Тема 1. ОБЩИЕ СВОЙСТВА СТАТИСТИЧЕСКИХ ОЦЕНОК

Теоретическое введение

Простейшая постановка задачи оценивания неизвестного параметра такова.

Пусть $x_1, \ldots, x_n$ — выборка, т.е. n независимых испытаний случайной величины X с функцией распределения $F(x;a)$, где a - параметр, значение которого неизвестно. Требуется по выборке оценить значение параметра.

Оценкой $\hat{a} = \varphi(x_1, \ldots, x_n)$ *называется функция наблюдений, используемая для приближенного определения неизвестного параметра*. Значение $\hat{a}$ оценки является *случайной величиной*, поскольку $(x_1, \ldots, x_n)$ — случайная величина, вообще говоря, многомерная.

Свойства оценок

1. Оценка $\hat{a} = \varphi(x_1, \ldots, x_n)$ называется *состоятельной,* если при $n \to \infty$ $\hat{a} \to a$ по вероятности при любом значении a.

2. Оценка $\hat{a} = \varphi(x_1, \ldots, x_n)$ называется *несмещенной*, если при любом a

$$M\hat{a} = M\varphi(x_1, \ldots, x_n) = a.$$

Состоятельность — обязательное свойство используемых оценок. Свойство несмещенности является желательным; многие применяемые оценки свойством несмещенности не обладают.

3. Оценка φ^* называется *оптимальной*, если для неё средний квадрат ошибки

$$M(\hat{a} - a)^2 = M[\varphi^*(x_1, \ldots, x_n) - a]^2 = \min M[\varphi(x_1, \ldots, x_n) - a]^2 = R(a)$$

минимален среди всех оценок $\{\varphi\}$; здесь критерием качества оценки принят квадрат ошибки $(\hat{a} - a)^2$. В более общей ситуации критерием качества служит некоторая величина $L(\hat{a}, a)$, называемая функцией потерь. Ясно, что оптимальной оценки может не существовать (так как характеристикой является функция, а не число).

Примеры

Пример 1. С целью оценить дисперсию σ^2 выборка

$$\xi_1, \xi_2, \ldots, \xi_n, \quad M\xi_i = m, \quad D\xi_i = \sigma^2 = ?,$$

Пример 1. Произведено 20 опытов над величиной X; результаты приведены в таблице 14.3.2.

Таблица 14.3.2

i	x_i	i	x_i	i	x_i	i	x_i
1	10,5	6	10,6	11	10,6	16	10,9
2	10,8	7	10,9	12	11,3	17	10,8
3	11,2	8	11,0	13	10,5	18	10,7
4	10,9	9	10,3	14	10,7	19	10,9
5	10,4	10	10,8	15	10,8	20	11,0

min=10.3, max=11.3

подвергается обработке по *разностям соседних наблюдений*. Применяется следующая процедура:

$$\tilde{\sigma}^2 = k \cdot \sum_{j=1}^{n-1} \left(\xi_{j+1} - \xi_j\right)^2; \tag{1}$$

т.е. суммируются квадраты разностей «соседних» наблюдений. Очевидно, $M\left(\xi_{j+1} - \xi_j\right) = 0$, $M\left(\xi_{j+1} - \xi_j\right)^2 = D\left(\xi_{j+1} - \xi_j\right) = 2\sigma^2$.

Из лекционного курса хорошо известна несмещенная оценка s^2 для σ^2:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2, \quad \bar{\xi} = \frac{1}{n}\cdot\sum_{j=1}^{n}\xi_j \tag{2}$$

однако, она может нас не устраивать по вычислительной реализации, т.к. сначала нужно вычислить среднее, затем все отклонения от среднего, возвести их в квадрат, и квадраты просуммировать.

Процедура (1) более удобна в этом смысле: к предыдущей сумме добавляется очередное слагаемое. Ясно также, что, поскольку $M\left(\xi_{j+1} - \xi_j\right)^2$ пропорционально σ^2, эта статистика оценивает величину разброса, надо лишь подобрать коэффициент *k* так, чтобы получить несмещенную оценку:

$$M\tilde{\sigma}^2 = \sigma^2 \tag{3}$$

Вычисляем $M\tilde{\sigma}^2$:

$$M\tilde{\sigma}^2 = k\cdot\sum_{j=1}^{n-1} M\left(\xi_{j+1} - \xi_j\right)^2 = k(n-1)\cdot M\left(\xi_{j+1} - \xi_j\right)^2 = k\left(n-1\right)\cdot 2\sigma^2$$

Из (3) получаем:

$$k\cdot\left(n-1\right)\cdot 2\sigma^2 = \sigma^2 \Rightarrow k = \frac{1}{2\left(n-1\right)}$$

Пример 2. Пусть $\xi_1, \xi_2, ..., \xi_n$ — выборка; наблюдения подчиняются закону Пуассона: $\xi_i \sim Po(a)$, (напомним, $M\xi_i = m_1 = a$, $D\xi_i = \sigma^2 = a$). Параметр a неизвестен.

