2.10. Определение доверительного интервала

Пусть θ – некоторый неизвестный параметр распределения. По выборке $X_1, \dots, X_n$ из данного распределения построим интервальную оценку параметра θ распределения, то есть найдем такой интервал, внутри которого с заданной (высокой) вероятностью $1-\alpha$ находится истинное значение неизвестного параметра θ. Указанную вероятность $1-\alpha$ называют доверительной вероятностью, а величину α – уровнем значимости.

В качестве значений доверительной вероятности обычно выбирают величины 0,9, 0,95, 0,99, достаточно близкие к 1. В каждом конкретном случае выбор величины доверительной вероятности определяется спецификой решаемой практической задачи.

Итак, пусть $X_1, \dots, X_n$ – выборка из данного распределения и задана величина доверительной вероятности $1-\alpha$.

def| Интервал $(\vartheta_1;\vartheta_2)$ называют доверительным интервалом для параметра θ, отвечающим доверительной вероятности $1-\alpha$, если его границы $\vartheta_1 = \vartheta_1(X_1, \dots, X_n)$ и $\vartheta_2 = \vartheta_2(X_1, \dots, X_n)$ – две статистики такие, что верно равенство: $P(\vartheta_1 < \theta < \vartheta_2) = 1-\alpha$.

Заметим, что границы доверительного интервала ϑ_1 и ϑ_2 – случайные величины (функции выборки $X_1, \dots, X_n$), параметр θ – неслучайная величина, так что интервал $(\vartheta_1;\vartheta_2)$ “накрывает“ величину θ с вероятностью $1-\alpha$ (соответственно, “не накрывает“ с вероятностью α).

Длина интервала $\vartheta_2 - \vartheta_1$ характеризует точность, а доверительная вероятность $1-\alpha$ – надежность интервальной оценки. Очевидно, что точность и надежность взаимосвязаны: увеличение надежности приводит к уменьшению точности – увеличению длины интервала $(\vartheta_2 - \vartheta_1)$. Выбирая величину доверительной вероятности $1-\alpha$, принимают соглашение: считать события, вероятность которых $P \geq 1-\alpha$, – практически достоверными, а события, вероятность которых $P \leq \alpha$ – практически невозможными.

Практически достоверно
$\theta \in (\vartheta_1; \vartheta_2)$
(вероятность $1-\alpha$)

Практически невозможно
$\theta \notin (\vartheta_1; \vartheta_2)$
(вероятность α)

2.11. Основные этапы процедуры построения доверительных интервалов

Напомним, что границы доверительного интервала $\vartheta_1(X_1, \ldots, X_n)$ и $\vartheta_2(X_1, \ldots, X_n)$ – случайные величины (функции выборки $X_1, \ldots, X_n$). Результатом эксперимента (серии n независимых наблюдений данной случайной величины) является реализация выборки $x_1, x_2, \ldots, x_n$. Соответственно, значения статистик ϑ_1 и ϑ_2 в точке $(x_1, x_2, \ldots, x_n)$ – это *числа* $\vartheta_{1e} = \vartheta_1(x_1, x_2, \ldots, x_n)$ и $\vartheta_{2e} = \vartheta_2(x_1, x_2, \ldots, x_n)$.

Таким образом, будем различать, идет речь о *доверительном интервале*, границы которого по смыслу – случайные величины, или о *реализации доверительного интервала*, границами которого являются конкретные числа.

Для построения *реализации* доверительного интервала на основе данной *реализации* выборки $x_1, x_2, \ldots, x_n$ выполняют следующие действия:

(1) формулируют предположения о распределении и о выборке $X_1, \ldots, X_n$ (допущения, принимаемые при построении априорной теоретической модели).

(2) строят доверительный интервал, для чего:

а) выбирают значение доверительной вероятности $1-\alpha$ (или уровня значимости α), то есть принимают соглашение:

– *вероятность практически достоверного события* $= 1-\alpha$;

– *вероятность практически невозможного события* $= \alpha$.

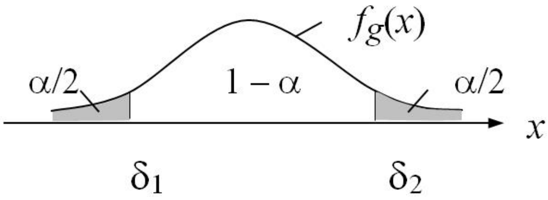
б) записывают вероятностное равенство:

$$P(\delta_1 < g(\theta; \hat{\theta}) < \delta_2) = \int_{\delta_1}^{\delta_2} f_g(x)\, dx = 1-\alpha,$$

где статистика g имеет известную (табулированную) плотность вероятности $f_g(x)$, θ – оцениваемый параметр, $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ – некоторая его оценка.

