

3. ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Математическая статистика изучает методы построения вероятностных моделей, не противоречащих результатам эксперимента. При этом априори предполагается, что эти результаты *могут быть описаны в вероятностных терминах*¹.

Мы ограничимся в этом курсе рассмотрением двух задач:

- 1) оценивание параметров распределения случайной переменной;
- 2) проверка статистических гипотез.

3.1. Одна содержательная задача

С помощью измерительного прибора проделаны n измерений некоторой физической величины. *Предполагается*, что за время измерений значение этой величины не изменяется. В то же время полученные числа x_1, \dots, x_n будут, вообще говоря, различны. *Предполагается*, что это различие объясняется наличием у измерительного прибора аддитивной случайной погрешности, т.е.

$$\xi = a + \eta,$$

где a – неизвестное значение измеряемой физической величины; η – случайная погрешность прибора; ξ – наблюдаемая физическая величина.

Задача состоит в оценивании значения a на основании полученных данных x_1, \dots, x_n .

В математической модели физическим величинам ξ и η соответствуют случайные переменные, которые мы также обозначим ξ и η .

При построении математической модели *удобно* считать, что имеется не n отсчетов, снятых с одного прибора, а n однотипных приборов, с каждого из которых снимается один отсчет. При этом числовой вектор $x = [x_1, \dots, x_n]^T$ оказывается значением случайного вектора $X = [X_1, \dots, X_n]^T$, координаты которого *независимы в совокупности* (это еще одно предположение!) и имеют одинаковое распределение, совпадающее с распределением случайной переменной ξ .

Далее, *предположим*, что у прибора отсутствует систематическая погрешность, т.е. $M(\eta) = 0$. Тогда $a = M(\xi)$. *Предположим*, наконец, как это часто делают, что случайная погрешность имеет нормальное распределение с известным (из паспорта измерительного прибора) среднеквадратическим отклонением σ .

Сделанные нами *предположения* позволяют сказать, что случайная переменная ξ принадлежит однопараметрическому семейству с.п. с плотностью распределения

$$f_{\xi}(t, a) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right),$$

параметр которого a подлежит оцениванию.

Оценкой параметра a распределения с.п. ξ мы будем, как обычно, называть интервал $[\hat{a} - \Delta, \hat{a} + \Delta]$, где Δ – число, задаваемое постановщиком задачи, \hat{a} – число, определяемое по полученным результатам измерений² ($\hat{a} = \varphi(x)$).

Серьезное предупреждение. Ранее, когда речь шла об оценивании (корня уравнения, суммы ряда, интеграла и т.д.), интервал-оценка покрывал оцениваемое число *гарантированно*. В математической статистике, каково бы ни было назначенное нами число Δ , интервал $[\hat{a} - \Delta, \hat{a} + \Delta]$ в единичном эксперименте может содержать оцениваемый параметр a , а может не содержать. Однако если эксперимент в *неизменных* (еще одно предположение!) условиях повторять многократно, то можно говорить об *относительной частоте* накрытия интервалом $[\hat{a} - \Delta, \hat{a} + \Delta]$ оцениваемого параметра, и, следовательно, о *вероятности* накрытия. Таким образом, сказав, что моделью эксперимента является случайная переменная, мы сказали тем самым, что эксперимент будет *многократно* повторяться. Без этого предположения статистическая обработка результатов измерений *бессмысленна*.

При оценивании математического ожидания в качестве центра интервала чаще всего берут среднее арифметическое результатов измерений

$$\hat{a} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

¹Мы хотим еще раз подчеркнуть, что ответственность за применение математической модели (за вероятностную трактовку содержательной задачи) несет постановщик задачи!

²В математической статистике оценку именуют *интервальной оценкой*, а ее центр \hat{a} – *точечной оценкой* параметра распределения.