Оценить первые два момента. Первый момент m_1 совпадает с параметром, $m_1 = a = ?$, оценив m_1 (мы знаем, как оценивается первый момент), мы оценим параметр *a*.

$$\hat{a} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n}\xi_i = \bar{\xi}$$

Второй момент является функцией от параметра:

$$m_2 = D\xi_i + m_1^2 = a + a^2 = f(a) = ?.$$

Подставив вместо *a* ее оценку $\hat{a}$, мы должны получить разумную оценку (это типичный способ рассуждений в статистике)

$$\hat{m}_2 = f(\hat{a}) = \hat{a} + \hat{a}^2 = \bar{\xi} + (\bar{\xi})^2 = \bar{\xi}(1+\bar{\xi})$$

Проверим ее несмещённость:

$$M\hat{m}_2 = M\left(\overline{\xi} + \overline{\xi}^2\right) = M\overline{\xi} + \left(D\overline{\xi} + (M\overline{\xi})^2\right) = a + \left[\frac{a}{n} + a^2\right] = (a + a^2) + \frac{a}{n} = m_2 + \frac{a}{n} \neq m_2,$$

т.е. оценка смещённая. Можно попытаться исправить ее, сделать несмещенной; для этого нужно избавиться от смещения a/n, но вычесть его из оценки мы не можем, поскольку значение *a* неизвестно; вычтем из оценки $\hat{m}_2$ такую случайную величину (СВ) X, у которой $MX = \frac{a}{n}$; в качестве такой *X* возьмем $X = \hat{a}/n = \overline{\xi}/n$. Новая оценка

$$\tilde{m}_2 = \hat{m}_2 - \frac{\overline{\xi}}{n} = \overline{\xi}(1 + \overline{\xi}) - \frac{\overline{\xi}}{n} = \frac{n-1}{n}\overline{\xi} + \overline{\xi}^2,$$

очевидно, оценка $\tilde{m}_2$ несмещенная:

$$M\tilde{m}_2 = M\hat{m}_2 - M\frac{\overline{\xi}}{n} = \left[(a + a^2) + \frac{a}{n}\right] - \frac{a}{n} = a + a^2 = m_2$$

Дополнительный вопрос: можно ли предложить другую несмещённую оценку? Можно, а именно, дежурную:

$$m_2^* = \frac{1}{n}\sum_{i=1}^{n} \xi_i^2, \quad Mm_2^* = m_2 = a^2 + a$$

Ну, и какая оценка более точная? Нужно сравнить дисперсии. ??? Не вдохновляет. Пока ответить не просто.

Однако, не вычисляя дисперсии, в нашем случае можно заранее сказать, что $\tilde{m}_2$ более точная, т.к. она основана *на достаточной статистике* $\overline{\xi}$.

Пример 3. Произведено n=10 выстрелов и получено $x = 8$ попаданий. Количество попаданий — это СВ ξ, распределённая по биномиальному закону:

$$\xi \sim Bi(n, p), \text{ т.е. } P\{\xi = k\} = C_n^k p^k (1 - p)^{n-k}.$$

Считая параметр p неизвестным, требуется предложить несмещённые оценки $\hat{m}$ и $\hat{\sigma}^2$ для математического ожидания (МО)

$m = np$ и дисперсии $\sigma^2 = npq = np(1-p)$

этого распределения,. Поступим аналогично тому, как это сделано в предыдущем примере. Мы знаем оценку для вероятности p:

$$\hat{p} = \frac{\xi}{n},$$

а m и σ^2 - функции *p*. Рассмотрим оценки

$$\hat{m} = n\hat{p} = n\frac{\xi}{n} = \xi \text{ и } \hat{\sigma}^2 = n\hat{p}(1-\hat{p}) = \xi\left(1 - \frac{\xi}{n}\right) = \xi - \frac{\xi^2}{n}$$

Несмещённость первой очевидна:

$$M\hat{m} = M\xi = np = m.$$

Для второй имеем

$$M\hat{\sigma}^2 = M\left(\xi - \frac{\xi^2}{n}\right) = np - \frac{1}{n}\ D\xi + (M\xi)^2\ = np - \frac{1}{n}\ npq + (np)^2\ =$$

$$= np - np^2 - pq = npq - pq = (n-1)\,pq \neq npq$$

$\Rightarrow \hat{\sigma}^2$ смещённая оценка: мешает множитель $(n-1) \neq n$. Исправим введением подходящего множителя:

$$\tilde{\sigma}^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{n}{n-1}\xi\left(1 - \frac{\xi}{n}\right);$$