Существует бесконечное множество значений величин δ_1 и δ_2, обеспечивающих справедливость указанного равенства, однако, если использовать дополнительное условие

$P\,(g(\theta;\hat{\theta}) < \delta_1\,) = P\,(g(\theta;\hat{\theta}) > \delta_2) = \alpha/2$, то решение ($\delta_1$; δ_2) будет единственным:



в) преобразуют неравенство $\delta_1 < g(\theta;\hat{\theta}) < \delta_2$ к виду $\vartheta_1 < \theta < \vartheta_2$, где $\vartheta_1 = \vartheta_1\,(X_1, \ldots, X_n)$, $\vartheta_2 = \vartheta_2\,(X_1, \ldots, X_n)$ – функции выборки, тогда

$$P(\vartheta_1 < \theta < \vartheta_2) \equiv 1-\alpha.$$

(3) проводят эксперимент – получают конкретную реализацию выборки $x_1, x_2, \ldots, x_n$.

(4) вычисляют значения $\vartheta_{1e} = \vartheta_1\,(x_1, x_2, \ldots, x_n)$ и $\vartheta_{2e} = \vartheta_2\,(x_1, x_2, \ldots, x_n)$.

В результате перечисленных действий (1) – (4) получают реализацию доверительного интервала – числовой интервал (ϑ_{1e} ; ϑ_{2e}).

Степень уверенности в том, что полученный интервал (ϑ_{1e} ; ϑ_{2e}) в действительности содержит неизвестный параметр θ выражается *выбранной априори* величиной доверительной вероятности $1-\alpha$ (вероятностью практически достоверного события). Иными словами, априори допускается, что утверждение “данная реализация доверительного интервала (ϑ_{1e} ; ϑ_{2e}) содержит оцениваемый параметр θ” может оказаться ошибочным, однако, число таких случаев мало́ и может наблюдаться лишь в $\alpha \cdot 100\%$ общего числа случаев; при этом приемлемая доля указанных ошибок выражается уровнем значимости α – вероятностью практически невозможного события.

2.12. Доверительный интервал для математического ожидания нормального распределения при известной дисперсии

Пример 1_ди

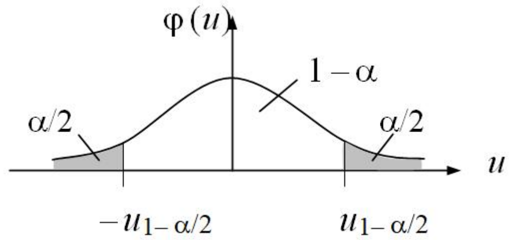
Пусть $X_1, \dots, X_n$ – выборка из нормального распределения $N(m;\sigma)$. Предполагается, что параметр σ известен: $\sigma = \sigma_o$ (например, этот параметр определен в результате специальных многократных измерений).

При указанных предположениях справедливо: $\bar{X} \sim N(m; \sigma_o / \sqrt{n})$

или иначе, $\dfrac{\bar{X} - m}{\sigma_o / \sqrt{n}} \sim N(0; 1)$.

Построим доверительный интервал для математического ожидания m. Зададим величину доверительной вероятности $1-\alpha$ и запишем вероятностное равенство $P\left(\left|\dfrac{\bar{X} - m}{\sigma_o / \sqrt{n}}\right| < u\right) = 1-\alpha$, откуда:

$u = u_{1-\alpha/2}$ – квантиль порядка $1-\alpha/2$ стандартного нормального распределения



Далее, $P\left(\left|\dfrac{\bar{X} - m}{\sigma_o / \sqrt{n}}\right| < u_{1-\alpha/2}\right) \equiv 1-\alpha$, откуда

$$P\left(\bar{X} - \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2} < m < \bar{X} + \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2}\right) \equiv 1-\alpha.$$

Таким образом, искомый доверительный интервал, отвечающий доверительной вероятности $1-\alpha$ имеет вид:

$$\left(\bar{X} - \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2}; \bar{X} + \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2}\right)$$

Заметим, что для данной величины $1-\alpha$ доверительной вероятности длина этого интервала равна $d = 2\dfrac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2}$.

При данном объеме выборки n длина d постоянна, от выборки к выборке меняется только положение центра интервала $\bar{X}$.

