Сравнение оценок

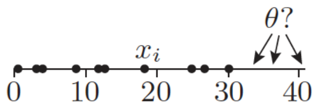
Анализируемые методами математической статистики данные обычно рассматриваются как реализация выборки из некоторого распределения, известного с точностью до параметра (или нескольких параметров). При таком подходе для определения распределения, наиболее подходящего для описания данных, достаточно уметь оценивать значение параметра по реализации.

Статистическая модель

Эксперимент. Пусть θ — некоторое *неизвестное* положительное число. Ниже приведены (с точностью до $0,1$) координаты x_i десяти точек, взятых наудачу из отрезка $[0,\theta]$.

$$3.5 \quad 3.2 \quad 25.6 \quad 8.8 \quad 11.6 \quad 26.6 \quad 18.2 \quad 0.4 \quad 12.3 \quad 30.1.$$

Попробуйте угадать значение параметра θ, на котором изображены эти точки.



С формальной точки зрения мы имеем дело со следующей моделью: набор x_i — это реализация независимых и равномерно распределенных на отрезке $[0,\theta]$ случайных величин X_i с функцией распределения

$$F_\theta(x) = \begin{cases} 0, & \text{если } x \le 0, \\ x/\theta, & \text{если } 0 < x < \theta, \\ 1, & \text{если } x \ge \theta. \end{cases}$$

Здесь $\theta \in \Theta = (0,+\infty)$ — неизвестный параметр масштаба.

Статистическая модель. В общем случае задается семейство функций распределения $\{F_\theta(x), \theta \in \Theta\}$, где Θ — множество возможных значений параметра; данные $x_1,\ldots,x_n$ рассматриваются как реализация выборки $X_1,\ldots,X_n$, элементы которой имеют функцию распределения $F_{\theta_0}(x)$ при некотором неизвестном значении $\theta_0 \in \Theta$. Задача состоит в том, чтобы оценить (восстановить) θ_0 по выборке $x_1,\ldots,x_n$, по возможности, *наиболее точно.*

Как «угадать» задуманное значение, основываясь на наблюдениях $x_1,\ldots,x_n$? Будем оценивать θ_0 при помощи некоторых функций $\hat\theta$ от n переменных $x_1,\ldots,x_n$, которые называются *оценками* или *статистиками.*

Для приведенных выше данных эксперимента в качестве оценок неизвестного параметра масштаба можно использовать, скажем, $\hat\theta_1 = x_{(n)} = \max\{x_1,\ldots,x_n\}$ и $\hat\theta_2 = 2(x_1+\ldots+x_n)/n$. Интуитивно понятно, что при увеличении n каждая из оценок будет приближаться именно к тому значению θ, с которым моделировалась выборка. Но какая из них точнее? Каким образом вообще можно сравнивать оценки? Прежде чем дать ответы на эти вопросы, познакомимся с важнейшими свойствами оценок — несмещенностью и состоятельностью.

Несмещенность и состоятельность

Определение. Оценка $\hat\theta(x_1,\ldots,x_n)$ параметра θ называется *несмещенной*, если $\mathbb{E}_\theta\, \hat\theta(X_1,\ldots,X_n) = \theta$ для всех $\theta \in \Theta$.

Здесь индекс θ у $\mathbb{E}_\theta$ означает, что имеется в виду математическое ожидание случайной величины

$\hat{\theta}(X_1, \ldots, X_n)$, где X_i распределены с функцией распределения $F_\theta(x)$. В дальнейшем этот индекс будет опускаться, чтобы формулы не выглядели слишком громоздко.

Замечание. Важно, чтобы условие несмещенности выполнялось для всех $\theta \in \Theta$. *Тривиальный контрпример*: оценка $\hat{\theta}(x_1, \ldots, x_n) = 1$, идеальная при $\theta = 1$, при других значениях θ имеет смещение

$$b(\theta) = \mathbb{E}\hat{\theta} - \theta = 1 - \theta.$$

Иногда представляет интерес получение оценки не для самого параметра θ, а для некоторой заданной функции $\phi(\theta)$.

