

3.7. Нормальное распределение

Нормальное распределение, распределение Гаусса – предельный закон распределения событий и явлений, являющихся результатом действия множества детерминированных факторов (физических причин, случайно сочетающихся), каждый из которых по интенсивности не выделяется на фоне других. Хаос и динамичность в причинно-следственных связях и их результат математически выражаются экспоненциальной функцией:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^x e^{-(x-M_x)^2/2\sigma_x^2} dx \quad (3.23)$$

Примеры геологических характеристик, подчиняющихся закону нормального распределения: топография, плотность морского песка, показатель сферичности для заданного размера частиц, показатель окатанности галек разного размера, уровень воды в скважине в течение времени, плотность гидросети, удельный вес образцов из гранитного массива, плотность заполнения зерен в песчанике, размеры беспозвоночных ископаемых организмов, угол наклона морского пляжа, угол склона речной долины, угол падения косой слоистости песчаника, пористость песчаника, содержание минералов в породах, содержание влаги в осадочных породах, содержание некоторых химических элементов.

Одним из исторических примеров демонстрации закона нормального распределения и центральной предельной теоремы является доска Гальтона (Рис 3.3.).

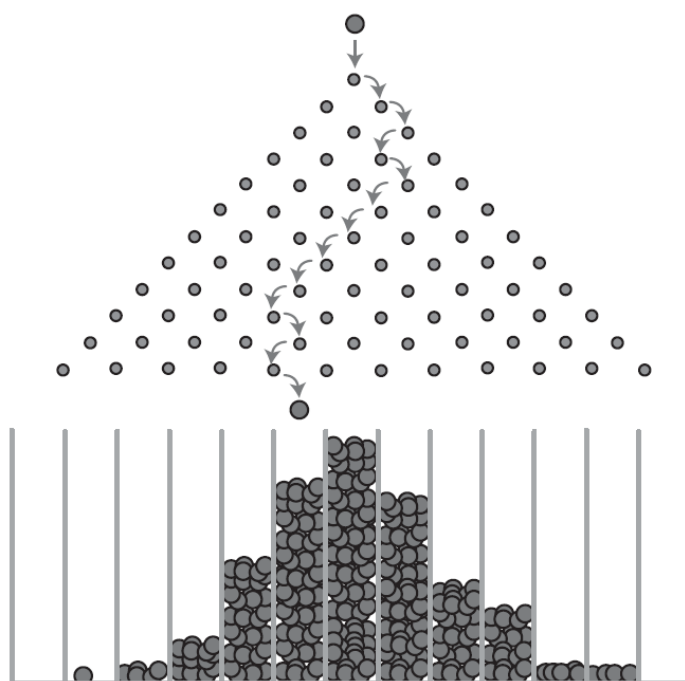


Рис.3.3. Доска Гальтона

Доска Гальтона демонстрирует, что сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет распределение, близкое к нормальному. Если сравнить результат распределения шариков на Рис 3.3. с нормальным распределением, Рис.3.4. то очевидна аналогия изображений. Самым простым и самым не точным способом проверки распределения на соответствие нормальному распределению является визуальная оценка. Поэтому для достоверной проверки изучаемого распределения на нормальность, следует пользоваться статистическими гипотезами.

3.8. Доверительный интервал для математического ожидания

Доверительный интервал - статистическая оценка параметра исследуемого вероятностного распределения, имеющая вид интервала, границами которого служат функции от результатов наблюдений и доверительной вероятности, который с вероятностью P "накрывает" неизвестное значение параметра. Так как математическое ожидание вычислить невозможно, его возможно только оценить, то при нахождении оценки математического ожидания, следует

указывать доверительный интервал с соответствующей доверительной вероятностью. *Доверительная вероятность* P - вероятность достоверности принимаемой гипотезы, характеристика надёжности, полученной по выборке оценки того или иного параметра. Сопутствующим параметром доверительной вероятности P является *уровень значимости* α - вероятность допущения ошибок, т.е. $\alpha = 1 - P$.

