

# **Глава 7.**

## **Точечное оценивание параметров распределений случайных величин**

# 7.1. Основные понятия, определения и критерии точечного оценивания

Пусть наблюдается СВ  $X$  с функцией распределения  $F(x)$  и плотностью распределения  $f(x)$ .

Случайная выборка измерения представлена вектором  $X^n = (X_1, \dots, X_n)$  с реализацией  $x^n = (x_1, \dots, x_n)$ .

Будем предполагать, что законы распределения элементов выборки  $X_i$  совпадают с законом распределения наблюдаемой случайной величины, а закон распределения случайного вектора  $X^n=(X_1, \dots, X_n)$  может быть найден по формулам теории вероятностей.

- *Параметром* распределения случайной величины назовем ее числовую характеристику (математическое ожидание, дисперсию, момент и т.п.) либо неизвестную константу, которая явно содержится в выражении функции распределения.

Параметр распределения будем обозначать  $a$ , имея в виду, что в общем случае это векторная величина с компонентами  $a_1, \dots, a_r$ .

Введем случайную величину  $\tilde{\theta} = g(X^n)$  с реализацией

$\tilde{\theta}_j = g(x^n)$ , где  $g$  – борелевская функция.

Эту функцию назовем *статистикой*. В общем случае статистика  $\tilde{\theta}$  – векторная случайная величина с компонентами  $\tilde{\theta}_j = g_j(X^n)$  и их реализацией  $\tilde{\theta} = g_j(x^n)$ .

Статистику  $\tilde{\theta}$ , реализация которой  $\tilde{\theta}$  принимается в качестве приближенного экспериментального значения параметра  $a$ , будем называть точечной оценкой этого параметра.

Ясно, что не всякая зависимость  $g_i(X^n)$  может давать удовлетворительную оценку неизвестного параметра  $a$ ; чтобы быть подходящей,  $g_i(X^n)$  должна обладать определенными свойствами. Именно, в соответствии с принципом оптимальности добиваются, чтобы оценка

$$\tilde{\theta} = g(X^n)$$

удовлетворяла следующим критериям точечного оценивания:

состоятельность, несмещенность, эффективность, достаточность и робастность.

Если все эти свойства обеспечить не удастся, то ограничиваются удовлетворением хотя бы какой-то их части.

*Состоятельность* оценки – это сходимость ее по вероятности к оцениваемому параметру при  $n \rightarrow \infty$ .

В случае состоятельной оценки вероятность сколь-нибудь существенного отличия  $\tilde{\theta}$  от  $a$  мала при достаточно большом объеме выборки измерений  $X$ . Поскольку сходимость по вероятности следует из сходимости почти наверное и в среднем квадратическом, то сходимость последнего вида также следует считать признаком состоятельности оценки  $\tilde{\theta}$ . В этом случае говорят о сильной состоятельности.

- *Несмещенность* оценки – это свойство вида:  $M(\tilde{\theta})=a$ .

Несмещенность гарантирует совпадение центра рассеяния возможных реализаций с оцениваемым параметром  $a$ .

Если это свойство не выполняется, т.е.  $M(\tilde{\theta}) - a = b(a) \neq 0$ , то оценку называют *смещенной*, при этом величину  $b(a)$  называют *систематической ошибкой* (смещением) оценки  $\tilde{\theta}$ .

Очевидно, что  $b(a)$  зависит от  $n$ . Если  $b(a) \rightarrow 0$  при  $n \rightarrow \infty$ , то оценку  $\tilde{\theta}$  называют асимптотически несмещенной. Не следует абсолютизировать это свойство. Во-первых, несмещенной оценки может не быть; во-вторых, требование несмещенности может прийти в противоречие с требованием минимума рассеяния  $\tilde{\theta}$  относительно  $a$ .

Оценку  $\tilde{\theta}$  называют *эффективной*, если ей соответствует минимальное значение ошибки.

В классе несмещенных оценок этот критерий означает минимальность дисперсии  $D(\tilde{\theta})$ .

