Лекция 16

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

ЦЕЛЬ ЛЕКЦИИ: ввести понятие доверительной вероятности и доверительного интервала, получить интервальные оценки математического ожидания и дисперсии.

Точность точечных оценок характеризуется их дисперсией. При этом отсутствуют сведения о том, насколько близки полученные оценки истинным значениям параметров. В ряде задач требуется не только найти для параметра a подходящее численное значение, но и оценить его точность и надежность. Необходимо узнать, к каким ошибкам может привести замена параметра a его точечной оценкой $\hat{a}$, и с какой степенью уверенности следует ожидать, что эти ошибки не выйдут за известные пределы.

Такие задачи особенно актуальны при малом числе опытов n, когда точечная оценка $\hat{a}$ в значительной степени случайна и приближенная замена a на $\hat{a}$ может привести к значительным ошибкам.

Более полный и надежный способ оценивания параметров распределений заключается в определении не единственного точечного значения, а интервала, который с заданной вероятностью накрывает истинное значение оцениваемого параметра.

Пусть по результатам n опытов получена несмещенная оценка $\hat{a} = u(x_1, x_2, \ldots, x_n)$ параметра a. Необходимо оценить возможную ошибку. Выбирается некоторая достаточно большая вероятность $\gamma = 1 - \alpha$ (например $\gamma = 0{,}95;\ 0{,}99;\ 0{,}9973$), такая, что событие с этой вероятностью можно считать практически достоверным событием, и находится такое значение ε, для которого

$$P\{|\hat{a} - a| < \varepsilon\} = \gamma . \tag{8.15}$$

В этом случае диапазон практически возможных значений ошибки, возникающей при замене a на $\hat{a}$, будет $\pm\varepsilon$, а большие по абсолютной величине ошибки будут появляться лишь с малой вероятностью α.

Выражение (8.15) означает, что с вероятностью $1 - \alpha$ неизвестное значение параметра a попадет в интервал

$$I_\alpha = (\hat{a} - \varepsilon;\ \hat{a} + \varepsilon). \tag{8.16}$$

Вероятность $\gamma = 1 - \alpha$ называется доверительной вероятностью, а интервал I_α, накрывающий с вероятностью γ истинное значение параметра, называется доверительным интервалом. Заметим, что неправильно говорить, что значение параметра лежит внутри доверительного интервала с вероятностью γ. Используемая формулировка (накрывает) означает, что хотя оцениваемый параметр и неизвестен, но он имеет постоянное значение и, следовательно, не имеет разброса, поскольку это не случайная величина.

Задача определения доверительного интервала может быть решена только тогда, когда удается найти закон распределения случайной величины $\hat{a}$. В общем случае этот закон зависит от закона распределения случайной величины X и, следовательно, и от его неизвестных параметров (в частности, и от самого оцениваемого параметра). Однако иногда удается перейти при получении оценки $\hat{a}$ к таким функциям опытных данных, закон распределения которых зависит только от величины n и закона распределения случайной величины X и не зависит от неизвестных параметров.

Пусть проведено n независимых испытаний над случайной величиной X, числовые характеристики которой – математическое ожидание m и дисперсия D – неизвестны. Для этих параметров получены точечные оценки:

$$m^* = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; D^* = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m^*)^2 . \tag{8.17}$$

Требуется найти доверительный интервал I_α, соответствующий доверительной вероятности $1-\alpha$, для математического ожидания m случайной величины X.

Так как случайная величина m^* представляет собой сумму n независимых и одинаково распределенных случайных величин x_i, то согласно центральной предельной теореме при достаточно больших n (на практике порядка 10÷20) ее закон распределения близок к нормальному. Таким образом получаем, что случайная величина m^* распределена по нормальному закону с математическим ожиданием m и дисперсией D/n (см. (7.3–7.4)). Если величина дисперсии D неизвестна, то в качестве ее оценки можно использовать D^*. В этом случае найдем такое ε_α, для которого

$$P\{|m^* - m| < \varepsilon_\alpha\} = \gamma = 1 - \alpha .$$

При использовании формулы (4.37) получаем

$$P\{|m^* - m| < \varepsilon_\alpha\} = 2\Phi\left(\frac{\varepsilon_\alpha}{\sigma_{m^*}}\right),$$

где $\sigma_{m^*} = \sqrt{D^*/n}$ – среднее квадратичное отклонение оценки m^*.

Из уравнения

$$2\Phi\left(\frac{\varepsilon_\alpha}{\sigma_{m^*}}\right) = 1 - \alpha$$

находим значение ε_α:

$$\varepsilon_\alpha = \sigma_{m^*}\Phi^{-1}\left(\frac{1-\alpha}{2}\right) = \sigma_{m^*}u_{\alpha/2}, \qquad (8.18)$$

где $\Phi^{-1}(x)$ – функция, обратная $\Phi(x)$, $u_{\alpha/2}$ – квантиль порядка $1-\alpha$ стандартного нормального распределения.

