021 г.

Оглавление

Точечная оценка

Definition

Будем называть **выборкой** набор случайных величин.

В теории вероятностей и статистике отдельно выделяют случай **независимых одинаково распределенных случайных величин** (н.о.р.с.в.). Каждая из них распределена так же как и остальные и все они независимы в совокупности. Это принято обозначать как $X_1, \ldots, X_n \sim F$.

Definition

Статистика — любая измеримая функция от выборки.

Definition

Точечная оценка параметра распределения — статистика, которую мы могли бы рассматривать как предполагаемое значение оцениваемого параметра.

Несмещённая оценка

$\hat{\theta}$ или $\hat{\theta}_n$ — оценка параметра θ.
$\hat{\theta}_n = g(X_1, \ldots, X_n)$ — случайная величина, т.к. зависит от данных.

Definition

Оценка $\hat{\theta}_n$ **несмещённая**, если $\mathbb{E}(\hat{\theta}_n) = \theta$.
$bias(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$ — смещение оценки.

Example

$X_1, \ldots, X_n \sim F$, покажем, что $\overline{X}$ является несмещённой оценкой $\mathbb{E}X = m$.

$$\mathbb{E}\hat{m} = \mathbb{E}\left(n^{-1}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i = \frac{1}{n}\sum_{i=1}^{n} m = \frac{nm}{n} = m$$

Состоятельная оценка

Сходимость по вероятности: $\forall \epsilon > 0$ выполняется $\mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) \overset{n \to \infty}{\to} 0$.

Definition

Оценка $\hat{\theta}_n$ **состоятельная**, если $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.

Example

Дана выборка независимых одинаково распределённых случайных величин $X_1, \dots, X_n \sim F$.

Покажем, что оценка математического ожидания $\hat{\theta} = \frac{1}{n} \sum_i X_i$ является состоятельной.

$$\mathbb{V}(\hat{\theta}) = \mathbb{V}\left(\frac{1}{n}\sum_i X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_i X_i\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Неравенство Чебышёва $\mathbb{P}(|\hat{\theta} - \theta| > \epsilon) \leq \frac{\mathbb{V}(\hat{\theta})}{\epsilon^2}$. Получаем

$$\mathbb{P}(|\hat{\theta} - \theta| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0, \;\; n \to \infty$$

Критерий состоятельности

Theorem

Если bias $\to 0$ и se $\to 0$ при $n \to \infty$, то оценка $\hat{\theta}_n$ состоятельная.

Example

Дана выборка из распределения Бернулли $X_1, \ldots, X_n \sim Bernoulli(p)$.

Оценка параметра распределения $\hat{p}_n = \dfrac{1}{n} \sum\limits_i X_i$.

$\mathbb{E}(\hat{p}_n) = p$, тогда при $n \to \infty$

$$bias = p - p = 0, \quad se = \sqrt{p(1-p)/n} \to 0.$$

Оценка $\hat{p}_n$ состоятельная.

Свойства оценок

Definition

Оценка является **асимптотически нормальной**, если

$$\frac{\hat{\theta}_n - \theta}{se} \rightsquigarrow \mathcal{N}(0, 1).$$

Definition

τ_θ — семейство несмещенных оценок для θ.
Оценка $\theta^*_{opt} \in \tau_\theta$ называется **оптимальной** оценкой для θ, если

$$\forall \theta^* \in \tau_\theta \quad \mathbb{V}\theta^*_{opt} \leq \mathbb{V}\theta^*$$

Оценка максимального правдоподобия

Дана выборка из параметрического распределения $X_1, \ldots, X_n \sim F_\theta, \quad \theta \in \Theta$.
Функция правдоподобия $L(\theta) = \prod_{i=1}^n p(x_i; \theta)$.
Оценка максимального правдоподобия параметра θ: $\hat{\theta} = \arg\max_\theta L(\theta)$

Пример ОМП. Нормальное распределение

Example

Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Оценить параметры распределения μ, σ.

Функции правдоподобия (без домножения на константу)

$$L(\mu, \sigma) = \prod_i \frac{1}{\sigma} \exp\left\{-\frac{(X_i - \mu)^2}{2\sigma^2}\right\} = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (X_i - \mu)^2\right\}$$

$$= \sigma^{-n} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{X} - \mu)^2}{2\sigma^2}\right\},$$

$$\text{где} \quad \overline{X} = \frac{1}{n}\sum_i X_i, \quad S^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2.$$

Последнее равенство верно, так как $\sum_i (X_i - \mu)^2 = nS^2 + n(\overline{X} - \mu)^2$,

что легко доказать из равенства $\sum_i (X_i - \mu)^2 = \sum_i (X_i - \overline{X} + \overline{X} - \mu)^2$.