Оценку  $\tilde{\theta}$  называют *асимптотически эффективной*, если свойство минимальности относительно  $\tilde{\theta}$  достигается в пределе  $n \rightarrow \infty$ .

Оценку  $\tilde{\theta}$  называют *достаточной*, если она содержит в себе столько же информации о параметре  $a$ , сколько её в выборке

$$X^n = (X_1, \dots, X_n).$$

Оценку  $\tilde{\theta}$  называют *робастной*, если она в каком-то смысле слабо зависит от изменения выборки  $X^n = (X_1, \dots, X_n)$ .

Введенные признаки оптимизации оценок относятся к фиксированному состоянию параметра  $a$ .

• Если эти признаки имеют место для всякого  $a$  из области его возможных реализаций, то говорят о *равномерной* состоятельности, несмещенности, эффективности, достаточности и робастности оценки  $\tilde{\theta}$  .

Оценки параметров распределений, удовлетворяющих хотя бы одному из перечисленных критериев оптимальности, могут строиться по разному.

Основная *задача* теории точечного оценивания состоит в разработке *методов* построения оценок, обладающих всеми или какими-то из перечисленных оптимальных свойств.

## 7.2. Точечные оценки математического ожидания и дисперсии

Пусть имеется случайная величина  $X$  с математическим ожиданием  $m$  и дисперсией  $D$ ; оба параметра неизвестны.

- Над величиной  $X$  произведено  $n$  независимых опытов, давших результаты  $X_1, \dots, X_n$ . Требуется найти состоятельные и несмещенные оценки для параметров  $m$  и  $D$ .

В качестве оценки для математического ожидания естественно предложить среднее арифметическое наблюдаемых значений  $m^*$ :

$$\tilde{m} = m^* = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

- Нетрудно убедиться, что эта оценка является состоятельной: согласно закону больших чисел, при увеличении  $n$  величина  $\tilde{m}$  сходится по вероятности к  $m$ . Оценка  $\tilde{m}$  является также и несмещенной, так как

$$M[\tilde{m}] = \frac{\sum_{i=1}^n m}{n} = m \quad . \quad (2)$$

Дисперсия этой оценки равна:

$$D[\tilde{m}] = \frac{1}{n} D \quad . \quad (3)$$

- Эффективность или неэффективность оценки зависит от вида закона распределения величины  $X$ . Можно доказать, что если величина  $X$  распределена по нормальному закону, дисперсия будет минимально возможной, т. е. оценка  $\tilde{m}$  является эффективной. Для других законов распределения это может быть и не так.

- Перейдем к оценке дисперсии  $D$ . На первый взгляд наиболее естественной оценкой представляется статистическая дисперсия:

$$D^* = \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n}, \quad (4)$$

где

$$\tilde{m} = \bar{x} = \frac{\sum_{i=1}^n X_i}{n} \quad (5)$$

- Проверим, является ли эта оценка состоятельной. Выразим ее через второй начальный момент

$$D^* = \frac{\sum_{i=1}^n X_i^2}{n} - \tilde{m}^2 \quad (6)$$

Первый член в правой части есть среднее арифметическое  $n$  наблюдаемых значений случайной величины  $X_i^2$ ; он сходится по вероятности к  $M[X^2] = \alpha_2[X]$ .

Второй член сходится по вероятности к  $m^2$ ;  
вся величина сходится по вероятности к  
величине

$$\alpha_2[X] - m^2 = D. \quad (7)$$

Это означает, что оценка  $D^*$  состоятельна.  
Проверим, является ли оценка  $D^*$  также и  
несмещенной. Подставим в формулу (6)  
вместо  $\tilde{m}$  его выражение (5) и произведем  
указанные действия:

$$\dot{D}^* = \frac{\sum_{i=1}^n X_i^2}{n} - \left( \frac{\sum_{i=1}^n X_i}{n} \right)^2 = \frac{n-1}{n^2} \sum_{i=1}^n X_i^2 - 2 \frac{\sum_{i<j} X_i X_j}{n^2}$$

(7)