Пример 1. Для выборочного контроля из партии готовой продукции отобраны n приборов. Пусть величины $X_1, \ldots, X_n$ — их времена работы до поломки. Допустим, что X_i одинаково показательно распределены с неизвестным параметром $\theta : F_\theta(x) = 1 - e^{-\theta x}, x > 0$. Требуется оценить *среднее время до поломки прибора*

$$\phi(\theta) = \mathbb{E}X_1 = \theta \int_0^{+\infty} x e^{-\theta x} dx = \frac{1}{\theta} \int_0^{+\infty} y e^{-y} dy = \frac{1}{\theta}.$$

По свойствам математического ожидания *выборочное среднее* $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$ будет несмещенной оценкой для функции $\phi(\theta)$: $\mathbb{E}\overline{X} = \phi(\theta)$.

Пример 2. Рассмотрим выборку из какого-либо распределения с двумя параметрами μ и σ, где $\mu = \mathbb{E}X_1$ и $\sigma^2 = \mathrm{Var}X_1$ (скажем, нормального закона $\mathcal{N}(\mu, \sigma^2)$). По свойствам математического ожидания выборочное среднее $\overline{X}$ несмещенно оценивает параметр μ. В качестве оценки для неизвестной дисперсии $\phi(\sigma) = \sigma^2$ можно взять *выборочную дисперсию*

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

Однако, оценка S^2 *имеет смещение.* Действительно, так как случайные величины X_i независимы и одинаково распределены, то, применяя свойства математического ожидания, получаем:

$$\begin{aligned} \mathbb{E}S^2 &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n^2} \sum_{i,j=1}^{n} X_i X_j \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}X_i^2 - \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}X_i^2 - \frac{1}{n^2} \sum_{i \neq j}^{n} \mathbb{E}X_i \mathbb{E}X_j \\ &= \mathbb{E}X_1^2 - \frac{1}{n}\mathbb{E}X_1^2 - \frac{n-1}{n}(\mathbb{E}X_1)^2 = \frac{n-1}{n}\mathrm{Var}X_1. \end{aligned}$$

Чтобы устранить смещение, достаточно домножить S^2 на $n/(n-1)$.

Само по себе свойство несмещенности *не достаточно* для того, чтобы оценка хорошо приближала неизвестный параметр. Например, первый элемент X_1 выборки из закона Бернулли служит несмещенной оценкой для θ: $\mathbb{E}X_1 = 0 \cdot (1 - \theta) + 1 \cdot \theta = \theta$. Однако, его возможные значения 0 и 1 даже не принадлежат $\Theta = (0, 1)$. Необходимо, чтобы погрешность приближения стремилась к нулю с увеличением размера выборки. Это свойство в математической статистике называется *состоятельностью.*

Определение. Оценка $\hat{\theta}(x_1, \ldots, x_n)$ параметра θ называется *состоятельной*, если для всех $\theta \in \Theta$ последовательность

$$\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n) \xrightarrow{P} \theta \quad \text{при } n \to \infty.$$

Состоятельность оценки (а точнее — последовательности оценок $\{\hat{\theta}_n\}$) означает концентрацию вероятностной массы около истинного значения параметра с ростом размера выборки n.

Как установить, будет ли данная оценка состоятельной? Обычно оказывается полезным один из следующих трех способов:

1. Иногда удается доказать состоятельность, непосредственно вычисляя функцию распределения оценки.
2. Другой способ проверки состоит в использовании закона больших чисел и свойства сходимости: если случайные величины ξ_n сходятся по вероятности к случайной величине ξ, то есть $\xi_n \xrightarrow{P} \xi$, и функция $\phi(x)$ непрерывна, то $\phi(\xi_n) \xrightarrow{P} \phi(\xi)$.
3. Часто установить состоятельность помогает следующая лемма.

Лемма. *Если оценка $\hat{\theta}_n$ не смещена, $\mathbb{E}\hat{\theta}_n = \theta$, и дисперсия $\mathrm{Var}\hat{\theta}_n$ стремятся к нулю при $n \to \infty$, то оценка $\hat{\theta}_n$ состоятельна.*

Доказательство. Согласно неравенству Чебышева:

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \le \frac{\mathrm{Var}\hat{\theta}_n}{\varepsilon^2} \to 0 \text{ при } n \to \infty.$$

Лемма доказана. □

Упражнение 1. *Для случайных величин $X_i, i = 1, \ldots, n$, взятых наудачу из отрезка $[0, \theta]$, докажите состоятельность оценки $\hat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$ первым способом.*

Упражнение 2. *Для случайных величин, распределенных экспоненциально с неизвестным параметром $\theta > 0$ (см. Пример 1), докажите состоятельность оценки $\hat{\theta} = 1/\overline{X}$ параметра θ вторым способом.*

Упражнение 3. *Для случайных величин $X_i, i = 1, \ldots, n$, взятых из распределения $\mathcal{N}(\theta, 1)$, докажите состоятельность оценки $\hat{\theta} = \overline{X} = (X_1 + \ldots + X_n)/n$ третьим способом.*

Методы получения оценок

В этой главе рассматриваются несколько методов получения оценок параметров статистических моделей, в том числе — метод моментов и метод максимального правдоподобия.