Соответствующая с.п. $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, согласно результатам п.2.6, имеет при сделанных предположениях нормальное распределение с параметрами

$$M(\bar{X}) = a; \quad D(\bar{X}) = \frac{\sigma^2}{n},$$

т.е.

$$f_{\bar{X}}(t) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(t-a)^2 n}{2\sigma^2}\right).$$

Итак, $M(\bar{X}) = a$, т.е. математическое ожидание среднего арифметического совпадает с оцениваемым параметром. В силу неравенства Чебышева *случайные* значения среднего арифметического, полученные в различных сериях измерений, будут группироваться около его математического ожидания, т.е. около оцениваемого параметра. Разброс *случайных* значений среднего арифметического относительно $M(\bar{X}) = a$ характеризуется дисперсией $D(\bar{X}) = \frac{\sigma^2}{n}$, которую можно сделать как угодно малой за счет увеличения длины обрабатываемой серии измерений.

При заданном числе Δ можно вычислить вероятность накрытия оцениваемого параметра a интервалом $[\hat{a} - \Delta, \hat{a} + \Delta]$:

$$\begin{aligned} \beta &= \mathbf{P}(|\bar{X} - a| < \Delta) = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\Delta}^{\Delta} \exp\left(-\frac{nx^2}{2\sigma^2}\right) dx = \operatorname{erf}\left(\sqrt{n} \frac{\Delta}{\sigma\sqrt{2}}\right), \end{aligned} \quad (3.1.1)$$

где erf – уже упоминавшаяся в курсе математического анализа "функция ошибок"

$$\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x^2) dx,$$

имеющаяся во всех средах конечного пользователя и библиотеках ФОРТРАНа.

В таблице 3.1 приведены значения переменной $\sqrt{n} \frac{\Delta}{\sigma}$ для некоторых часто используемых значений вероятности накрытия оцениваемого параметра.

Таблица 1:

β	0.9	0.95	0.99	0.995	0.999
$\sqrt{n} \frac{\Delta}{\sigma}$	1.64	1.96	2.58	2.81	3.29

Проанализируем формулу (3.1.1). Она связывает три числа: n – количество обрабатываемых измерений ("стоимость" эксперимента), Δ – полуширина интервала-оценки ("качество" оценки) и β – вероятность накрытия оцениваемого параметра³ ("надежность" оценки). Видно, что при фиксированной стоимости эксперимента (n) повышение качества оценки (уменьшение Δ) приводит к уменьшению ее надежности (β): интеграл от положительной функции убывает при уменьшении длины промежутка интегрирования. Увеличить надежность оценки при сохранении ее качества можно лишь за счет увеличения затрат и т.д.

3.2. Оценивание параметров распределения случайной переменной

Обобщим задачу, рассмотренную в предыдущем пункте. Единственные *объективные* данные, имеющиеся в нашем распоряжении – это экспериментально полученные n чисел w_1, \dots, w_n . Числовой вектор, составленный из этих чисел $w = [w_1, \dots, w_n]^T$, называется *выборкой*, а число n – *объемом выборки*.

Возможность статистической обработки выборки обеспечивается следующими *предположениями*:

1) выборка w является значением случайного вектора

$$X = [X_1 \dots X_n]^T;$$

2) координаты случайного вектора X независимы в совокупности;

³В математической статистике β именуется *доверительной вероятностью*, а интервал-оценка с соответствующей Δ – *доверительным интервалом*.

- 3) координаты случайного вектора X являются копиями наблюдаемой случайной переменной ξ ;
 4) наблюдаемая с.п. ξ принадлежит известному семейству случайных переменных с k параметрами $\vartheta_1, \dots, \vartheta_k$.
 За справедливость этих предположений ответственность несет постановщик задачи (экспериментатор).