Найдем математическое ожидание величины (7):

$$M[D^*] = \frac{n-1}{n} \sum_{i=1}^n M[X_i^2] - \frac{2}{n^2} \sum_{i<j} M[X_i X_j] \quad (8)$$

Так как дисперсия  $D^*$  не зависит от того, в какой точке выбрано начало координат, выберем его в точке  $m$ .

$$M[X_i^2] = D \quad \sum_{i=1}^n M[X_i^2] = nD \quad (9)$$

- $$M[X_i X_j] = K_{ij} = 0 \quad (10)$$

Последнее равенство следует из того, что опыты независимы. Подставляя (9) и (10) в (8), получим:

$$M[D^*] = \frac{n-1}{n} D \quad (11)$$

Отсюда видно, что величина  $D^*$  не является несмещенной оценкой для  $D$ : ее математическое ожидание не равно  $D$ , а несколько меньше.

Пользуясь оценкой  $D^*$  вместо дисперсии  $D$ , мы будем совершать некоторую систематическую ошибку в меньшую сторону. Чтобы ликвидировать это смещение, достаточно ввести поправку, умножив величину  $D^*$  на  $\frac{n-1}{n}$ . Получим:

$$\begin{aligned}
 \frac{n-1}{n} D^* &= \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n} \frac{n}{n-1} = \\
 &= \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n-1}
 \end{aligned}$$

Такую «исправленную» статистическую дисперсию мы и выберем в качестве оценки для  $D$ :

$$\tilde{D} = \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n-1} \quad (12)$$

Так как множитель  $\frac{n}{n-1}$  стремится к единице при  $n \rightarrow \infty$ , а оценка  $D^*$  состоятельна, то оценка  $\tilde{D}$  также будет состоятельной.

На практике часто вместо формулы (12) бывает удобнее применять другую, равносильную ей, в которой статистическая дисперсия выражена через второй начальный момент:

$$\tilde{D} = S^2 = \left[ \frac{\sum_{i=1}^n X_i^2}{n} - \tilde{m}^2 \right] \frac{n}{n-1} \quad (13)$$

- При больших значениях  $n$ , естественно, обе оценки - смещенная  $D^*$  и несмещенная  $\tilde{D}$  - будут различаться очень мало и введение поправочного множителя теряет смысл.

Таким образом, мы пришли к следующим правилам обработки ограниченного по объему статистического материала.

- Если даны значения  $x_1, x_2, \dots, x_n$ , принятые в  $n$  независимых опытах случайной величиной  $X$  с неизвестными математическим ожиданием  $m$  и дисперсией  $D$ , то для определения этих параметров следует пользоваться приближенными значениями (оценками):

$$\tilde{m} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$
$$\tilde{D} = S^2 = \frac{\sum_{i=1}^n (x_i - \tilde{m})^2}{n - 1}$$

или

$$\tilde{D} = S^2 = \left[ \frac{\sum_{i=1}^n x_i^2}{n} - \tilde{m}^2 \right] \frac{n}{n-1}$$

# **7.3. Методы получения оценок параметров распределения**

Часто на практике на основании анализа физического механизма, порождающего случайную величину  $X$ , можно сделать вывод о законе распределения этой случайной величины. Однако параметры этого распределения неизвестны, и их необходимо оценить по результатам эксперимента, обычно представленных в виде конечной выборки  $x_1, x_2, \dots, x_n$ . Для решения такой задачи чаще всего применяются два метода: метод моментов и метод максимального правдоподобия.

# 7.3.1. Метод моментов

- Метод моментов состоит в приравнивании теоретических моментов соответствующим эмпирическим моментам того же порядка.

Эмпирические начальные моменты  $k$ -го порядка определяются формулами:

$$\alpha_k^* = \frac{1}{n} \sum_{i=1}^n (x_i)^k,$$

а соответствующие им теоретические начальные моменты  $k$ -го порядка – формулами:

$$\alpha_k = \sum_{i=1}^n (x_i)^k p(x, a)$$

для дискретных случайных величин,

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x, a) dx$$

для непрерывных случайных величин,

где  $a$  – оцениваемый параметр распределения.

