

# Лекция 5

Доверительные интервалы и  
доверительная вероятность.

Распределение Стьюдента

# ***Точечная оценка***

***Точечной оценкой*** называют оценку, которая определяется одним числом. При малом количестве обрабатываемых измерений  $n$  точечная оценка может значительно отличаться от оцениваемого параметра. Поэтому при небольшом объёме выборки необходимо рассмотреть надёжность этой оценки, которую можно оценить ***неслучайным интервалом***, расположенным вокруг точечной оценки, в который результат измерения попадет с заданной доверительной вероятностью (обычно в измерениях называемой «надежностью» и обозначаемой  $\alpha$ ).

Оценка  $a^*(x_1, x_2, \dots, x_n)$  называется **состоятельной**, если с увеличением объема выборки  $n$  она стремится (по вероятности) к оцениваемому параметру  $a$ .

Оценка  $a^*(x_1, x_2, \dots, x_n)$  называется **несмещенной**, если ее математическое ожидание при любом объеме выборки равно оцениваемому параметру  $a$ , т. е.  $M[a^*] = a$ .

Выборочные параметры распределения, определяемые по серии измерений, являются *случайными величинами*, следовательно, и их отклонения от генеральных параметров также будут случайными. Оценка этих отклонений носит вероятностный характер — при статистическом анализе можно лишь указать *вероятность той или иной погрешности*.

Пусть для генерального параметра  $a$  получена из опыта несмещенная оценка  $a^*$ . Назначим достаточно большую вероятность  $\beta$  (такую, что событие с вероятностью  $\beta$  можно считать практически достоверным) и найдем такое значение  $\varepsilon_\beta = f(\beta)$ , для которого

$$P(a^* - a \leq \varepsilon_\beta) = \beta$$

Диапазон практически возможных значений ошибки, возникающей при замене  $a$  на  $a^*$ , будет  $\pm\varepsilon_\beta$ .

Большие по абсолютной величине ошибки будут появляться только с малой вероятностью

$$p = 1 - \beta$$

называемой уровнем значимости. Иначе выражение можно интерпретировать как вероятность того, что истинное значение параметра  $a$  лежит в пределах

$$a^* - \varepsilon_\beta \leq a \leq a^* + \varepsilon_\beta$$

Вероятность  $\beta$  называется доверительной вероятностью и характеризует надежность полученной оценки.

Доверительный интервал при данной доверительной вероятности определяет точность оценки. Величина доверительного интервала зависит от доверительной вероятности, с которой гарантируется нахождение параметра  $a$  внутри доверительного интервала: чем больше величина  $\beta$ , тем больше интервал  $I_\beta$  (и величина  $\varepsilon_\beta$ ). Увеличение числа опытов проявляется в сокращении доверительного интервала при постоянной доверительной вероятности или в повышении доверительной вероятности при сохранении доверительного интервала.

# построение доверительного интервала

- Рассмотрим построение доверительного интервала для математического ожидания нормально распределенной случайной величины  $X$  с известным генеральным стандартом  $\sigma$  по выборке объемом  $n$ . Наилучшей оценкой для математического ожидания  $m_x$  является среднее выборки  $\bar{x}$  со стандартным отклонением среднего

$$\sigma(\bar{x}) = \frac{\sigma}{n}$$

Используя функцию Лапласа, получаем

$$P(|\bar{x} - m_x| \leq \varepsilon_\beta) = \beta = 2\Phi\left(\frac{\varepsilon_\beta}{\sigma(\bar{x})}\right)$$



Задавшись доверительной вероятностью  $\beta$ , определим по таблице функции Лапласа величину

$$k_\beta = \frac{\varepsilon_\beta}{\sigma(x)}$$

Тогда доверительный интервал для математического ожидания принимает вид

$$x - k_\beta \sigma(x) \leq m_x \leq x + k_\beta \sigma(x)$$

или

$$\bar{x} - k_\beta \frac{\sigma}{\sqrt{n}} \leq m_x \leq \bar{x} + k_\beta \frac{\sigma}{\sqrt{n}}$$

уменьшение доверительного интервала обратно пропорционально корню квадратному из числа опытов.

При отсутствии грубых и систематических ошибок математическое ожидание случайной величины совпадает с истинным результатом наблюдений. Легче всего оценить математическое ожидание при известной дисперсии генеральной совокупности. Однако значение  $\sigma^2$  нельзя получить из наблюдений, ее можно только оценить при помощи выборочной дисперсии  $s^2$ . Ошибка от этой замены будет тем меньше, чем больше объем выборки  $n$ . На практике эту погрешность не учитывают при  $n \leq 50$  и в формуле для доверительного интервала генеральный параметр  $\sigma$  заменяют выборочным стандартом.