Чем выше гарантия надёжности оценки, тем больше величина интервала, в котором может находиться генеральный параметр. В исследованиях прикладного характера доверительная вероятность обычно принимается $P=0,95$. Соответственно, *уровень значимости* $\alpha = 1 - P = 0.05$ в случае *нормально распределённой* случайной величины это соответствует вероятности попадания случайной величины в интервал $M_x - 2\sigma \leq X \leq M_x + 2\sigma$ (*правило двух сигма*). Для оценки математического интервала, согласно *правилу двух сигма* можно записать доверительный интервал,:

$$\bar{x} - 2s < M_x < \bar{x} + 2s, \quad (3.24)$$

с доверительной вероятностью не более 0.95.

По аналогии с вероятностью 0,997 – *правило* $M_x - 3\sigma \leq X \leq M_x + 3\sigma$ *трёх сигма*, можно записать доверительный интервал:

$$\bar{x} - 3s < M_x < \bar{x} + 3s. \quad (3.25)$$

Правилу одного сигма $M_x - \sigma < X < M_x + \sigma$ соответствует доверительная вероятность 0,683. Интервал соответствующий *правилу одного сигма* ещё называют *приблизительным интервалом*, в котором с той или иной вероятностью (но не выше 0,683) может находиться математическое ожидание выражается стандартными границами:

$$\bar{x} - s < M_x < \bar{x} + s \quad (3.26)$$

Иллюстрация этих правил для нормального распределения представлено на рис 3.4.

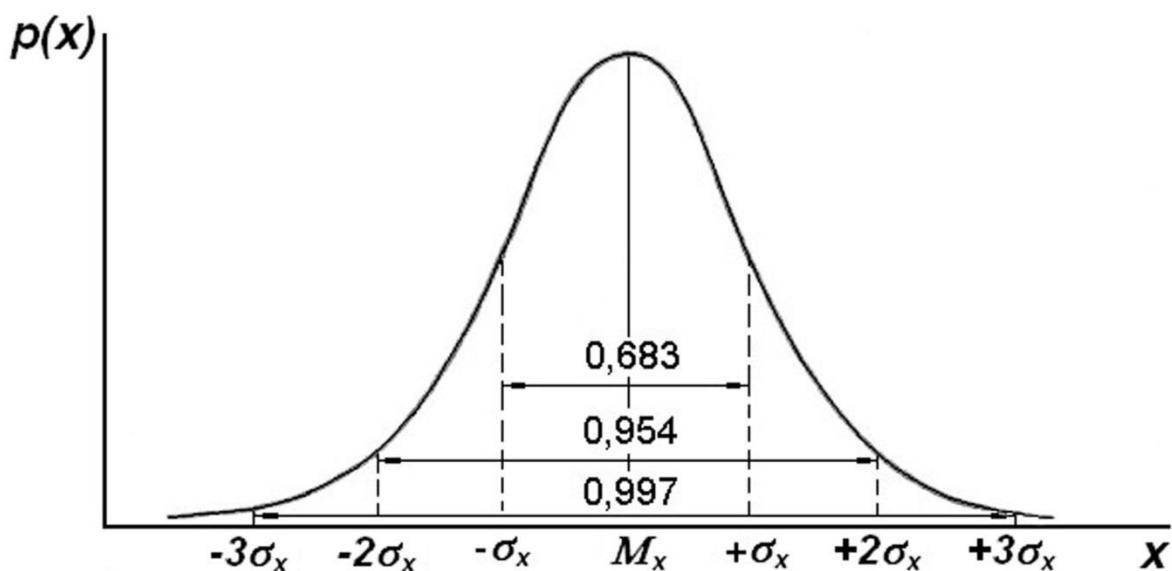


Рис 3.4. Правило одного, двух и трёх сигма примере нормального распределения.