Пример ОМП. Нормальное распределение

Логарифм функции правдоподобия

$$\ln L(\mu, \sigma) = -n \ln \sigma - \frac{nS^2}{2\sigma^2} - \frac{n(\overline{X} - \mu)^2}{2\sigma^2}$$

Возьмём производные по параметрам и приравняем их к нулю

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = \frac{n(\overline{X} - \mu)}{\sigma^2} = 0 \qquad \Rightarrow \qquad \hat{\mu} = \overline{X}$$

$$\frac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{nS^2}{\sigma^3} + \frac{n(\overline{X} - \mu)^2}{\sigma^3} = 0 \qquad \Rightarrow \qquad \hat{\sigma} = S$$

Пример ОМП. Равномерное распределение

Example

Пусть $X_1, \ldots, X_n \sim Unif(0, \theta)$. Оценить θ.

Плотность равномерного распределения

$$f(x;\theta) = \begin{cases} 1/\theta, & 0 \leq x \leq \theta \\ 0, & \text{иначе} \end{cases}$$

Функция правдоподобия

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & \theta \geq X_{(n)} \\ 0, & \theta < X_{(n)} \end{cases}$$

где $X_{(n)} = \max\{X_1, \ldots, X_n\}$.
Получаем $\hat{\theta} = X_{(n)}$.

Свойства ОМП

Основные свойства ОМП:

1. ОМП состоятельная: $\hat{\theta} \xrightarrow{\mathbb{P}} \theta^*$, где θ^* - истинное значение параметра θ.
2. Инвариантность ОМП: если $\hat{\theta}$ - ОМП параметра θ, то $g(\hat{\theta})$ - ОМП для $g(\theta)$.
3. ОМП асимптотически нормальная: $(\hat{\theta} - \theta^*)/\hat{se} \rightsquigarrow N(0,1)$.
4. ОМП является асимптотически оптимальной или эффективной.
   Среди всех хороших оценок ОМП имеет наименьшую дисперсию, по крайней мере, для больших выборок.

Оценка максимального правдоподобия

Дана выборка из параметрического распределения $X_1, \ldots, X_n \sim F_\theta, \quad \theta \in \Theta$.
Функция правдоподобия $L(\theta) = \prod_{i=1}^n p(x_i; \theta)$.
Оценка максимального правдоподобия параметра θ: $\hat{\theta} = \arg\max_\theta L(\theta)$

Example (Смесь нормальных распределений)

$$p(x) = \sum_{i=1}^{2} w_i \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(x-a_i)^2}{2\sigma_i^2}}, \qquad \text{где} \sum_{i=1}^{2} w_i = 1$$

Экспоненциальное семейство распределений

Многие известные и популярные распределения могут быть представлены в обобщенном виде:

$$f(x) = \frac{1}{h(\theta)} g(x) e^{\theta^T u(x)}$$

$h(\theta) \in \mathbb{R}^1, \quad x \in \mathbb{R}^m, \quad \theta \in \mathbb{R}^\alpha, \quad \alpha \ll m, \quad u(x) = (u_1(x), \ldots, u_\alpha(x))^T.$

Распределение	Плотность	$\boldsymbol{u(x)}$	$\boldsymbol{\theta}$
Бернулли	$q^x(1-q)^{1-x}$	x	$\log \frac{q}{1-q}$
Мультиномиальное	$\prod_k \mu_k^{x_k}$	$[x_1, \ldots, x_{K-1}]$	$\theta_i = \log \frac{\mu_i}{1-\sum_j \mu_j}$
Нормальное	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$[x, x^2]$	$\left[-\frac{1}{2\sigma}, \frac{\mu}{\sigma^2}\right]$
Гамма	$\frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$	$[\log x, x]$	$[a-1, -b]$
Бета	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$	$[\log(x), \log(1-x)]$	$[a-1, b-1]$
Пуассон	$\exp(-\lambda)\frac{\lambda^x}{x!}$	$[x, \log\Gamma(x+1)]$	$[k, -1]$

Экспоненциальное семейство распределений

$$f(x) = \frac{1}{h(\theta)} g(x) e^{\theta^T u(x)}$$

Theorem

Для распределений из экспоненциального семейства ОМП существует и единственна.

