

Математические основы искусственного интеллекта.

Метод максимального правдоподобия

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

Март 2022

Важной является задача оценивания параметров генеральной совокупности $X(\theta)$.

Что дано?

- 1 Априорные сведения о виде распределения генеральной совокупности.
- 2 Выборка из генеральной совокупности x_1, x_2, \dots, x_n .

Что надо найти?

Оценку $\hat{\theta}(x_1, x_2, \dots, x_n)$ параметра θ .

Метод максимального правдоподобия или метод наибольшего правдоподобия (ММП, ML, MLE — англ. maximum likelihood estimation) в математической статистике — это метод оценивания неизвестного параметра путем максимизации функции правдоподобия.

Основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия.

Определение

Функция правдоподобия в математической статистике — это совместное распределение выборки из параметрического распределения, рассматриваемое как функция параметра.

При этом используется

- 1 либо совместная функция плотности (в случае выборки из непрерывного распределения),
- 2 либо совместная вероятность (в случае выборки из дискретного распределения)

вычисленные для данных выборочных значений.

Понятия вероятности и правдоподобия тесно связаны. Сравните два предложения:

«Какова вероятность выпадения 12 очков в каждом из ста бросков двух костей?»

«Насколько правдоподобно, что кости не шулерские, если из ста бросков в каждом выпало 12 очков?»

Если распределение вероятности зависит от параметра θ , то

- 1 с одной стороны, можно рассматривать условную вероятность событий x при заданном параметре θ ,
- 2 с другой стороны — вероятность заданного события X при различных значениях параметра θ .

Первый случай соответствует функции, зависящей от события x :

$$P(x) = P(x | \theta),$$

а второй — функции, зависящей от параметра θ при фиксированном событии X :

$$L(\theta) = L(\theta | x = X).$$

Последнее выражение и есть функция правдоподобия и показывает, насколько правдоподобно выбранное значение параметра θ при известном событии X .

Неформально: если вероятность позволяет предсказывать неизвестные результаты, основанные на известных параметрах, то правдоподобие позволяет оценивать неизвестные параметры, основанные на известных результатах.

$$L(\theta \mid x) = p_{\theta}(x) = P_{\theta}(X = x)$$

Важно понимать, что по абсолютному значению правдоподобия нельзя делать никаких вероятностных суждений. Правдоподобие позволяет сравнить несколько вероятностных распределений с разными параметрами и оценить в контексте какого из них наблюдаемые события наиболее вероятны.

Пусть дано параметрическое семейство распределений вероятности $\{\mathbb{P}_\theta\}_{\theta \in \Theta}$, и дана выборка $X_1, \dots, X_n \sim \mathbb{P}_\theta$ для некоторого $\theta \in \Theta$. Предположим, что совместное распределение этой выборки задаётся функцией $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$, $\mathbf{x} \in \mathbb{R}^n$, где $f_{\mathbf{X}}$ является либо плотностью вероятности, либо функцией вероятности случайного вектора $\mathbf{X} = (X_1, \dots, X_n)^\top$.

Для фиксированной реализации выборки $\mathbf{X} = \mathbf{x}$ функция

$$f_{\mathbf{X}}(\theta \mid \mathbf{x}): \Theta \rightarrow \mathbb{R}$$

называется функцией правдоподобия.

Пусть есть выборка X_1, \dots, X_n из распределения \mathbb{P}_θ , где $\theta \in \Theta$ — неизвестные параметры. Пусть $L(\mathbf{x} \mid \theta): \Theta \rightarrow \mathbb{R}$ — функция правдоподобия, где $\mathbf{x} \in \mathbb{R}^n$. Точечная оценка

$$\hat{\theta}_{\text{МП}} = \hat{\theta}_{\text{МП}}(X_1, \dots, X_n) = \operatorname{argmax}_{\theta \in \Theta} L(X_1, \dots, X_n \mid \theta)$$

называется оценкой максимального правдоподобия параметра θ .

Таким образом оценка максимального правдоподобия — это такая оценка, которая максимизирует функцию правдоподобия при фиксированной реализации выборки.

