МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
Лекция 2

§ 3. Нижняя граница для дисперсии несмещенной оценки

3.1. Информационное неравенство Рао-Крамера

Оказывается, никаким выбором оценочной функции невозможно сделать дисперсию ошибки меньше, чем некоторое определенное значение.

Пусть $x \equiv (x_1, x_2 ... x_n)$ — результаты *n* наблюдений (не обязательно независимых), являющиеся конкретными значениями многомерной случайной величины $\xi \equiv (\xi_1, \xi_2 \ldots \xi_n)$. Закон распределения $p(x;a)$ известен с точностью до параметра *a* (будем считать $p(x;a)$ плотностью распределения, если ξ непрерывна, и вероятностью, если ξ дискретна). Мы рассматриваем различные несмещенные оценки $\varphi(\xi)$ для $f(a)$ (может быть, нам нужна оценка не для *a*, а для $a^2 = f(a)$ или для $\sin a = f(a)$; функция $f(\cdot)$ известна, но значение параметра *a* неизвестно). ***Как обычно, рассматривая статистические задачи, мы отвлекаемся от конкретного значения*** $x \equiv (x_1, x_2 ... x_n)$***, считая его одной из возможных реализаций случайной величины*** $\xi \equiv (\xi_1, \xi_2 \ldots \xi_n)$***, а оценку*** $\varphi(\xi)$ ***рассматриваем как случайную величину.***

Теорема. Для любой оценки $\varphi(\xi)$, несмещенно оценивающей $f(a)$, при условиях, оговариваемых ниже, справедливо неравенство:

$$D\varphi(\xi) \geq \sigma_0^2 \equiv \frac{\left(df(a)/da\right)^2}{I_\xi(a)}, \tag{1}$$

где обозначено

$$I_\xi(a) = \mathrm{M}\left[\frac{\partial \ln p(\xi;a)}{\partial a}\right]^2 = \int_X \left[\frac{\partial \ln p(x;a)}{\partial a}\right]^2 p(x;a)dx\,. \tag{2}$$

Т. е. дисперсия несмещенной оценки не может быть меньше величины σ_0^2, которая определяется правой частью неравенства и которую можно вычислить заранее, если известен закон распределения. Условия, при которых справедливо это неравенство, выпишем после проведения выкладок, из которых будет ясно, что именно нужно потребовать.

Обозначим через X множество тех значений *x*, при которых $p(x;a) \neq 0$, т. е. $X \equiv \{x : p(x;a) \neq 0\}$. Условие несмещенности

$$\mathrm{M}\varphi(\xi) = \int_X \varphi(x) p(x;a)dx = f(a) \tag{3}$$

продифференцируем по *a*. Учитывая, что $p' = p(\ln p)'$, получим

$$\int_X \varphi(x)\frac{\partial l \ln p(x;a)}{\partial a}\, p(x;a)dx = \frac{df(a)}{da}. \qquad (4)$$

Здесь имеются в виду многомерные интегралы. Продифференцируем по *a* условие

$$\int_X p(x;a)dx = 1. \qquad (5)$$

Учтем, что $p' = p(\ln p)'$. Затем, умножив результат на $f(a)$, получим

$$\int_X f(a)\frac{\partial \ln p(x;a)}{\partial a}\, p(x;a)dx = 0. \qquad (6)$$

Вычтем из (4) равенство (6):

$$\int_X [\varphi(x) - f(a)]\frac{\partial \ln p(x;a)}{\partial a}\, p(x;a)dx = \frac{df(a)}{da}. \qquad (6a)$$

Введем обозначения

$$h(x) = [\varphi(x) - f(a)]\sqrt{p(x;a)},$$

$$g(x) = \frac{\partial \ln p(x;a)}{\partial a}\sqrt{p(x;a)}.$$

Возведя (6а) в квадрат, получим:

$$\left[\int h(x)g(x)dx\right]^2 = \left(f_a'\right)^2. \qquad (7)$$

Применим к левой части (7) неравенство Коши – Буняковского – Шварца:

$$\int h^2(x)dx \cdot \int g^2(x)dx \ge \left[\int h(x)g(x)dx\right]^2 = \left(f_a'\right)^2 \qquad (8)$$

Получим:

$$\int_X [\varphi(x) - f(a)]^2 p(x;a)dx \cdot \int_X \left[\frac{\partial \ln p(x;a)}{\partial a}\right]^2 p(x;a)dx \ge \left(\frac{df(a)}{da}\right)^2. \qquad (9)$$

Здесь первый множитель есть $D\varphi(\xi)$, а второй $I_\xi(a)$, так что (9) совпадает с (1).