Замечание

Проиллюстрируем связь точности и надежности интервальной оценки при фиксированных значениях n и σ_o:

$$1-\alpha=0{,}9 \qquad u_{1-\alpha/2}=u_{0,95}=1{,}65 \qquad d=3{,}3\frac{\sigma_o}{\sqrt{n}};$$

$$1-\alpha=0{,}95 \qquad u_{1-\alpha/2}=u_{0,975}=1{,}96 \qquad d=3{,}92\frac{\sigma_o}{\sqrt{n}};$$

$$1-\alpha=0{,}99 \qquad u_{1-\alpha/2}=u_{0,995}=2{,}58 \qquad d=5{,}16\frac{\sigma_o}{\sqrt{n}}.$$

Видим, что при данном объеме выборки n с ростом надежности оценки $1-\alpha$ ее точность убывает (длина d соответствующего интервала растет). Таким образом, как уже упоминалось, плата за повышение надежности – уменьшение точности интервальной оценки.

Пусть заданы $\varepsilon>0$ и $1-\alpha$ – величины, характеризующие соответственно точность и надежность оценки. Найдем объем выборки, достаточный для обеспечения одновременно заданных значений точности и надежности оценки.

Из условия $d \le 2\varepsilon$ получаем: $n \ge (\dfrac{\sigma_o}{\varepsilon} u_{1-\alpha/2})^2$.

Пример 1_ди (доверительный интервал для m при известном σ, $1-\alpha = 0{,}95$)

Точность прибора известна (в паспорте прибора указано σ_o = 0,02). С помощью этого прибора проведено n независимых повторных измерений ($n = 25$) некоторой физической величины. По результатам измерений $x_1, x_2, \ldots, x_{25}$ вычислено среднее выборочное, оказавшееся равным $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = 2,42. Считая числа $x_1, x_2, \ldots, x_{25}$ реализацией выборки из нормального распределения $N(m; \sigma_o)$, где $\sigma_o = 0{,}02$, найти реализацию доверительного интервала для математического ожидания m; доверительная вероятность $1-\alpha = 0{,}95$.

<u>Решение</u>

Выражение для реализации доверительного интервала для математического ожидания m нормального распределения (при $\sigma = \sigma_o$), отвечающего доверительной вероятности $1-\alpha$, имеет вид:

$$\left(\bar{x} - \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2};\ \bar{x} + \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha/2}\right).$$

Квантиль порядка $1-\alpha/2$ стандартного нормального распределения: $u_{1-\alpha/2} = u_{0{,}975} = 1{,}96$. Подставляя $\bar{x} = 2{,}42$, $\sigma_o = 0{,}02$, $n = 25$ в выражение для реализации доверительного интервала, получаем:

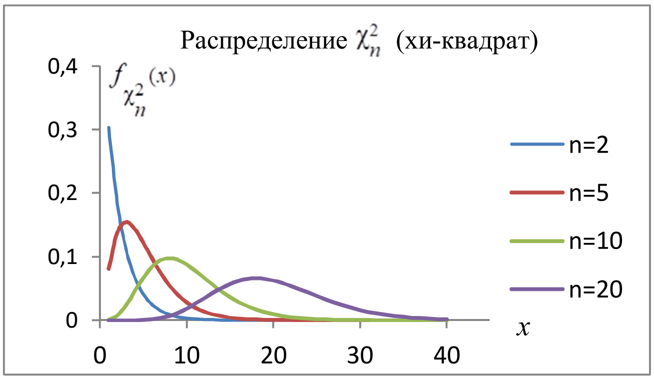
$$(2{,}420 - 0{,}008;\ 2{,}420 + 0{,}008) = (2{,}412;\ 2{,}428).$$

2.13. Распределение χ^2, распределение Стьюдента, лемма Фишера

Распределение χ^2

$\underline{def|}$ Сумма квадратов n независимых стандартных нормальных случайных величин $\sum_{i=1}^{n} X_i^2 = \chi_n^2$ $(X_i \sim N(0;1))$ называется случайной величиной χ_n^2 (с n степенями свободы).

Плотность вероятности распределения χ_n^2 табулирована, ее график имеет вид:



Для распределения χ_n^2 имеют место следующие соотношения: $M[\chi_n^2] = n$, $D[\chi_n^2] = 2n$, $Mo[\chi_n^2] = n - 2$. Заметим, что с ростом n форма кривой $f_{\chi_n^2}(x)$ становится более симметричной, а ее максимум смещается вправо.