При небольших объемах выборок для построения доверительного интервала математического ожидания используют распределение Стюдента, или  $t$ -распределение. Распределение Стюдента имеет величина  $t$

$$t = \frac{\bar{x} - m_x}{s_x} \sqrt{n}$$

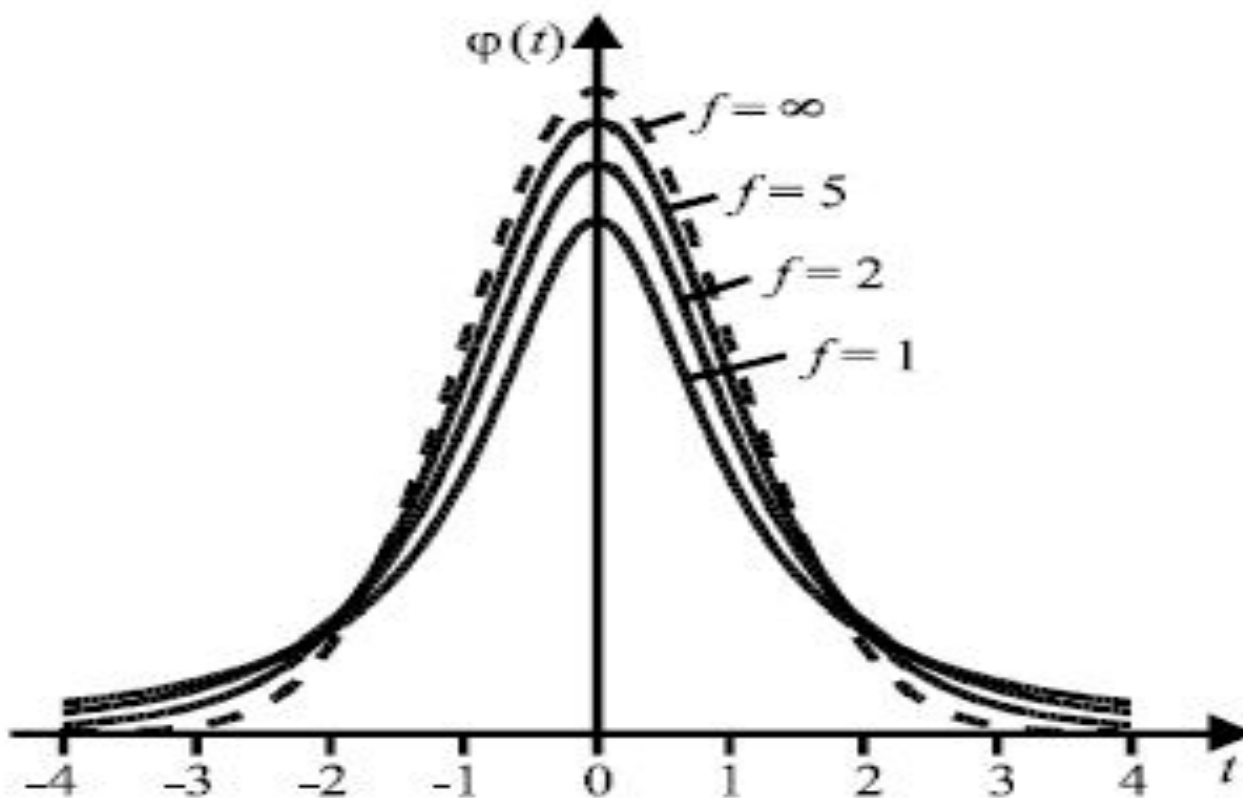
с плотностью вероятности

$$\varphi(t) = \frac{1}{\sqrt{2\pi f}} \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \left(1 + \frac{t^2}{f}\right)^{-\left(\frac{f+1}{2}\right)}, \quad -\infty < t$$

$< +\infty$

где  $\Gamma(f)$  — гамма-функция Эйлера:

Распределение Стьюдента зависит только от числа степеней свободы  $f$ , с которым определена выборочная дисперсия. На рис. приведены графики плотности  $t$ -распределения для нескольких чисел свободы  $f$  и нормальная кривая.



Плотность распределения  
Стьюдента

Кривые  $t$ -распределения по своей форме напоминают нормальную кривую, но при малых  $f$  они медленнее сближаются с осью абсцисс при  $t \rightarrow \infty$ . При  $f \rightarrow \infty$   $s^2 \rightarrow \sigma^2$ , поэтому распределение Стьюдента сближается (в пределе соответствует) с нормальным распределением. Вероятность того, что случайная величина попадет в интервал  $(t_{p/2}; t_{1-p/2})$ , определяется выражением

$$P(t_{p/2}; t_{1-p/2}) = 1 - p = \beta$$

Распределение Стьюдента симметрично относительно нуля, поэтому

$$t_{p/2} = -t_{1-p/2}$$