Условия, которые нужно наложить на $p(x;a)$ для справедливости (1):

1) множество X не должно зависеть от *a*, иначе нельзя дифференцирование в (3) и (5) внести под знак интеграла;
2) функция $p(x;a)$ должна быть дифференцируемой по *a*;
3) интегралы в (9) должны существовать.

Следствие. Если оценивается само значение параметра *a*, т. е. $f(a) \equiv a$, то

$$D\varphi(\xi) \ge \sigma_0^2 \equiv \frac{1}{I(a)}. \qquad (10)$$

3.2. Информация Фишера

Величина $I_\xi(a)$ *в формуле (2) называется* ***информацией Фишера****, содержащейся в выборке* ξ *относительно параметра а.*

Свойства информации.

1. *О вычислении*.

Справедлива следующая формула:

$$I_\xi(a) \equiv \mathrm{M}\left[\frac{\partial \ln p(\xi;a)}{\partial a}\right]^2 = -\mathrm{M}\frac{\partial^2}{\partial a^2}\ln p(\xi;a). \qquad (11)$$

Действительно,

$$(\ln p)'' = \left(\frac{p'}{p}\right)' = \frac{pp'' - p'p'}{p^2} = \frac{p''}{p} - \left(\frac{p'}{p}\right)^2 = \frac{p''}{p} - \left(\frac{\partial \ln p}{\partial a}\right)^2,$$

$$M\,\frac{p''}{p} = \int \frac{p''(x;a)}{p(x;a)}\,p(x;a)dx = \left[\int p(x;a)dx\right]'' = 0.$$

2. *Аддитивность информации Фишера.*

Информация, содержащаяся в n независимых наблюдениях, равна сумме информаций, содержащихся в отдельных наблюдениях.

Действительно, пусть $\xi = (\xi_1, \xi_2 \ldots \xi_n)$, где ξ_i — независимы и распределены по закону $q_j(x_j;a)$. Тогда, если обозначено $x = (x_1, x_2 \ldots x_n)$, то $p(x;a) = \prod_{j=1}^{n} q_j(x_j;a)$. Вычислим информацию в наблюдениях ξ:

$$I_\xi(a) = -\mathrm{M}\frac{\partial \ln p(\xi;a)}{\partial a^2} = \sum_{j=1}^{n} -\mathrm{M}\frac{\partial^2}{\partial a^2}\ln q_j(\xi_j;a) = \sum_{j=1}^{n} I_j(a), \qquad (12)$$

где $I_j(a)$ — информация, содержащаяся в одном наблюдении ξ_j.

Если все наблюдения распределены одинаково, т. е. $q_j(\cdot) = q(\cdot)$ для всех j (это означает, что мы имеем дело с выборкой $(\xi_1, \xi_2 \ldots \xi_n)$), то

$$I_\xi(a) = nI_1(a), \qquad (13)$$

где $I_1(a) = \mathrm{M}\left[\frac{\partial \ln q(\xi_i;a)}{\partial a}\right]^2$ — информация в одном наблюдении.

Следствие. Если $\xi = (\xi_1, \xi_2 \ldots \xi_n)$ — выборка, то из (10) и (13) следует:

$$D\varphi \geq c(a)/n. \qquad (14)$$

Это означает, что в условиях, в которых справедливо неравенство Рао – Крамера, дисперсию оценки нельзя сделать убывающей быстрее, чем $1/n$.

Замечания.

1. Об обобщениях неравенства (1).

Неравенство Рао – Крамера обобщалось в различных направлениях — снимались ограничения, в основном, первое ограничение. Различные обобщения, в том числе на многомерный случай, можно найти в книгах [5], [6].

2. Об использовании неравенства.

С помощью (1) можно выносить суждения о качестве имеющихся оценок. Далеко не всегда удается найти оптимальную оценку, т.к. или она не существует (неравенство не утверждает, что существует оптимальная оценка), или, если существует, то трудно реализуема в вычислительном отношении.