Для получения оценок параметров распределения, содержащего два неизвестных параметра  $a_1$  и  $a_2$ , составляется система из двух уравнений

$$\begin{cases} \alpha_1(\hat{a}_1, \hat{a}_2) = \alpha_1^*(\hat{a}_1, \hat{a}_2) \\ \mu_2(\hat{a}_1, \hat{a}_2) = \mu_2^*(\hat{a}_1, \hat{a}_2) \end{cases}$$

где  $\mu_2$  и  $\mu_2^*$  – теоретический и эмпирический центральные моменты второго порядка.

- Решением системы уравнений являются оценки  $\hat{a}_1$  и  $\hat{a}_2$  неизвестных параметров распределения  $a_1$  и  $a_2$ .

Приравняв теоретический эмпирический начальные моменты первого порядка, получаем, что оценкой математического ожидания случайной величины  $X$ , имеющей произвольное распределение, будет выборочное среднее, т. е.

$$M[X] = m^* = \frac{1}{n} \sum_{i=1}^n x_i.$$

- Затем, приравняв теоретический и эмпирический центральные моменты второго порядка, получим, что оценка дисперсии случайной величины  $X$ , имеющей произвольное распределение, определяется формулой

$$D[X] = D^* = \frac{1}{n} \sum_{i=1}^n (x_i - m^*)^2.$$

Подобным образом можно найти оценки теоретических моментов любого порядка.

Метод моментов отличается простотой и не требует сложных вычислений, но полученные этим методом оценки часто являются неэффективными.

## 7.3.2. Метод максимума правдоподобия

Метод максимального правдоподобия точечной оценки неизвестных параметров распределения сводится к отысканию максимума функции одного или нескольких оцениваемых параметров.

Пусть  $X$  – непрерывная случайная величина, которая в результате  $n$  испытаний приняла значения  $x_1, x_2, \dots, x_n$ .

Для получения оценки неизвестного параметра  $a$  необходимо найти такое значение  $\hat{a}$ , при котором вероятность реализации полученной выборки была бы максимальной. Так как  $x_1, x_2, \dots, x_n$  представляют собой взаимно независимые величины с одинаковой плотностью вероятности  $f(x)$ , то функцией правдоподобия называют функцию аргумента  $a$ :

$$L(x_1, x_2, \dots, x_n; a) = f(x_1; a) \cdot f(x_2; a) \dots f(x_n; a).$$

Оценкой максимального правдоподобия параметра  $a$  называется такое значение  $\hat{a}$ , при котором функция правдоподобия достигает максимума, т. е. является решением уравнения

$$\left. \frac{dL(a)}{da} \right|_{a=\hat{a}} = 0,$$

которое явно зависит от результатов испытаний  $x_1, x_2, \dots, x_n$ .

Поскольку функции  $L(a)$  и  $\ln L(a)$  достигают максимума при одних и тех же значениях  $\hat{a} = u(x_1, x_2, \dots, x_n)$ , то часто для упрощения расчетов используют логарифмическую функцию правдоподобия и ищут корень соответствующего уравнения

$$\left. \frac{d \ln L(a)}{da} \right|_{a=\hat{a}} = 0,$$

которое называется уравнением правдоподобия.

Если необходимо оценить несколько параметров  $a_1, a_2, \dots, a_k$  распределения  $f(x; a_1, a_2, \dots, a_k)$ , то функция правдоподобия будет зависеть от этих параметров. Для нахождения оценок  $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k$  параметров распределения необходимо решить систему  $k$  уравнений правдоподобия

$$\frac{d}{da} \ln L(a_1, a_2, \dots, a_k) \Big|_{a_i = \hat{a}_i} = 0, i = 1, 2, \dots, k.$$

Метод максимального правдоподобия дает состоятельные и асимптотически эффективные оценки. Однако получаемые методом максимального правдоподобия оценки бывают смещенными, и, кроме того, для нахождения оценок часто приходится решать достаточно сложные системы уравнений.