Таким образом, приближенно решена задача построения доверительного интервала в виде

$$I_\alpha = (m^* - \varepsilon_\alpha; m^* + \varepsilon_\alpha),$$

где ε_α определяется формулой (8.18).

Чтобы избежать при вычислении ε_α обратного интерполирования в таблицах функции $\Phi(x)$, обычно составляется небольшая таблица, в которой приводятся значения квантилей $u_{\alpha/2}$ в зависимости от наиболее часто используемых значений доверительной вероятности $1-\alpha$ (табл. 8.4).

Таблица 8.4

$1-\alpha$	$u_{\alpha/2}$
0,9	1,643
0,95	1,960
0,99	2,576
0,9973	3,000
0,999	3,290

Величина $u_{\alpha/2}$ определяет для нормального закона распределения число средних квадратичных отклонений, которое нужно отложить вправо и влево от центра рассеивания для того, чтобы вероятность попадания на этот участок была равна $1-\alpha$.

С использованием величины $u_{\alpha/2}$ доверительный интервал будет иметь вид

$$I_\alpha = (m^* - u_{\alpha/2}\sigma_{m^*}; m^* + u_{\alpha/2}\sigma_{m^*}).$$

Интервальные оценки математического ожидания и дисперсии нормальных случайных величин

Для случайной величины X, имеющей гауссово распределение, найдены точные методы построения доверительных интервалов оценок математического ожидания и дисперсии.

Если случайная величина X распределена нормально с математическим ожиданием m и дисперсией $D = \sigma^2$, то случайная величина

$$\chi^2 = \frac{(n-1)D^*}{D} \qquad (8.19)$$

имеет χ^2 распределение с $n-1$ степенями свободы, а случайная величина

$$T = \sqrt{n}\,\frac{m^* - m}{\sqrt{D^*}} \qquad (8.20)$$

подчиняется закону распределения Стьюдента с $r = n-1$ степенями свободы.

В формулах (8.19–8.20) m^* и D^* – точечные оценки математического ожидания и дисперсии в соответствии с (8.17).

Для обоих неизвестных параметров m и D необходимо построить доверительные интервалы.

Для математического ожидания величину ε_α (половину длины доверительного интервала) выбираем из условия

$$P\{|m^* - m| < \varepsilon_\alpha\} = 1 - \alpha. \qquad (8.21)$$

В левой части выражения (8.21) перейдем от случайной величины m^* к величине T, распределенной по закону Стьюдента. Для этого умножим обе части неравенства $\left|m^* - m\right| < \varepsilon_\alpha$ на положительную величину $\sqrt{n}/\sqrt{D^*}$ и получим

$$P\left\{\frac{\sqrt{n}\left|m^*-m\right|}{\sqrt{D^*}}<\frac{\varepsilon_\alpha}{\sqrt{D^*/n}}\right\}=1-\alpha\,,$$

а при использовании (8.20)

$$P\left\{\left|T\right|<\frac{\varepsilon_\alpha}{\sqrt{D^*/n}}\right\}=P\{\left|T\right|<t_{\alpha/2}\}=1-\alpha\,,$$

где величину $t_{\alpha/2}=\varepsilon_\alpha \Big/ \sqrt{D^*/n}$ находим из условия

$$P\{\left|T\right|<t_{\alpha/2}\}=\int_{-t_{\alpha/2}}^{t_{\alpha/2}} f_{n-1}(t)dt=1-\alpha \text{ или } P\{\left|T\right|>t_{\alpha/2}\}=\alpha/2\,.$$

По таблице процентных точек распределения Стьюдента (прил. 4) находим значение $t_{\alpha/2}=t_{n-1,\alpha/2}$ и получаем

$$\varepsilon_\alpha=t_{n-1,\alpha/2}\sqrt{D^*/n}\,,$$

и соответственно доверительный интервал оценки математического ожидания будет иметь вид

$$I_\alpha=\left(m^*-t_{n-1,\alpha/2}\sqrt{D^*/n}\,;\,m^*+t_{n-1,\alpha/2}\sqrt{D^*/n}\right). \tag{8.22}$$

Для нахождения доверительного интервала оценки дисперсии выразим случайную величину D^* через величину χ^2 в соответствии с (8.19):

$$D^*=\chi^2\frac{D}{n-1}\,.$$

Знание закона распределения случайной величины χ^2 позволяет найти доверительный интервал, в который эта величина попадает с вероятностью $1-\alpha$. Поскольку распределение χ^2 асимметрично (см. рис. 8.8), брать интервал I_α симметричным, как для нормального распределения или распределения Стьюдента, неправомерно. Поэтому доверительный интервал строят так, чтобы площади под кривой распределения от 0 до z_1 и от z_2 до бесконечности были равны $\alpha/2$:

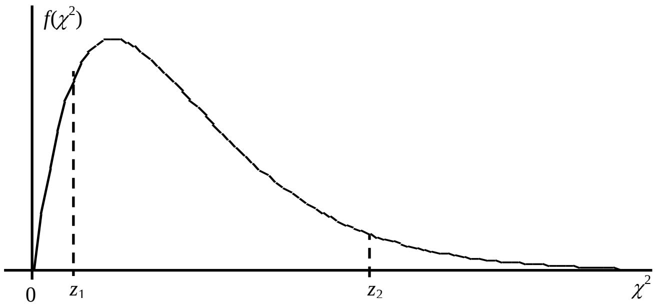

Рис. 8.8. Доверительный интервал распределения χ^2

$$P\{\chi^2_{n-1} < z_1\} = \int_0^{z_1} f_{n-1}(\chi^2)d\chi^2 = \alpha/2; \tag{8.23}$$

$$P\{\chi^2_{n-1} > z_2\} = \int_{z_2}^{\infty} f_{n-1}(\chi^2)d\chi^2 = \alpha/2. \tag{8.24}$$

Для интеграла (8.24) при заданном $1-\alpha$ по таблице процентных точек χ^2 распределения (прил. 3) находят $z_2 = \chi^2_{n-1,\alpha/2}$. Для получения z_1 перепишем выражение (8.23) в виде

$$P\{\chi^2_{n-1} < z_1\} = 1 - \int_{z_1}^{\infty} f_{n-1}(\chi^2)d\chi^2 = 1-\alpha/2,$$

откуда

$$z_1 = \chi^2_{n-1,1-\alpha/2}.$$

Таким образом, получаем для случая неизвестного математического ожидания

$$P\{\chi^2_{n-1,1-\alpha/2} \le \chi^2_{n-1} \le \chi^2_{n-1,\alpha/2}\} = P\left\{\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)D^*}{D} \le \chi^2_{n-1,\alpha/2}\right\} =$$

$$= P\left\{\frac{(n-1)D^*}{\chi^2_{n-1,\alpha/2}} \le D \le \frac{(n-1)D^*}{\chi^2_{n-1,1-\alpha/2}}\right\},$$

а доверительный интервал

$$I_\alpha = \left(\frac{(n-1)D^*}{\chi^2_{n-1,\alpha/2}}; \frac{(n-1)D^*}{\chi^2_{n-1,1-\alpha/2}} \right) \tag{8.25}$$

накрывает неизвестную дисперсию с заданной вероятностью $1-\alpha$.

<u>Пример</u>. Проведено $n = 25$ независимых измерений случайной величины X, имеющей нормальное распределение. Получены следующие результаты: 20, 21, 21, 25, 19, 22, 23, 23, 18, 21, 21, 17, 18, 24, 20, 22, 21, 19, 19, 22, 18, 23, 22, 18, 20. Необходимо определить 90 %-ные доверительные интервальные оценки математического ожидания и дисперсии измеренной случайной величины.

Точечные оценки математического ожидания и дисперсии:

$$m^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{25}\sum_{i=1}^{25} x_i = 20{,}68\,;$$

$$D^* = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m^*)^2 = \frac{1}{24}\sum_{i=1}^{25}(x_i - m^*)^2 = 4{,}39\,.$$

По таблице процентных точек t-распределения Стьюдента для $n = 25$ и $1-\alpha = 0{,}9$ (прил. 4) находим, что $t_{\alpha/2} = \; = t_{24;0{,}05} = 1{,}711$. Поэтому в соответствии с (8.22) получаем интервальную оценку математического ожидания в виде

$$I_\alpha = \left(m^* - 1{,}711\sqrt{D^*/25};\; m^* + 1{,}711\sqrt{D^*/25} \right) = (19{,}96;\, 21{,}40)\,.$$

По таблице процентных точек распределения χ^2 для $n = 25$ и $1-\alpha = 0{,}9$ (прил. 3) находим, что $\chi^2_{n-1;\alpha/2} = \chi^2_{24;0{,}05} = 36{,}42$ и $\chi^2_{n-1;1-\alpha/2} = \; = \chi^2_{24;0{,}95} = 13{,}85$. Таким образом, согласно (8.25) интервальная оценка дисперсии гауссовой случайной величины X будет иметь вид

$$I_\alpha = \left(\frac{24D^*}{36{,}42}; \frac{24D^*}{13{,}85} \right) = (2{,}895;\, 7{,}613)\,.$$