Доверительное отклонение - функция от результатов наблюдений и доверительной вероятности, позволяющая оценить доверительный интервал, который с вероятностью P «накрывает» неизвестное значение параметра:

$$s_x^\alpha = \pm \frac{s_x}{\sqrt{n}} \cdot t_v^\alpha, \quad (3.27)$$

где s_x - *квадратичное отклонение*, n -количество наблюдений; t_v^α - критерий *Стьюдента*, случайная величина, характеризующая соотношение неизвестной ошибки определения M_x и квадратичного отклонения s_x :

$$t_v^\alpha = \frac{(\bar{x} - M_x) \sqrt{n}}{s_x} \quad (3.28)$$

Используя t_v^α - критерий *Стьюдента*, можно с заданным уровнем значимости $\alpha = 1 - P$ указывать доверительный интервал для математического ожидания, по следующему правилу:

$$\bar{x} - s_x^\alpha \leq M_x \leq \bar{x} + s_x^\alpha \quad (3.29)$$

Этим соотношением пользуются практически при обработке экспериментальных данных для определения доверительного интервала. Необходимо отметить, что *уровнем значимости* α исследователь задаётся независимым путём, т.е. исходя из сущности задачи, природы процесса,

явления. При вычислении s_x^α , параметр t_v^α берется из таблиц, для *двустороннего* или *одностороннего критерия Стьюдента*, в зависимости от физической сущности задачи.

t-критерий - критерий значимости, основанный на распределении Стьюдента и используемый для проверки гипотез о средних значениях нормально распределённых физических величин и для проверки на значимость оценок параметров. t_v^α для одностороннего критерия берётся из таблицы. t_v^α для двустороннего критерия берётся из таблицы для одностороннего критерия, для вдвое меньшего уровня значимости $t_v^{\alpha/2}$. На рис.3.5. представлена графическая интерпретация одностороннего критерия Стьюдента – плотность вероятностей *t*-критерия. Случай одностороннего критерия – заштрихована критическая область, составляющая α площади под кривой. На Рис. 3.6. представлена Плотность вероятностей двустороннего *t*-критерия. На графике отмечены левая $-t^{\alpha/2}$ и правая $t^{\alpha/2}$ критические области двустороннего *t*-критерия. Сумма площадей этих областей соответствует уровню значимости α . Так как ошибки измерений подлежат нормальному распределению, то при вычислении доверительного интервала для математического ожидания следует использовать двусторонний критерий Стьюдента (Приложение 1.).