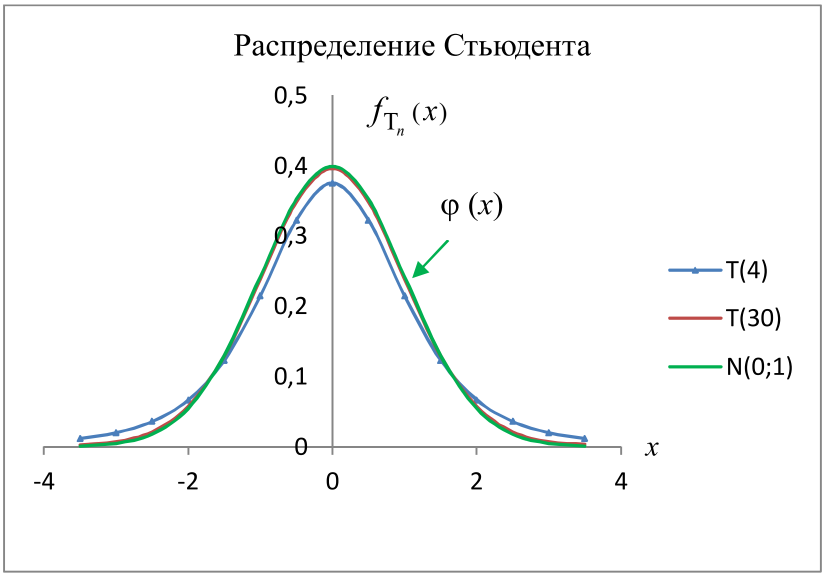
Заметим также, что в силу центральной предельной теоремы центрированная и нормированная случайная величина $\frac{\chi_n^2 - n}{\sqrt{2n}} \underset{n\to+\infty}{\sim} N(0;1)$ асимптотически нормальна, поэтому в таблицах для распределения χ_n^2 приводятся квантили только для $n \le 30$.

<u>Распределение Стьюдента</u>

<u>*def*</u>| Пусть случайные величины $U \sim N(0;1)$ и χ_n^2 – независимы. Отношением Стьюдента называется случайная величина

$$\mathrm{T}_n = \frac{U}{\sqrt{\frac{\chi_n^2}{n}}} \quad (t - \text{отношение}).$$

Плотность вероятности распределения Стьюдента табулирована, ее график имеет вид, представленный на рисунке:



Распределение $f_{\mathrm{T}_n}(x)$ симметрично, $\mathrm{M}[\mathrm{T}_n]=0$ и имеет место асимптотическое свойство: $\mathrm{T}_n \underset{n\to+\infty}{\sim} N(0;1)$.

При малых значениях n распределение Стьюдента заметно отличается от стандартного нормального распределения, однако при $n > 30$ эти распределения близки.

T| *<u>Лемма Фишера</u> (о распределении $\bar{X}$ и S^2 для выборки из нормального распределения)*

Пусть $X_1, \ldots, X_n$ – выборка из нормального распределения $N(m;\sigma)$ тогда:

- выборочное среднее $\bar{X}$ и выборочная дисперсия S^2 (или исправленная выборочная дисперсия S^{*2}) – взаимно независимы;
- выборочное среднее $\bar{X}$ подчиняется нормальному распределению: $\bar{X} \sim N(m; \frac{\sigma}{\sqrt{n}})$;
- случайная величина $\frac{nS^2}{\sigma^2}$ (или $\frac{(n-1)\,S^{*2}}{\sigma^2}$) распределена по закону χ^2_{n-1} (с $n-1$ степенью свободы).

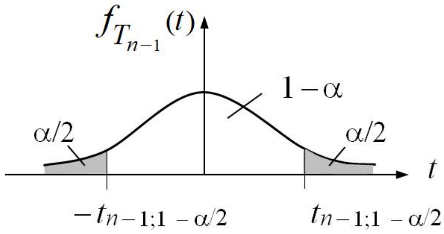
Из леммы Фишера следует независимость $\bar{X}$ и $\frac{nS^2}{\sigma^2}$, а также $\bar{X}$ и $\frac{(n-1)\,S^{*2}}{\sigma^2}$.

2.14. Доверительный интервал для математического ожидания нормального распределения при неизвестной дисперсии Примеры 2_ди и 3_ди

Пусть $X_1, \ldots, X_n$ – выборка из нормального распределения $N(m; \sigma)$, параметры которого m и σ *неизвестны*. Найдем интервальную оценку параметра m, отвечающую заданной величине доверительной вероятности $1-\alpha$.

Согласно лемме Фишера статистики $U = \dfrac{\bar{X} - m}{\sigma / \sqrt{n}} \sim N(0; 1)$ и $\dfrac{(n-1)\, S^{*2}}{\sigma^2} = \chi^2_{n-1}$ – независимы. Запишем отношение Стьюдента

$$T_{n-1} = \frac{U}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}} = \frac{\bar{X} - m}{S^* / \sqrt{n}}.$$



Из вероятностного равенства

$$P\,(|T_{n-1}| < t) = 1 - \alpha$$

имеем: $t = t_{n-1;1-\alpha/2}$ – квантиль порядка $1 - \alpha/2$ распределения Стьюдента.