Таким образом, задача свелась к оцениванию по имеющейся выборке $w = [w_1 \dots w_n]^T$ параметра (векторного) $\vartheta = [\vartheta_1 \dots \vartheta_k]^T$. Оценкой параметра ϑ мы будем называть k -мерный параллелепипед J с центром в точке $\hat{\vartheta} = [\hat{\vartheta}_1 \dots \hat{\vartheta}_k]^T$ и "ребрами"

$$[\hat{\vartheta}_1 - \Delta_1, \hat{\vartheta}_1 + \Delta_1[, \dots, [\hat{\vartheta}_k - \Delta_k, \hat{\vartheta}_k + \Delta_k[.$$

Центр параллелепипеда-оценки находится по имеющейся выборке w :

$$\hat{\vartheta} = \varphi(w),$$

где $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ – заданная функция, именуемая *статистикой*. Так как предполагается, что процедура оценивания будет повторяться многократно, рассмотрим случайную переменную $T = \varphi(X)$, которую также называют статистикой.

Предположим для определенности, что случайная переменная ξ абсолютно непрерывна. Поскольку ее плотность распределения $f_\xi(x; \vartheta)$ известна с точностью до оцениваемого параметра, можно построить плотность распределения случайного вектора X :

$$f_X(x; \vartheta) = f_\xi(x_1; \vartheta) \cdot f_\xi(x_2; \vartheta) \dots f_\xi(x_n; \vartheta).$$

Отсюда следует *принципиальная* возможность найти плотность распределения статистики T $f_T(t, \vartheta)$.

Поскольку большие отклонения случайного вектора от его математического ожидания маловероятны, можно надеяться, что случайные значения статистики $T = \varphi(X)$ будут концентрироваться около ее математического ожидания. Отсюда вытекает естественное требование: математическое ожидание статистики должно совпадать с оцениваемым параметром ($M(T) = \vartheta$) или хотя бы приближаться к нему при увеличении объема выборки ($\lim_{n \rightarrow +\infty} M(T) = \vartheta$).

Статистика, удовлетворяющая условию $M(T) = \vartheta$, называется *несмещенной*; использование такой статистики при обработке эксперимента не приводит к появлению дополнительной *систематической* ошибки.

Статистика, удовлетворяющая условию $\lim_{n \rightarrow +\infty} M(T) = \vartheta$, называется *асимптотически несмещенной*. Практическое значение асимптотической несмещенности состоит в возможности получения пренебрежимо малой дополнительной систематической ошибки за счет увеличения объема выборки.

Серьезное предупреждение. Обратите внимание, что мы изначально предполагаем отсутствие *систематической погрешности наблюдения*. Если такая погрешность имеется, то ее невозможно устранить никакой математической обработкой!

Предположим, что статистика является несмещенной. Тогда для всех компонент $M(T_j) = \vartheta_j$, и мы можем представить случайное событие "оцениваемый параметр ϑ накрывается параллелепипедом J " так:

$$(\vartheta \in J) = \bigcap_{j=1}^k A_j,$$

где A_j – случайное событие $(|T_j - M(T_j)| < \Delta_j)$.

Оценим вероятность события $\vartheta \in J$ в предположении, что все с.п. T_j имеют дисперсии. В силу неравенства Чебышева

$$\mathbf{P}(|T_j - M(T_j)| < \Delta_j) \geq 1 - \frac{D(T_j)}{\Delta_j^2}, \quad j = 1, \dots, k. \quad (3.2.1)$$

Если $k = 2$, то по теореме сложения вероятностей

$$\mathbf{P}(\vartheta \in J) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cup A_2) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) - 1,$$

откуда с учетом (3.2.1) получаем

$$\mathbf{P}(\vartheta \in J) \geq 1 - \frac{D(T_1)}{\Delta_1^2} - \frac{D(T_2)}{\Delta_2^2}.$$

Аналогично при любом k имеем

$$\beta = \mathbf{P}(\vartheta \in J) \geq 1 - \sum_{j=1}^k \frac{D(T_j)}{\Delta_j^2}. \quad (3.2.2)$$

Предположим вдобавок, что дисперсии всех компонент статистики T стремятся к нулю при неограниченном росте объема выборки. Тогда неравенство (3.2.2) показывает, что вероятность накрытия оцениваемого параметра параллелепипедом-оценкой с заданными длинами ребер может быть сделана как угодно близкой к единице за счет увеличения объема выборки⁴. Такая статистика называется *состоятельной*.