Часто вместо функции правдоподобия L используют логарифмическую функцию правдоподобия $l = \ln L$. Так как функция $x \rightarrow \ln x$, $x > 0$ монотонно возрастает на всей области определения, максимум любой функции $L(\theta)$ является максимумом функции $\ln L(\theta)$, и наоборот. Таким образом,

$$\hat{\theta}_{\text{МП}} = \operatorname{argmax}_{\theta \in \Theta} l(X_1, \dots, X_n \mid \theta),$$

Максимум функции правдоподобия

Если функция правдоподобия дифференцируема, то необходимое условие экстремума — равенство нулю ее градиента:

$$\frac{\partial L(\mathbf{x}, \theta_0)}{\partial \theta} = 0.$$

В покоординатной форме необходимое условие условием МП-оценок:

$$\begin{cases} \frac{\partial}{\partial \theta_1} \ln L_n(x, \theta) &= 0 \\ \dots & \dots \\ \frac{\partial}{\partial \theta_k} \ln L_n(x, \theta) &= 0 \end{cases}$$

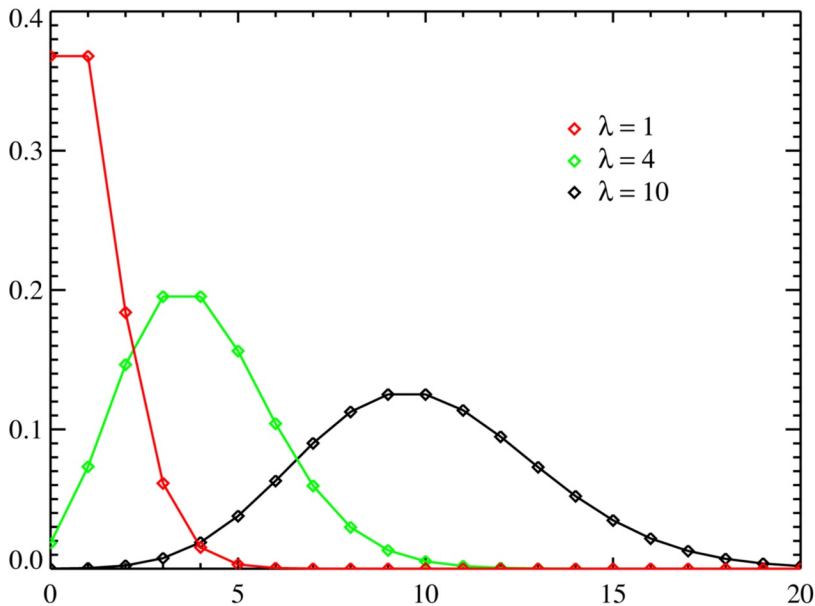
где $L_n(x, \theta) = \prod_{i=1}^n f(x_i, \theta)$ — функция правдоподобия выборки x объема n .

Достаточное условие экстремума может быть сформулировано как отрицательная определенность гессиана — матрицы вторых производных:

$$H = \frac{\partial^2 l(\mathbf{x}, \theta_0)}{\partial \theta \partial \theta^T}.$$

Оценки максимального правдоподобия, вообще говоря, могут быть смещенными, но являются состоятельными, асимптотически эффективными и асимптотически нормальными оценками.

Распределение Пуассона



Метод максимального правдоподобия для дискретных случайных величин

Пусть X_1, X_2, \dots, X_n — случайная выборка объема n . Пусть известен вид закона распределения генеральной совокупности X , но неизвестен параметр θ , которым определяется этот закон. Требуется найти его точечную оценку.

Обозначим вероятность того, что в результате испытания величина X_i примет значение i , $i = 1, 2, \dots, n$, через $p(i; \theta)$.

Функция правдоподобия дискретной случайной величины X имеет вид

$$L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta)p(x_2; \theta) \dots p(x_n; \theta),$$

где x_1, x_2, \dots, x_n — фиксированные числа.