оценка, конечно, несмещенная, проверим:

$$M\tilde{\sigma}^2 = \frac{n}{n-1}M\hat{\sigma}^2 = \frac{n}{(n-1)}(n-1)\,pq = npq = \sigma^2$$

Для нашего наблюдения $x = 80$: $\tilde{\sigma}^2 = \frac{10}{9}\cdot 8(1-0{,}8) = \frac{10}{9}\cdot 1.6,\quad \tilde{\sigma} = \frac{4}{3}$

Для небольших n поправку $\frac{n}{n-1}$ нужно учитывать

Пример 4. Для определения вероятности события A производится n независимых опытов, в каждом из которых

$$P(A) \equiv p$$

одна и та же. Определить, при каком значении p дисперсия оценки вероятности будет максимальной.

Решение: ξ — число появлений случайного события A. Оценка

$$\hat{p} = \frac{\xi}{n};\qquad \xi \sim Bi(n, p)$$

$$D\hat{p} = D\frac{\xi}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Это выражение достигает максимума при $p^* = \frac{1}{2}$.

Пример 5. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — результаты независимых измерений неизвестной постоянной величины. Ошибки измерений подчиняются одному и тому же закону нормального распределения.

Стандартное отклонение σ ошибок измерений оценивается по формуле:

$$\hat{\sigma} = k\sum_{j=1}^{n}\left|\xi_j - \bar{\xi}\right|,\ \text{где}\ \bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$$

Определить значение коэффициента k, при котором оценка $\hat{\sigma}$ является несмещенной:

$$k = ?: M\hat{\sigma} = \sigma$$

Решение. Определим МО:

$$M\hat{\sigma} = k\sum_{j=1}^{n}M\left|\xi_j - \bar{\xi}\right| = knM\left|\xi_j - \bar{\xi}\right| = kn\cdot M\left|\eta\right|,$$

где обозначено $\eta \equiv \xi_j - \bar{\xi} \sim N(0, \sigma_1^2)$.

$M|\eta| = ?$, но σ_1^2 неизвестна, $\sigma_1^2 = ?$ - дисперсия разности зависимых с.в..

Вынесем из средне-арифметического $\bar{\xi}$ слагаемое с номером *j*:

$$\eta \equiv \xi_j - \frac{\xi_j}{n} - \frac{1}{n}\sum_{i=1, i\neq j}^{n} \xi_i = \xi_j \frac{n-1}{n} - \frac{1}{n}\sum_{i=1, i\neq j}^{n} \xi_i ,$$

чтобы иметь для η сумму независимых. Теперь для η имеем:

$$\begin{cases} M\eta = 0; \\ D\eta = \left(\frac{n-1}{n}\right)^2 \sigma^2 + \frac{1}{n^2}(n-1)\sigma^2 = \frac{\sigma^2}{n^2}(n-1)\left[(n-1)+1\right] = \sigma^2 \frac{n-1}{n} = \sigma_1^2. \end{cases}$$

$$\eta \sim N\left(0, \sigma_1^2 = \sigma^2 \frac{n-1}{n}\right).$$

Определим

$$M|\eta| = \int_{-\infty}^{\infty} |y| \frac{e^{-y^2/2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}} dy = \frac{2}{\sqrt{2\pi\sigma_1^2}} \int_0^{\infty} y \cdot e^{-y^2/2\sigma_1^2} dy =$$

$$= \frac{2}{\sqrt{2\pi\sigma_1^2}} \sigma_1^2 \int_0^{\infty} e^{-y^2/2\sigma_1^2} d\left(\frac{y^2}{2\sigma_1^2}\right) = \sigma_1 \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-z} dz = \sigma_1 \sqrt{\frac{2}{\pi}}$$

Условие несмещенности:

$$M\hat{\sigma} = kn \cdot M|\eta| = kn\sqrt{\frac{2}{\pi}} \cdot \sigma_1 = kn\sqrt{\frac{2}{\pi}} \cdot \sigma \sqrt{\frac{n-1}{n}} = \sigma.$$

$$k = \frac{1}{n} \cdot \sqrt{\frac{\pi}{2}} \cdot \sqrt{\frac{n}{n-1}} = \sqrt{\frac{\pi}{2n(n-1)}}$$

Домашнее задание: гл.19, №№ 102, 103, 104.

Сборник задач по математике для втузов. Ч.4. / Под редакцией А.В.Ефимова и А.С.Поспелова. М.: Изд. ФизМатЛит., 2003.- 423 с.