Очевидно, "хорошая" статистика должна быть несмещенной (хотя бы асимптотически) и состоятельной. В следующем параграфе мы рассмотрим один из приемов построения таких "хороших" статистик.

Естественно считать, как и в п.3.1, что размеры *доверительного параллелепипеда* J характеризуют *качество* оценки: чем эти размеры меньше, тем оценка точнее.

Точно так же *доверительная вероятность* β (вероятность накрытия оцениваемого параметра ϑ доверительным параллелепипедом) характеризует *надежность* оценки: чем больше β , тем реже доверительный параллелепипед *не будет* накрывать оцениваемый параметр. Аналогично п.3.1, уменьшение размеров параллелепипеда J приводит к уменьшению доверительной вероятности β .

И, наконец, объем выборки n (количество экспериментов) характеризует *стоимость* получения оценки.

Понятно, что лозунг "лучше, больше, дешевле!" не может быть реализован. Если фиксировать затраты (n), то *увеличение точности* оценки неизбежно приведет к *уменьшению ее надежности*. Желая увеличить надежность, мы будем вынуждены либо пожертвовать точностью, либо заплатить большую цену и т.д.

3.3. Метод максимального правдоподобия

Вообще говоря, статистики изобретаются (и поэтому обычно носят имена их авторов). Метод максимального правдоподобия является одним из регулярных методов построения "хороших" статистик. Его изобрел Р.А. Фишер⁵.

Если $f_X(x; \vartheta)$ – семейство плотностей распределения с векторным параметром ϑ , то, подставив вместо переменного вектора x полученную в эксперименте выборку w , получим новую функцию, заданную на множестве допустимых значений параметра ϑ . Эта функция называется *функцией правдоподобия*⁶ выборки:

$$lik(\vartheta) = f_X(x; \vartheta)|_{x=w}.$$

Поскольку координаты случайного вектора X независимы в совокупности и имеют одинаковые плотности распределения, то

$$lik(\vartheta) = \prod_{j=1}^n f_{\xi}(w_j; \vartheta).$$

Аналогично, в дискретном случае имеем

$$lik(\vartheta) = \prod_{j=1}^n \mathbf{P}(\xi = w_j; \vartheta).$$

Метод максимального правдоподобия предлагает принять в качестве центра доверительного параллелепипеда точку глобального максимума функции правдоподобия.

Примеры. 1. При оценивании математического ожидания нормальной наблюдаемой с.п. ξ с известной дисперсией σ^2 плотность распределения принадлежит однопараметрическому семейству

$$f_{\xi}(t; a) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right).$$

Функция правдоподобия равна

$$lik(a) = f_X(w; a) = \left(\frac{1}{\sqrt{2\pi} \cdot \sigma}\right)^n \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (w_j - a)^2\right).$$

В силу *возрастания* логарифма точка максимума функции правдоподобия совпадает с точкой максимума ее логарифма

$$l(a) = \ln(lik(a)) = -n \cdot \ln(\sqrt{2\pi} \cdot \sigma) - \frac{1}{2\sigma^2} \sum_{j=1}^n (w_j - a)^2.$$

⁴Можно показать, что то же самое справедливо, если заменить несмещенность статистики ее *асимптотической несмещенностью*.

⁵Рональд Айлмер ФИШЕР (1890-1962) – английский биолог, математик и статистик. С его именем связаны многие понятия и утверждения в математической статистике.

⁶Обозначение lik – от английского слова *likelihood* (правдоподобие). Мы настоятельно рекомендуем считать название „функция-правдоподобия“ *одним* словом, чтобы избежать соблазна найти связь с интуитивным представлением о правдоподобии.

Решив уравнение

$$l'(a) \equiv \frac{1}{\sigma^2} \sum_{j=1}^n (w_j - a) = 0,$$

найдем единственную стационарную точку функции правдоподобия

$$\bar{w} = \hat{a} = \frac{1}{n} \sum_{j=1}^n w_j.$$

Поскольку $l''(a) \equiv -\frac{n}{\sigma^2} < 0$, то $\hat{a} = \bar{w}$ – точка *локального максимума*, который, очевидно, является и *глобальным*.