Plug-in оценки

Если нам необходимо оценить параметр θ и $\theta = \mathbb{E}\phi(X_1)$ для некоторой функции $\phi(x)$, то мы можем рассмотреть оценку вида

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i).$$

Нам уже встречались оценки такого типа. Например, в модели Бернулли, когда X_i имею вероятность успеха с неизвестным параметром $\theta \in (0, 1)$, мы оценивали $\theta = \mathbb{E}X_1$ с помощью среднего $\overline{X}$. По закону больших чисел такие оценки оказываются несмещенными и состоятельными. Если $\mathrm{Var}\phi(X_1) < +\infty$, то можно построить доверительный интервал для plug-in оценки (как мы делали в главе про Монте-Карло).

Метод моментов

Моментом k-го порядка случайной величины X называется величина $\alpha_k = \mathbb{E}X^k$. Моменты существуют не всегда. Например, у закона Коши математическое ожидание α_1 не определено.

Положим $A_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^k$. Если момент α_k существует, то в силу закона больших чисел $A_k \xrightarrow{P} \alpha_k$ (это plug-in оценки моментов). Поэтому для реализации $x_1, \ldots, x_n$ выборки достаточно большого размера можно утверждать, что $a_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^k \approx \alpha_k$, т. е. эмпирические моменты k-го порядка a_k близки к теоретическим моментам α_k. На этом соображении основывается так называемый метод моментов.

Допустим, что распределение элементов выборки зависит от m неизвестных параметров $\theta_1, \ldots, \theta_m$, где вектор $\theta = (\theta_1, \ldots, \theta_m)$ принадлежит некоторой области Θ в $\mathbb{R}^m$. Пусть $\mathbb{E}|X|^m < \infty$ для всех $\theta \in \Theta$ (отсюда следует конечность всех моментов до m из неравенства Ляпунова). Тогда существуют все $\alpha_k = \alpha_k(\theta)$, $k = 1, \ldots, m$, и можно записать систему из m (вообще говоря, нелинейных) уравнений

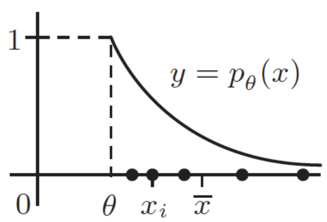
$$\alpha_k(\theta) = a_k, \quad k = 1, \ldots, m.$$

Предположим, что левая часть системы задает взаимно однозначное отображение $g : \Theta \to B$, где B — некоторая область в $\mathbb{R}^m$, и что обратное отображение $g^{-1} : B \to \Theta$ непрерывно. Другими словами, для всех $(y_1, ..., y_m)$ из B система имеет единственное решение, которое непрерывно зависит от правой части. Компоненты решения $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_m)$ при $y_k = A_k$ называются *оценками метода моментов*.

Пример 3. Рассмотрим модель сдвига показательного закона, в которой плотностью распределения величин X_i служит функция $p_\theta(x) = e^{-(x-\theta)} \cdot \mathbf{1}_{\{x \geq \theta\}}$. Здесь

$$\alpha_1(\theta) = \mathbb{E}X_1 = \int_\theta^{+\infty} x e^{-(x-\theta)} dx = \int_0^{+\infty} (y + \theta) e^{-y} dy = 1 + \theta.$$

Из уравнения $1 + \theta = A_1 = \overline{X}$, находим по методу моментов оценку $\hat{\theta} = \overline{X} - 1$.



Какими статистическими свойствами обладают оценки, полученные методом моментов? Их состоятельность вытекает из непрерывности определенного выше отображения g^{-1}.

Метод максимального правдоподобия

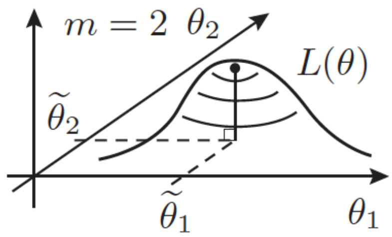
Метод получил распространение после появления в 1912 г. статьи Р. Фишера, где было доказано, что получаемые этим методом оценки являются асимптотически наиболее точными при выполнении некоторых условий регулярности модели.