Пусть из каких-либо соображений построена оценка φ. Как оценить ее качество? Если $D\varphi$ близка к σ_0^2, это означает, что оценка «хорошая». Если же нет, то с некоторой осторожностью можно считать, что оценка φ «плохая», поскольку известно, что при широких условиях существуют оценки с дисперсией, асимптотически (с ростом *n*) приближающейся к значению σ_0^2 (см. раздел 5.2).

3.3. Эффективные оценки. Экспонентные семейства распределений

Оценка, дисперсия которой достигает нижней границы, определенной неравенством Рао – Крамера, называетя эффективной.

Эффективность — это самое лучшее в смысле точности, что мы можем ожидать. Возникают вопросы: при каких условиях существуют эффективные оценки, и как их искать? Ответом на второй вопрос является следующее утверждение.

Утверждение. Эффективная оценка $\varphi(x)$ для $f(a)$, если она существует, равна

$$\varphi(x) = f(a) + \frac{df\ /\ da}{I_\xi(a)} \cdot \frac{\partial \ln p(x;a)}{\partial a}, \tag{15}$$

причем зависимость правой части от параметра *a* фиктивна.

Действительно, равенство в (8) возможно тогда и только тогда, когда

$$h(x) = Cg(x), \tag{16}$$

где *C* = *C*(*a*) может зависеть от *a*. Подставив это в (6а), найдем *C*(*a*):

$$C(a)\int g^2(x)dx = f_a{}', \qquad C(a) = f_a{}'/I(a).$$

Подставив это в (16) и сократив на $\sqrt{p(x;a)}$, получим:

$$\varphi(x) - f(a) = \frac{f_a{}'}{I(a)} \frac{\partial \ln p(x;a)}{\partial a}, \tag{17}$$

что эквивалентно (15). Заметим, что слева в (15) стоит оценка, т.е. функция наблюдений. Следовательно, зависимость правой части от параметра *a* фиктивна.

Пример 1. Проверка на эффективность оценки для вероятности случайного события.

Пусть имеется случайное событие A. Его вероятность q неизвестна. Проведено n независимых испытаний, число выпадений события ξ — случайная величина, распределенная по биномиальному закону $Bi(n,q)$ с параметрами n и q. Роль неизвестного параметра играет q. Рассмотрим оценку $\hat{q}=\xi/n$, ее дисперсия $\mathrm{D}\hat{q}=q(1-q)/n$. Эффективна ли оценка $\hat{q}$? Нижняя граница Рао – Крамера $\sigma_0^2=1/I_\xi(q)$. Учитывая, что

$$p_\xi(x;q)=P\{\xi=x\}=C_n^x q^x(1-q)^{n-x},\ x = 0, 1\ldots n,$$

вычислим информацию $I(q)$:

$$I_\xi(q)=\mathrm{M}\left[\frac{\partial \ln p_\xi(\xi;q)}{\partial q}\right]^2=\mathrm{M}\left[\frac{\xi-qn}{q(1-q)}\right]^2=\frac{\mathrm{D}\xi}{q^2(1-q)^2}=\frac{n}{q(1-q)},$$

откуда $\sigma_0^2=q(1-q)/n=\mathrm{D}\hat{q}$. Т. е. оценка $\hat{q}=\xi/n$ эффективна, лучшей несмещенной оценки не существует.

Возникает вопрос: какие распределения допускают существование эффективных оценок? Ответ получим, если соотношение (15) понимать как дифференциальное уравнение относительно $p(x,a)$:

$$\frac{\partial \ln p(x;a)}{\partial a}=\frac{I_\xi(a)}{f'(a)}\left[\varphi(x)-f(a)\right].$$

После интегрирования по a получим

$$p(x;a)=C(x)e^{A(a)\varphi(x)}g(a).$$

Распределения такого вида называют *экспонентными*. Здесь существенно то, что логарифм множителя, зависящего от a и от x, есть произведение функции $A(a)$ от параметра a на функцию $\varphi(x)$ от наблюдений. Если распределение представлено в этом виде, то $\varphi(x)$ есть эффективная оценка для своего математического ожидания $\mathrm{M}\varphi(\xi) = f(a)$. Заметим, что нормальное распределение, распределения Пуассона, биномиальное, геометрическое, гамма являются экспонентными семействами распределений.