$\ln L(X^n, \theta) = -n \ln h(\theta) + \sum_{i=1}^n \ln g(X_i) + \sum_{i=1}^n \theta^T u(X_i)$

$$\frac{\partial \ln L(X^n, \theta)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \left(-\ln h(\theta) + \theta^T \left(\frac{1}{n} \sum_{i=1}^n u(X_i) \right) \right)$$

$$= -\frac{1}{h(\theta)} \frac{\partial (h\theta)}{\partial \theta_i} + \frac{1}{n} \sum_{j=1}^n u_i(X_j)$$

$$\frac{\partial^2 \ln L(X^n, \theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{h^2(\theta)} \frac{\partial h(\theta)}{\partial \theta_i} \frac{\partial h(\theta)}{\partial \theta_j} - \frac{1}{h(\theta)} \frac{\partial^2 h(\theta)}{\partial \theta_i \partial \theta_j}$$

Экспоненциальное семейство распределений

$$f(x) = \frac{1}{h(\theta)} g(x) e^{\theta^T u(x)}$$

Распишем нормировочный множитель

$$h(\theta) = \int g(x) e^{\theta^T u(x)} dx$$

$$\frac{\partial h(\theta)}{\partial \theta_i} = \int u_i(x) g(x) e^{\theta^T u(x)} dx$$

$$\frac{1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta_i} = \int \frac{1}{h(\theta)} u_i(x) g(x) e^{\theta^T u(x)} dx = \mathbb{E} u_i(x)$$

Тогда

$$\frac{\partial^2 \ln L(X^n, \theta)}{\partial \theta_i \partial \theta_j} = \mathbb{E} u_i \mathbb{E} u_j - \mathbb{E}(u_i u_j) = -cov(u)$$

Предельные теоремы

Закон больших чисел

Пусть $X_1, \ldots, X_n$ н.о.р.с.в. с конечным вторым моментом $\mathbb{E}X_1^2 < \infty$.
Тогда

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1$$

Центральная предельная теорема

Пусть $X_1, \ldots, X_n$ н.о.р.с.в.:
$m = \mathbb{E}X_1,\ \sigma^2 = \mathbb{V}X_1,\ \sigma \in (0, \infty)$
Тогда $\forall y \in \mathbb{R}$

$$P\left(\frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}} < y\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} dx = \Phi(y)$$

Доверительный интервал

Definition

Доверительным интервалом с доверительной вероятностью $1 - \alpha$ для параметра θ называется интервал $C_n = (a, b)$, где $a = a(X_1, \ldots, X_n)$ и $b = b(X_1, \ldots, X_n)$ - такие функции выборки, что $\mathbb{P}(\theta \in C_n) \geq 1 - \alpha$.

Пример

Дана выборка $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.
Построить доверительный интервал для μ.

Точечная оценка $\hat{\mu} = \overline{X}$.
Распределение точечной оценки $\hat{\mu} \sim N(\mu, \sigma^2/n)$.
Тогда

$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} \sim N(0,1)$$

Получаем доверительный интервал

$$\mathbb{P}\left\{-z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} \leq z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

$$C_n = \left(\hat{\mu} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ \hat{\mu} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

Пример

$L = p_n(X, \theta) = \prod_{i=1}^n p(x_i, \theta)$

$\dfrac{\partial \ln L}{\partial \theta} = \sum_{i=1}^n \dfrac{\partial \ln p(x_i, \theta)}{\partial \theta}$

Покажем, что $\mathbb{E}\dfrac{\partial \ln p(x, \theta)}{\partial \theta} = 0$

$$1 = \int p(x, \theta) dx$$

$$0 = \int \frac{\partial p(x, \theta)}{\partial \theta} dx = \int \frac{\frac{\partial p(x, \theta)}{\partial \theta} p(x, \theta)}{p(x, \theta)} dx$$

$$= \int \frac{\partial \ln p(x, \theta)}{\partial \theta} p(x, \theta) dx = \mathbb{E}\frac{\partial \ln p(x, \theta)}{\partial \theta}$$

Дисперсия

$$\mathbb{V}\left(\frac{\partial \ln p(x, \theta)}{\partial \theta}\right) = \mathbb{E}\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = -\mathbb{E}\frac{\partial^2 \ln p(x, \theta)}{\partial \theta^2}$$

Пример

Дана выборка $X_1, \dots, X_n \sim Pois(\lambda)$.
Построить доверительный интервал для параметра λ.

$$L(X^n|\lambda) = \prod_{i=1}^{n} \mathbb{P}(X = x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

$$\ln L = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - c$$

$$\frac{\partial \ln L}{\partial \lambda} = -n + \left.\frac{\sum_{i=1}^{n} x_i}{\lambda}\right|_{\lambda=\hat{\lambda}_{ML}} = 0$$

$$\hat{\lambda}_{ML} = \overline{X}$$

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

$$\mathbb{V}\left(\frac{\partial \ln L(X^n|\lambda)}{\partial \lambda}\right) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} x_i}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

Пример

По центральной предельной теореме

$$\frac{\frac{n}{\lambda}\left(\overline{X}-\lambda\right)-0}{\sqrt{\frac{n}{\lambda}}}=\sqrt{\frac{n}{\lambda}}\left(\overline{X}-\lambda\right)\to N(0,1)$$

Доверительный интервал

$$\mathbb{P}\left\{-z_{\frac{\alpha}{2}}\leq\sqrt{\frac{n}{\lambda}}\left(\overline{X}-\lambda\right)\leq z_{\frac{\alpha}{2}}\right\}=1-\alpha$$

Дополнения: Plug-in estimators

Оцениваемый параметр как функция от распределения

Мы можем представить параметры как функцию от распределения:

- **Среднее значение.**

$$\mu(F) = \int x \mathrm{d}F(x)$$

- **Дисперсия.**

$$\sigma^2(F) = \int x^2 \mathrm{d}F(x) - \mu^2(F)$$

- **Медиана.**

$$m(F) = \inf\{x | F(x) \geqslant 1/2\}$$

Plug-In оценивание превращает проблему получения оценки θ в проблему оценивания распределения F. Но как это сделать?