Далее, $P\left(\left|\dfrac{\bar{X} - m}{S^* / \sqrt{n}}\right| < t_{n-1;\,1-\alpha/2}\right) \equiv 1 - \alpha$, откуда

$$P\,(\bar{X} - \frac{S^*}{\sqrt{n}}\, t_{n-1;1-\alpha/2} < m < \bar{X} + \frac{S^*}{\sqrt{n}}\, t_{n-1;1-\alpha/2}) \equiv 1 - \alpha.$$

Таким образом, искомый доверительный интервал, отвечающий доверительной вероятности $1-\alpha$, имеет вид:

$$(\bar{X} - \frac{S^*}{\sqrt{n}}\, t_{n-1;1-\alpha/2}\,;\ \bar{X} + \frac{S^*}{\sqrt{n}}\, t_{n-1;1-\alpha/2}).$$

Заметим, что положение центра этого интервала, как и его длина – случайные величины.

Пример 2_ди (доверительный интервал для m при неизвестном σ, $1-\alpha = 0{,}6827$)

По результатам пяти измерений ($n=5$) вычислено выборочное среднее $\bar{x}$ и исправленная выборочная дисперсия s^{*2}, соответственно: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 4$ и $s^{*2} = \frac{1}{4}\sum_{i=1}^{5}(x_i - \bar{x})^2 = 10/4$.

Считая числа $x_1, x_2, \ldots, x_5$ реализацией выборки из нормального распределения $N(m;\sigma)$, где σ неизвестно, найти реализацию доверительного интервала для математического ожидания m, приняв величину доверительной вероятности равной $1-\alpha = 0{,}6827$.

Решение

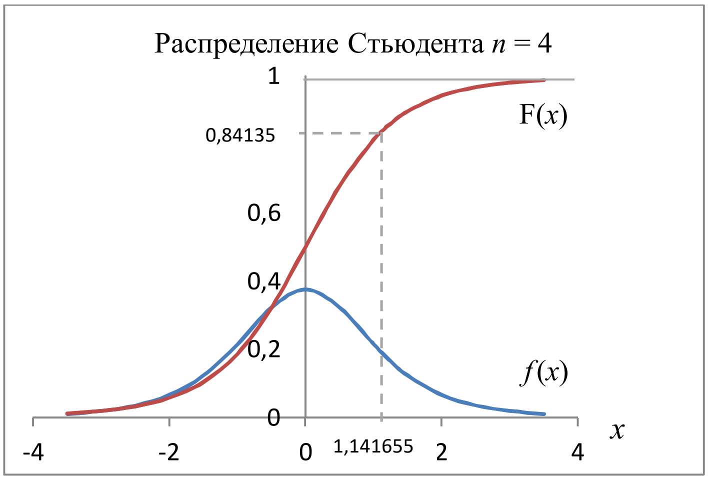
Реализация доверительного интервала, отвечающего доверительной вероятности $1-\alpha$, для математического ожидания m нормального распределения имеет вид:

$$\left(\bar{x} - \frac{s^*}{\sqrt{n}} t_{n-1;1-\alpha/2}\,;\, \bar{x} + \frac{s^*}{\sqrt{n}} t_{n-1;1-\alpha/2}\right)$$

Здесь $\frac{s^*}{\sqrt{n}}$ – стандартная ошибка, $\frac{s^*}{\sqrt{n}} t_{n-1;1-\alpha/2}$ – погрешность.

Правило записи реализации доверительного интервала (для $1-\alpha =$
$= 0{,}6827$):

1) округление значения погрешности – округляемая цифра отбрасывается, если она <5; если она ≥ 5 то в предыдущий разряд добавляется единица;
2) в погрешности оставляют одну значащую цифру, если она ≥ 4 и две, если первая из них <4;
3) в значении $\bar{x}$ оставляют последнюю значащую цифру в том же разряде, что и в погрешности;
4) общий множитель вида 10^k выносят за скобки.



Находим значение квантили $t_{n-1;1-\alpha/2} = t_{4;0,84135} = 1,141655$, вычисляем значение погрешности

$\dfrac{s^*}{\sqrt{5}} t_{4;\,0,84135} = \dfrac{1}{\sqrt{2}} \cdot 1,141655 = 0,8$ и получаем реализацию доверительного интервала: $(4,0 - 0,8; 4,0 + 0,8) = (3,2; 4,8)$.