2. Если семейство нормальных плотностей распределения содержит два *оцениваемых* параметра (a и $\mathcal{D} = \sigma^2$), то

$$\begin{aligned} f_{\xi}(t; a, \mathcal{D}) &= \frac{1}{\sqrt{2\pi\mathcal{D}}} \cdot \exp\left(-\frac{(t-a)^2}{2\mathcal{D}}\right); \\ \text{lik}(a, \mathcal{D}) = f_X(w; a, \mathcal{D}) &= \left(\frac{1}{\sqrt{2\pi\mathcal{D}}}\right)^n \cdot \exp\left(-\frac{1}{2\mathcal{D}} \sum_{j=1}^n (w_j - a)^2\right); \\ l(a, \mathcal{D}) = \ln(\text{lik}(a, \mathcal{D})) &= -\frac{n}{2} \cdot \ln(2\pi\mathcal{D}) - \frac{1}{2\mathcal{D}} \sum_{j=1}^n (w_j - a)^2. \end{aligned}$$

Решая уравнение $\nabla l = \theta$, т.е. систему

$$\begin{cases} \frac{1}{\mathcal{D}} \sum_{j=1}^n (w_j - a) = 0 \\ -\frac{n}{2\mathcal{D}} + \frac{1}{2\mathcal{D}^2} \sum_{j=1}^n (w_j - a)^2 = 0 \end{cases},$$

найдем *единственную* стационарную точку:

$$\bar{w} = \hat{a} = \frac{1}{n} \sum_{j=1}^n w_j; \quad s^2 = \hat{\mathcal{D}} = \frac{1}{n} \sum_{j=1}^n (w_j - \bar{w})^2.$$

Вычислив в этой точке матрицу Гессе:

$$l''(\hat{a}, \hat{\mathcal{D}}) = \begin{bmatrix} -\frac{n}{\hat{\mathcal{D}}} & 0 \\ 0 & -\frac{n}{2\hat{\mathcal{D}}} \end{bmatrix},$$

убедимся в ее *отрицательной определенности*. Отсюда следует, что стационарная точка – точка *локального максимума*, который, очевидно, является и *глобальным*.

3. Наблюдаемая с.п. ξ принадлежит семейству равномерных распределений с двумя параметрами a и b ($b > a$):

$$f_{\xi}(t; a, b) = \begin{cases} \frac{1}{b-a} & \text{при } t \in [a, b]; \\ 0 & \text{при } t \notin [a, b]. \end{cases}$$

Функция правдоподобия выборки

$$\text{lik}(a, b) = \begin{cases} \frac{1}{(b-a)^n} & \text{при } a \leq \min_j(w_j) \text{ и } \max_j\{w_j\} \leq b; \\ 0 & \text{при } a > \min_j(w_j) \text{ или } \max_j\{w_j\} > b. \end{cases}$$

Очевидно, дробь $\frac{1}{b-a}$ будет максимальна при минимальном значении ее знаменателя. Но a нельзя сделать больше, чем $\min_j(w_j)$, а b – меньше, чем $\max_j(w_j)$. Итак,

$$\hat{a} = \min_j(w_j), \quad \hat{b} = \max_j(w_j).$$

4. Дискретная случайная переменная ξ принадлежит семейству распределений с одним параметром p ($0 < p < 1$):

$$\mathbf{P}(\xi = t; p) = \begin{cases} p, & t = 1; \\ 1 - p, & t = 0. \end{cases}$$

Обозначив количество экспериментов n , а количество единиц ("успехов") в этих экспериментах m , получим

$$lik(p) = \mathbf{P}(X = w; p) = \prod_{j=1}^n \mathbf{P}(X_j = w_j; p) = p^m \cdot (1 - p)^{n-m}$$

(среди n сомножителей m равны p , а остальные равны $1 - p$). Поэтому

$$l(p) = \ln(lik(p)) = m \cdot \ln(p) + (n - m) \cdot \ln(1 - p).$$

Производная $l'(p) = \frac{m - np}{p(1 - p)}$ обращается в нуль только в одной точке, и эта единственная стационарная точка $\hat{p} = \frac{m}{n}$ дает, очевидно, глобальный максимум функции правдоподобия. Таким образом, *относительная частота успеха* является статистикой максимального правдоподобия для вероятности в случае двух исходов эксперимента.