Дополнения: Эмпирическая функция распределения

Definition

Эмпирическая функция распределения $\hat{F}_n$ выборки $X_1, \ldots, X_n$ имеет вид

$$\hat{F}_n(x) = \frac{\sum\limits_{i=1}^{n} I(X_i \le x)}{n}, \qquad \text{где} \quad I(X_i \le x) = \begin{cases} 1, X_i \le x \\ 0, X_i > x \end{cases}$$

Theorem

Пусть $\hat{F}_n$ *- эмпирическая функция распределения, построенная по выборке* $X_1, \ldots, X_n \sim F$*. Тогда*

$$\mathbb{E}(\hat{F}_n(x)) = F(x),$$

$$\mathbb{V}(\hat{F}_n(x)) = \frac{F(x)\,(1 - F(x))}{n},$$

$$\hat{F}_n(x) \xrightarrow{P} F(x).$$

Дополнения: Оценка для часто используемых функций

Использование ECDF в качестве оценки функции распределения

Мы можем использовать ECDF в качестве оценки функции распределения. При этом дифференциал превращается в сумму δ-функций,

$$\mathrm{d}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\delta(x - X_i)$$

которая при интегрировании трансформируется в обычную сумму.

Выборочные оценки

- **Выборочное среднее**:
$$\mu(\hat{F}_n) = \int x\mathrm{d}\hat{F}_n(x) = \int x\frac{1}{n}\sum_{i=1}^{n}\delta(x - X_i) = \frac{1}{n}\sum_{i=1}^{n}X_i.$$
- **Выборочная дисперсия**:
$$\sigma^2(\hat{F}_n) = \int x^2\mathrm{d}\hat{F}_n(x) - \mu^2(\hat{F}_n) = \ldots = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2.$$

Дополнения: Характеристические функции

Definition

Пусть задана случайная величина X с распределением $\mathbb{P}$, тогда **характеристическая функция** задается формулой

$$\phi_X(t) = \mathbb{E}\left[\exp(itX)\right] = \int_{-\infty}^{+\infty} \exp(itX)\mathbb{P}(x)x.$$

Дискретная с.в.

Для дискретной с.в. со значениями x_k и вероятностями p_k х.ф. принимает вид

$$\phi_X(t) = \sum_{k=1}^{N} p_k \exp(itx_k)$$

Пример: распределение Бернулли

$$\phi_X(t) = 1 + p \cdot (\exp(it) - 1)$$

Абсолютно непрерывная с.в.

Пусть с.в. X имеет плотность распределения f_X:

$$\phi_X(t) = \int_{-\infty}^{+\infty} \exp(itx) f_X(x)\mathrm{d}x$$

Пример: $X \sim U[0, 1]$

$$\phi_X(t) = \frac{\exp(it) - 1}{it}$$

Дополнения: применение Plug-In для характеристических функций

Характеристические функции для разных распределений

Различные параметрические семейства имеют разный вид характеристических функций. Мы можем построить характеристическую функцию на основе ECDF оценки и сравнить с настоящим видом распределения.

$$\phi_{ECDF}(t) = \frac{1}{N}\sum_{k=1}^{N} \exp(itX_k)$$

Нормальное распределение

$$\phi_X(t) = \exp(i \cdot \mu t - \sigma^2 t^2/2)$$

Распределение Коши

$$\phi_X(t) = \exp(i \cdot x_0 t - \gamma|t|)$$

Регрессия

Регрессионный анализ позволяет ответить на вопрос какому распределению в большей степени соответствует наша выборка.

Резюме

Сегодня мы узнали:

- Основные определения. Познакомились с точечным оцениванием. Узнали какие естественные требования нужно предъявлять к оценкам параметров.
- Обсудили метод максимального правдоподобия. Научились применять его. И поговорили о его сильных и слабых сторонах.
- Для экспоненциального семейства распределений доказали корректность применения метода максимального правдоподобия.
- Обсудили центральную предельную теорему и закон больших чисел. Они помогают нам доказывать сходимость наших оценок по вероятности.
- Научились строить доверительные интервалы для точечных оценок.

Материалы

Материалы для самостоятельного изучения

1. Larry Wasserman. All of Statistics.