Пример 3_ди (доверительный интервал для m при неизвестном σ, $1-\alpha = 0,95$)

Произведено $n = 19$ измерений $x_1, x_2, \ldots, x_{19}$ некоторой физической величины с помощью прибора, измеряющего с точностью до 0,1. Вычислено выборочное среднее $\overline{x} = \dfrac{1}{19}\sum_{i=1}^{19} x_i = 7,07$ и исправленная выборочная дисперсия $s^{*2} = \dfrac{1}{18}\sum_{i=1}^{19}(x_i - \overline{x})^2 = 0,041$.

Считая, что $x_1, x_2, \ldots, x_{19}$ – реализация выборки из нормального распределения $N(m;\sigma)$, где σ неизвестно, записать реализацию доверительного интервала для математического ожидания m при доверительной вероятности $1-\alpha = 0,95$.

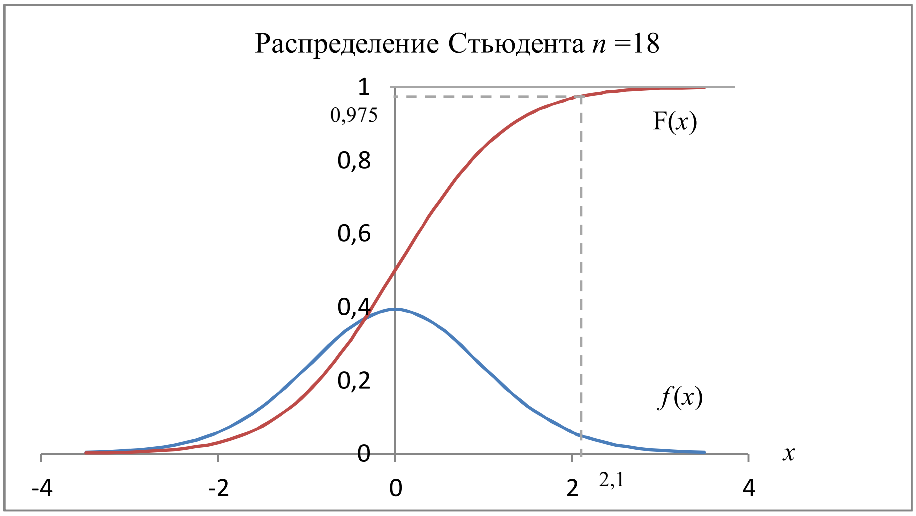
<u>Решение</u>

Правило записи реализации доверительного интервала для $1-\alpha = 0,95$:

1) выборочное среднее $\bar{x}$ вычисляется с точностью на порядок большей, чем точность измерений;
2) выборочное стандартное отклонение s^* (исправленное) – с точностью, на порядок большей, чем точность вычисления среднего;
3) правило округления: если округляемая цифра < 5, то она отбрасывается, если она ≥ 5 – в предыдущий разряд добавляется единица.

Реализация доверительного интервала, отвечающего доверительной вероятности $1-\alpha$, для математического ожидания m при неизвестном σ имеет вид:

$$(\bar{x} - \frac{s^*}{\sqrt{n}} t_{n-1;1-\alpha/2}; \bar{x} + \frac{s^*}{\sqrt{n}} t_{n-1;1-\alpha/2})$$



Находим значение квантили $t_{n-1;1-\alpha/2} = t_{18;\,0{,}975} = 2{,}10$, вычисляем значение погрешности $\frac{s^*}{\sqrt{19}} t_{18;\,0{,}975} = 0{,}098$ и получаем искомую реализацию доверительного интервала для m:

$$(7{,}07 - 0{,}098;\ 7{,}07 + 0{,}098) = (6{,}972;\ 7{,}168).$$

2.15. Доверительный интервал для дисперсии нормального распределения

Пусть $X_1, \dots, X_n$ – выборка из нормального распределения $N(m;\sigma)$. Параметры распределения m и σ неизвестны. Найдем интервальную оценку дисперсии распределения σ^2, отвечающую заданной величине доверительной вероятности $1-\alpha$.

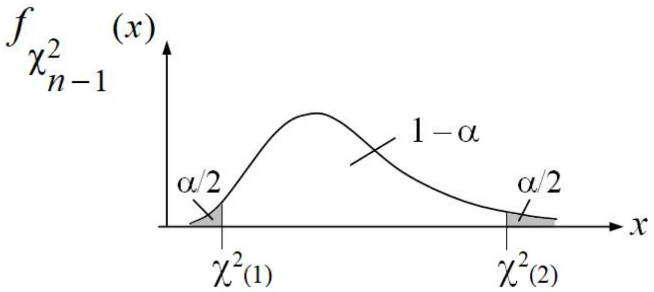
Известно (п. 2.13.), что статистика $\dfrac{(n-1)\,S^{*2}}{\sigma^2}$ распределена по закону χ^2_{n-1} (хи-квадрат с $n-1$ степенью свободы):

$$\frac{(n-1)\,S^{*2}}{\sigma^2} = \chi^2_{n-1}$$

Запишем равенство:

$$P\left(\chi^2_{(1)} < \frac{(n-1)\,S^{*2}}{\sigma^2} < \chi^2_{(2)}\right) = 1-\alpha,$$