Для знакомства с методом предположим для простоты, что элементы выборки X_i имеют дискретное распределение: $f(x, \theta) = P(X_1 = x)$ (здесь $\theta = (\theta_1, \ldots, \theta_m)$ — вектор неизвестных параметров модели). Тогда совместная вероятность выборки

$$f(x,\theta) = f(x_1,\theta)\cdot\ldots\cdot f(x_n,\theta)$$

зависит от $n+m$ аргументов (здесь $x = (x_1, ..., x_n)$). Рассматриваемая как функция от $\theta_1, \ldots, \theta_m$ при фиксированных значениях элементов выборки $x_1, \ldots, x_n$, она называется *функцией правдоподобия* и обычно обозначается через $L(\theta)$. Величину $L(\theta)$ можно считать мерой правдоподобия значения $L(\theta)$ при заданной реализации x.

Представляется разумным в качестве оценок параметров $\theta_1, \ldots, \theta_m$ взять наиболее правдоподобные значения $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_m$ которые получаются при максимизации функции $L(\theta)$. Такие оценки называются *оценками максимального правдоподобия* (ОМП).



Часто проще искать точку максимума функции $\ln L(\theta)$, которая совпадает с $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_m)$ в силу монотонности логарифма.

Пример 4. Для схемы Бернулли $X_1, \ldots, X_n$ с вероятностью «успеха» θ имеем:

$$f(x,\theta) = P(X_1 = x) = \theta^x(1-\theta)^{1-x},$$

где x принимает значения 0 или 1. Поэтому функция правдоподобия $L(\theta) = \theta^{s_n}(1-\theta)^{n-s_n}$, где $s_n = x_1 + \ldots + x_n$, представляет собой многочлен n-й степени. Найдем точку максимума

$$\ln L(\theta) = s_n \ln\theta + (n - s_n)\ln(1-\theta).$$

Дифференцируя по θ, получаем уравнение $s_n/\theta - (n - s_n)/(1-\theta) = 0$, откуда $\tilde{\theta} = s_n/n = \bar{x}$. Таким образом, ОМП в схеме Бернулли — это частота «успехов» в реализации $x_1, \ldots, x_n$.

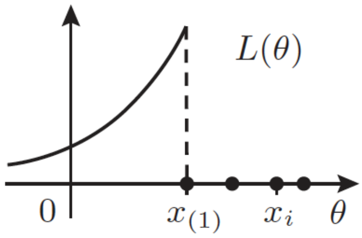
В случае непрерывных моделей будем использовать обозначение $f(x,\theta)$ для плотности распределения случайной величины X_1.

Пример 5. Рассмотрим модель сдвига показательного закона с плотностью $f(x,\theta) = e^{-(x-\theta)}\,\mathbf{1}_{\{x\geq\theta\}}$. В этом случае функция правдоподобия равна

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = e^{-(x_1+\ldots+x_n)}e^{n\theta}\,\mathbf{1}_{\{x_{(1)}\geq\theta\}}.$$

Отсюда (см. рис. ниже) получаем в качестве ОМП $\tilde{\theta} = x_{(1)}$, которая отлична от оценки метода моментов $\hat{\theta} = \bar{x} - 1$, найденной ранее для этой модели в Примере 3. Заметим также, что здесь $L(\theta)$ не является гладкой функцией, и поэтому ОМП нельзя вычислять, приравнивая нулю производную функции правдоподобия.

В случае, когда $L(\theta)$ гладко зависит от $\theta_1, \ldots, \theta_m$, оценки максимального правдоподобия являются



компонентами решения (вообще говоря, нелинейной) системы уравнений:

$$\frac{\partial}{\partial\theta_j}\ln L(\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\ln f(x_i,\theta) = 0, \quad j = 1,\ldots,m.$$

Доверительные интервалы

Вместо того, чтобы приближать неизвестный скалярный параметр θ с помощью «точечной» оценки $\hat\theta$, можно локализовать его иначе — указать случайный интервал $(\hat\theta_1, \hat\theta_2)$, который накрывает θ с вероятностью близкой к единице.