Найдем числовые характеристики статистик максимального правдоподобия, полученных в этих примерах.

Пример 1. Статистика $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$.

Как показано в п.3.1,

$$M(\bar{X}) = a, \quad D(\bar{X}) = \frac{\sigma^2}{n}.$$

Пример 2. Статистики

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2.$$

Аналогично примеру 1 получаем

$$M(\bar{X}) = a, \quad D(\bar{X}) = \frac{D}{n}.$$

Можно показать, что

$$M(S^2) = \frac{n-1}{n} D; \quad D(S^2) = \frac{(n-1)^2}{n^3} \cdot \mu_4(\xi) - \frac{(n-1)(n-3)}{n^3} \cdot D^2$$

(здесь $\mu_4(\xi) = M((\xi - a)^4)$ – четвертый центральный момент с.п. ξ).

Пример 3. Статистики

$$X_{max} = \max_j(X_j); \quad X_{min} = \min_j(X_j).$$

Построим функцию распределения

$$F_{X_{max}}(t) = \mathbf{P}(X_{max} < t).$$

Неравенство $X_{max} < t$ означает, что все координаты вектора X меньше, чем t . Поэтому

$$F_{X_{max}}(t) = (F_{\xi}(t))^n = \begin{cases} 0 & \text{при } t < a; \\ \left(\frac{t-a}{b-a}\right)^n & \text{при } a \leq t \leq b; \\ 1 & \text{при } t > b. \end{cases}$$

Несложные вычисления дают

$$M(X_{max}) = \frac{nb + a}{n+1}; \quad D(X_{max}) = \frac{n(b-a)^2}{(n+1)(n+2)}.$$

Аналогично, для статистики X_{min} получаем

$$M(X_{min}) = \frac{na + b}{n+1}; \quad D(X_{min}) = \frac{n(b-a)^2}{(n+1)(n+2)}.$$

Пример 4. Статистика "относительная частота успеха".

Обозначим с.п. "количество успехов" ν . Тогда, очевидно, относительная частота успеха равна $\frac{\nu}{n}$.

Как известно из п.2.3, с.п. ν имеет биномиальное распределение, причем $M(\nu) = n \cdot p$, $D(\nu) = n \cdot p \cdot (1 - p)$. Отсюда

$$M\left(\frac{\nu}{n}\right) = p, \quad D\left(\frac{\nu}{n}\right) = \frac{p(1-p)}{n}.$$

Отметим, что для статистики \bar{X} в примерах 1 и 2 математическое ожидание совпадает с оцениваемым параметром, т.е. \bar{X} – *несмещенная* статистика. То же можно сказать и об относительной частоте в примере 4. В то же время в примерах 2 и 3

$$M(S^2) = \frac{n-1}{n} \mathcal{D} \neq \mathcal{D},$$

$$M(X_{max}) = \frac{nb+a}{n+1} \neq b, \quad M(X_{min}) = \frac{na+b}{n+1} \neq a,$$

т.е. эти три статистики – *смещенные*. Однако они являются *асимптотически несмещенными*, так как

$$\lim_{n \rightarrow +\infty} M(S^2) = \mathcal{D},$$

$$\lim_{n \rightarrow +\infty} M(X_{max}) = b, \quad \lim_{n \rightarrow +\infty} M(X_{min}) = a.$$

Отметим также, что дисперсии всех рассмотренных статистик стремятся к нулю при $n \rightarrow \infty$. В силу неравенства (3.2.2) все эти статистики *состоятельны*.

Можно показать, что при достаточно общих допущениях статистики максимального правдоподобия являются состоятельными и асимптотически несмещенными.