здесь $\chi^2_{(1)}$ и $\chi^2_{(2)}$ – границы доверительного интервала, подлежащие определению. С учетом дополнительного условия $P(\chi^2_{n-1} < \chi^2_{(1)}) = P(\chi^2_{n-1} > \chi^2_{(2)}) = \alpha/2$ эти границы определяются единственным образом.



$\chi^2_{(1)} = \chi^2_{n-1;\,\alpha/2}$ и

$\chi^2_{(2)} = \chi^2_{n-1;\,1-\alpha/2}$ – квантили распределения случайной величины χ^2_{n-1} порядка $\alpha/2$ и $1-\alpha/2$, соответственно.

В итоге получаем:

$$P\left(\frac{(n-1)\,S^{*2}}{\chi^2_{(2)}} < \sigma^2 < \frac{(n-1)\,S^{*2}}{\chi^2_{(1)}}\right) \equiv 1-\alpha.$$

2.16. Приближенная интервальная оценка для математического ожидания произвольного распределения по выборке большого объема

Пусть $X_1, \dots, X_n$ – выборка из некоторого (*произвольного*) распределения, причем *объем выборки n достаточно велик*.

Введем обозначения: $MX_i = m$, $DX_i = \sigma^2$ $(i = 1, \dots, n)$, тогда $M\bar{X} = m$, $D\bar{X} = (\sigma/\sqrt{n}\,)^2$.

Согласно центральной предельной теореме имеем: $\dfrac{\bar{X} - m}{\sigma/\sqrt{n}} \sim N(0;1)$

при $n \to +\infty$. По условию n – велико, поэтому примем допущение, что указанная нормальность $\dfrac{\bar{X} - m}{\sigma/\sqrt{n}} \sim N(0;1)$ имеет место при данном объеме выборки n.

Зададимся величиной доверительной вероятности $1 - \alpha$ и запишем (формально): $P\left(\left|\dfrac{\bar{X} - m}{\sigma/\sqrt{n}}\right| < u\right) = 1 - \alpha$, откуда

$u = u_{1-\alpha/2}$ – квантиль порядка $1 - \alpha/2$ стандартного нормального распределения. Таким образом,

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} < m < \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right) \equiv 1 - \alpha.$$

Последнее равенство содержит *неизвестную величину* σ. Заменив σ ее несмещенной состоятельной оценкой S^*, получаем *приближенную* интервальную оценку для математического ожидания *m,* отвечающую доверительной вероятности $1 - \alpha$ при большом объеме выборки:

$$\left(\bar{X} - \frac{S^*}{\sqrt{n}} u_{1-\alpha/2} < m < \bar{X} + \frac{S^*}{\sqrt{n}} u_{1-\alpha/2}\right).$$

2.17. Приближенная интервальная оценка вероятности p в схеме Бернулли по выборке большого объема. Пример 4_ди

Пусть проводится n независимых испытаний, p – вероятность успеха в каждом испытании и $X_1,\dots,X_n$ – выборка из соответствующего распределения Бернулли, причем объем выборки n достаточно велик. Каждый элемент выборки – индикатор X_i появления успеха в i-м испытании. Случайная величина $X=\sum_{i=1}^{n} X_i$ – число успехов – подчиняется биномиальному распределению:

$$P(X=m)=C_n^m p^m (1-p)^{n-m} \quad (m=0,1,\dots,n).$$

Согласно теореме Муавра-Лапласа, центрированная и нормированная случайная величина $\frac{X-np}{\sqrt{npq}}=\frac{(X/n)-p}{\sqrt{pq/n}}$ является асимптотически нормальной, что можно символически записать так: $\frac{\hat{p}-p}{\sqrt{pq/n}} \sim N(0;1)$ при $n\to+\infty$. Здесь относительная частота $\hat{p}=\frac{X}{n}$ – оценка максимального правдоподобия вероятности p (несмещенная состоятельная асимптотически эффективная и асимптотически нормальная точечная оценка), $q=1-p$.

Это означает, что при достаточно больших значениях n распределение статистики $\frac{\hat{p}-p}{\sqrt{pq/n}}$ близко к стандартному нормальному. Будем считать, что нормальность распределения имеет место при данном (достаточно большом) объеме выборки n.