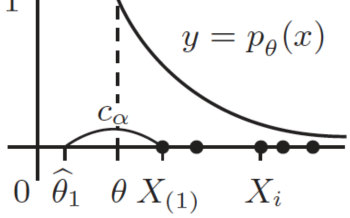
Определение. Пусть $\alpha \in (0,1)$. Две статистики $\hat\theta_1$ и $\hat\theta_1$ определяют границы *доверительного интервала для параметра θ с коэффициентом доверия* $1-\alpha$, если при всех $\theta \in \Theta$ для выборки $\mathbf{X} = (X_1,\ldots,X_n)$ из закона распределения $F_\theta(x)$ справедливо неравенство

$$\mathbb{P}\big(\hat\theta_1(\mathbf{X}) < \theta < \hat\theta_2(\mathbf{X})\big) \ge 1-\alpha.$$

Часто на практике полагают $\alpha = 0.05$. Если вероятность в левой части неравенства стремится к $1-\alpha$ при $n\to\infty$, то интервал называется *асимптотическим* (так у нас было, например, в главе про метод Монте-Карло). Как правило, длина доверительного интервала возрастает при увеличении коэффициента доверия $1-\alpha$ и стремится к нулю с ростом размера выборки n.

Пример 6. Для модели сдвига показательного закона с плотностью $p_\theta(x) = e^{-(x-\theta)}\,\mathbf{1}_{\{x\ge\theta\}}$ оценкой максимального правдоподобия является $X_{(1)} = \min\{X_1,\ldots,X_n\}$. Поскольку $\theta < X_{(1)}$, можно взять $X_{(1)}$ в качестве $\hat\theta_2$. Попробуем подобрать константу c_α так, чтобы для $\hat\theta_1 = X_{(1)} - c_\alpha$ (см. рис.) при всех θ выполнялось тождество

$$\mathbb{P}(X_{(1)} - c_\alpha < \theta < X_{(1)}) = \mathbb{P}(X_{(1)} - c_\alpha < \theta) = 1-\alpha.$$



Используя независимость и показательность величин $X_i - \theta$, перепишем условие:

$$\alpha = \mathbb{P}(X_{(1)} - \theta \ge c_\alpha) = \mathbb{P}(X_i - \theta \ge c_\alpha,\ i = 1,\ldots,n) = e^{-nc_\alpha}.$$

Откуда находим, что длина интервала $c_\alpha = (-\ln\alpha)/n$. Отметим, что $c_\alpha \to \infty$ при $\alpha \to 0$ и $c_\alpha \to 0$ при $n \to \infty$.

Пример 7. Допустим, что элементы выборки X_i распределены по закону $\mathcal{N}(\theta, \sigma^2)$, причем параметр масштаба σ известен, а параметр сдвига θ — нет. Эту модель часто применяют к данным, полученным при независимых измерениях некоторой величины θ с помощью прибора (или метода), имеющего известную среднюю погрешность (стандартную ошибку) σ.

Пусть $\Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$. Для $0 < \alpha < 1$ обозначим через x_α так называемую α-квантиль этого закона, т. е. решение уравнения $\Phi(x_\alpha) = \alpha$. Приведем некоторые значения $x_{1-\alpha/2}$:

α	0,05	10^{-2}	10^{-3}	10^{-5}
$x_{1-\alpha/2}$	1,96	2,58	3,29	4,26

ОМП оценкой для θ служит $\overline{X}$. Известно, что $\overline{X} \sim \mathcal{N}(\theta, \sigma^2/n)$. Тогда $\sqrt{n}(\overline{X} - \theta)/\sigma \sim \mathcal{N}(0,1)$. Поэтому в качестве границ интервала с коэффициентом доверия $1-\alpha$ можно взять

$$\hat{\theta}_1 = \overline{X} - \frac{\sigma x_{1-\alpha/2}}{\sqrt{n}} \quad \text{и} \quad \hat{\theta}_2 = \overline{X} - \frac{\sigma x_{\alpha/2}}{\sqrt{n}}.$$

Мы получили

$$\mathbb{P}(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \mathbb{P}(x_{\alpha/2} < \sqrt{n}(\overline{X} - \theta)/\sigma < x_{1-\alpha/2}) = 1 - \alpha.$$

В силу четности плотности закона $\mathcal{N}(0,1)$ верно равенство $x_{\alpha/2} = -x_{1-\alpha/2}$. Таким образом, из приведенной выше таблицы видим, что с вероятностью $0,95$ истинное значение параметра сдвига θ находится в интервале $\overline{X} \pm 1.96\sigma/\sqrt{n} \approx 2\sigma/\sqrt{n}$ (правило двух сигм).

Список литературы

[1] М. Б. Лагутин. Наглядная математическая статистика. Бином, 2009.