В качестве точечной оценки параметра θ принимают такое его значение $\theta^* = \theta^*(x_1, x_2, \dots, x_n)$, при котором функция правдоподобия достигает максимума. Оценку θ^* называют оценкой наибольшего правдоподобия.

Пример. Оценка параметра λ распределения Пуассона

Найти методом наибольшего правдоподобия оценку параметра λ распределения Пуассона.

Составим функцию правдоподобия, учитывая, что $\theta = \lambda$:

$$\begin{aligned} L &= p(x_1; \lambda) p(x_2; \lambda) \dots p(x_n; \lambda) = \\ &= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \dots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}. \end{aligned}$$

Найдем логарифмическую функцию правдоподобия:

$$\ln L = \ln \lambda \sum x_i - n\lambda - \ln(x_1! x_2! \dots x_n!).$$

Найдем первую производную по λ :

$$\frac{d \ln L}{d \lambda} = \frac{\sum x_i}{\lambda} - n.$$

Напишем уравнение правдоподобия, для чего приравняем первую производную нулю:

$$\frac{\sum x_i}{\lambda} - n = 0.$$

Найдем критическую точку, для чего решим полученное уравнение относительно λ :

$$\frac{\sum x_i}{n} = \lambda.$$

Найдем вторую производную по λ :

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{\sum x_i}{\lambda^2}.$$

Легко видеть, что при $\lambda = \frac{\sum x_i}{n}$ вторая производная отрицательна; следовательно, $\lambda = \frac{\sum x_i}{n}$ — точка максимума и, значит, в качестве оценки наибольшего правдоподобия параметра λ распределения Пуассона надо принять выборочную среднюю $\lambda^* = \frac{\sum x_i}{n}$.

Метод максимального правдоподобия для непрерывных случайных величин

Пусть X_1, X_2, \dots, X_n — случайная выборка объема n . Допустим, что вид плотности распределения $f(x)$ задан, но неизвестен параметр θ , которым определяется эта функция. Требуется найти его точечную оценку.

Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta),$$

где x_1, x_2, \dots, x_n — фиксированные числа.

Оценку наибольшего правдоподобия неизвестного параметра распределения непрерывной случайной величины ищут так же, как в случае дискретной величины.

Пример. Оценка параметра λ показательного распределения

Найти методом наибольшего правдоподобия оценку параметра λ , показательного распределения.

Составим функцию правдоподобия, учитывая, что $\theta = \lambda$:

$$\begin{aligned} L &= f(x_1; \lambda) f(x_2; \lambda) \dots f(x_n; \lambda) = \\ &= \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda(x_1 + x_2 + \dots + x_n)}. \end{aligned}$$

Найдем логарифмическую функцию правдоподобия:

$$\ln L = n \ln \lambda - \lambda \sum x_i.$$

Пример. Оценка параметра λ показательного распределения

Найдем первую производную по λ :

$$\frac{d \ln L}{d \lambda} = n \frac{1}{\lambda} - \sum x_i.$$

Напишем уравнение правдоподобия, для чего приравняем первую производную нулю:

$$\frac{n}{\lambda} - \sum x_i = 0.$$

Найдем критическую точку, для чего решим полученное уравнение относительно λ :

$$\frac{n}{\sum x_i} = \lambda.$$

Пример. Оценка параметра λ показательного распределения

Найдем вторую производную по λ :

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{n}{\lambda^2}.$$

Легко видеть, что при $\lambda = \frac{n}{\sum x_i}$ вторая производная отрицательна; следовательно, $\lambda = \frac{n}{\sum x_i}$ — точка максимума и, значит, в качестве оценки наибольшего правдоподобия параметра λ показательного распределения надо принять выборочную среднюю $\lambda^* = \frac{n}{\sum x_i}$.

- 1 Сформулируйте
- 2 Пусть $X_1, \dots, X_n \sim U[0, \theta]$ — независимая выборка из непрерывного равномерного распределения на отрезке $[0, \theta]$, где $\theta > 0$ — неизвестный параметр. Какаой вид имеет функция правдоподобия? Будет ли МП-оценка параметра θ несмещенной?