Зададимся величиной доверительной вероятности $1-\alpha$ и запишем:

$P\left(\left|\frac{\hat{p}-p}{\sqrt{pq/n}}\right|<u\right)=1-\alpha$, откуда $u=u_{1-\alpha/2}$ – квантиль порядка $1-\alpha/2$ стандартного нормального распределения.

$$\mathrm{P}\left(\left|\frac{\hat{p}-p}{\sqrt{pq/n}}\right|<u_{1-\alpha/2}\right)\equiv 1-\alpha,$$

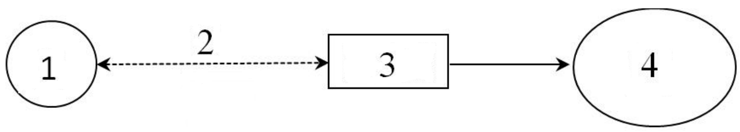
$$\mathrm{P}\left(\hat{p}-u_{1-\alpha/2}\sqrt{p(1-p)/n}<p<\hat{p}+u_{1-\alpha/2}\sqrt{p(1-p)/n}\right)\equiv 1-\alpha.$$

Заменив в последнем тождестве неизвестную величину p ее оценкой $\hat{p}$, получим выражение для приближенного (асимптотического) доверительного интервала, отвечающего доверительной вероятности $1-\alpha$:

$$\left(\hat{p}-u_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n};\ \hat{p}+u_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\right)$$

Отметим еще раз, что полученное выражение используют для нахождения *приближенной интервальной оценки вероятности* p *при достаточно больший значениях* n.

Пример 4_ди



Приемное устройство (3) охранной системы находится в режиме ожидания (дежурном режиме) и может получать сообщения (пакеты) фиксированной длительности $\Delta\tau$ от датчика (1) по радиоканалу (2). Если в каком-либо интервале времени $\Delta\tau$ от датчика поступает сообщение, то на исполнительное устройство (4) передается извещение “тревога”.

Под *ложной тревогой* понимают ошибочное формирование извещения “тревога” при условии, что сообщение от датчика отсутствует; ложная тревога обусловлена только наличием собственного шума в системе, в частности – в радиоканале.

Для оценки вероятности ложной тревоги осуществили наблюдение за работой системы в течение интервала времени, равного $10^4\Delta\tau$. Подсчитали частоту события “ложная тревога” (число интервалов $\Delta\tau$, таких, в которых было ошибочно сформировано извещение “тревога”) и его относительную частоту, оказавшуюся равной 0,006.

Задача: найти *реализацию* приближенной интервальной оценки вероятности ложной тревоги для доверительной вероятности $1-\alpha=0{,}95$.

Указание: Принять допущение, что вероятностной моделью эксперимента по оценке вероятности p ложной тревоги служит последовательность независимых испытаний (схема Бернулли), то есть, что выборка объема $n=10^4$ извлечена из распределения Бернулли с параметром p.

<u>Решение</u>

Выражение для реализации приближенного доверительного интервала для неизвестного параметра p, отвечающего доверительной вероятности $1-\alpha$, имеет вид:

$$\left(\hat{p}_e-u_{1-\alpha/2}\sqrt{\hat{p}_e(1-\hat{p}_e)/n};\ \hat{p}_e+u_{1-\alpha/2}\sqrt{\hat{p}_e(1-\hat{p}_e)/n}\right), \qquad (*)$$

где $\hat{p}_e$ – значение оценки $\hat{p}=\mathrm{X}/n$, вычисленное по реализации выборки (по результатам эксперимента), по условию задачи $\hat{p}_e=0{,}006$.

Квантиль стандартного нормального распределения $u_{1-\alpha/2}=u_{0,975}=$ $=1{,}96$. Подставим числовые данные в (*):

$$u_{1-\alpha/2}\sqrt{\hat{p}_e(1-\hat{p}_e)/n}=1{,}96\ \sqrt{0{,}006\,(1-0{,}006)/10000}=\ 0{,}0015,$$

$$\hat{p}_e-u_{1-\alpha/2}\sqrt{\hat{p}_e(1-\hat{p}_e)/n}=0{,}006-0{,}0015=0{,}0045,$$

$$\hat{p}_e+u_{1-\alpha/2}\sqrt{\hat{p}_e(1-\hat{p}_e)/n}=0{,}006+0{,}0015=0{,}0075.$$

В итоге получаем искомую *реализацию приближенного доверительного интервала* $(0{,}0045;0{,}0075)$, *соответствующего доверительной вероятности* $1-\alpha=0{,}95$.