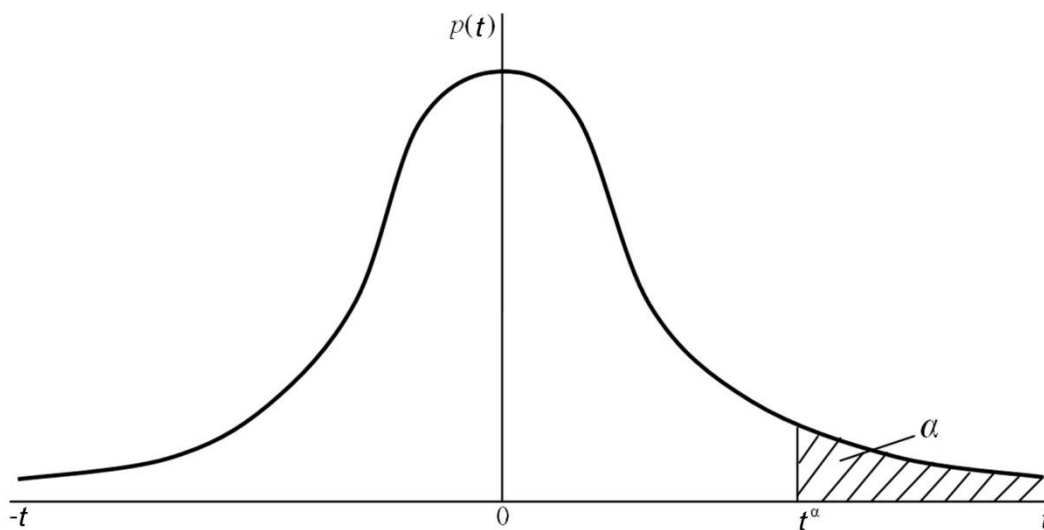


Рис.3.5.Плотность вероятностей одностороннего *t*-критерия

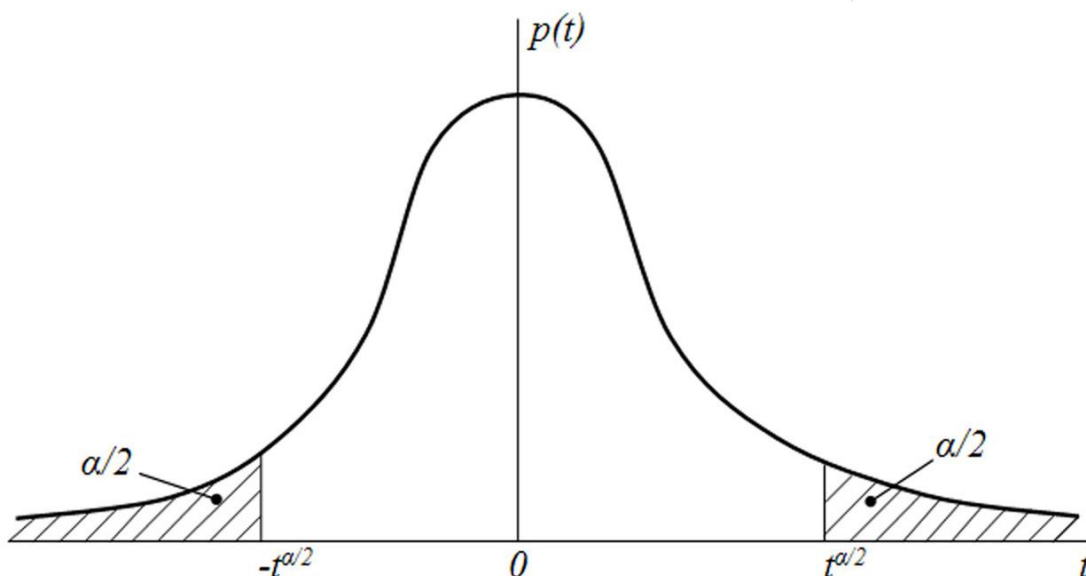


Рис.3.6. Плотность вероятностей двустороннего t -критерия

ПРИМЕР 6. (Построение доверительных интервалов, коэффициент вариации. Отсев выпадающих точек).

Рассмотрим *процедуру оценки* математического ожидания и *дисперсии* случайной величины на примере *определения коэффициента* физической проницаемости цементного камня. Коэффициент проницаемости определён по *результатам* лабораторного исследования шести образцов цементного камня. В этом примере речь идёт только об определении коэффициента физической проницаемости цементного камня. При этом предполагается, что *ошибки измерений* подчиняются *закону нормального распределения*. Исходные данные представлены в следующей таблице:

i	Коэффициент физической проницаемости $k \cdot 10^{13} \text{ м}^2$
1	0,239
2	0,2472
3	0,2421
4	0,2525

5	0,2469
6	0,2508

В качестве среднего значения выберем среднее арифметическое и по формуле (3.16) вычислим его, $\bar{k} = 0,2464 \cdot 10^{-13}$. Определим дисперсию по формуле (3.6) $S_{om}^2 = 0,26092 \cdot 10^{-30}$, следовательно квадратичное отклонение по согласно (3.7) равняется $S_k = 0,5108 \cdot 10^{-15}$ является квадратичным отклонением или стандартным отклонением. Вычислим коэффициент вариации по формуле (3.11), $V_x = s_k / \bar{k} = 0,0207$. Полученный коэффициент вариации характеризует относительную ошибку эксперимента. Его можно выразить в процентах - 2,07%. Что является достаточно хорошей точностью результатов.

Приблизительный интервал, в котором с той или иной вероятностью (но не выше 0,683) может находиться математическое ожидание коэффициента физической проницаемости исследованного образца цементного камня, выражается стандартными границами (3.26):

$$0,2464 \cdot 10^{-13} - 0,5108 \cdot 10^{-15} \leq M_k \leq 0,2464 \cdot 10^{-13} + 0,5108 \cdot 10^{-15}$$

Следует заметить, что математическое ожидание может находиться и за пределами стандартных границ. Для того чтобы можно было указать надёжность, достоверность интервала, необходимо воспользоваться соотношениями (3.27), (3.29). Вначале нужно задаться вероятностью ошибки или, точнее, доверительной вероятностью обозначения интервала, в котором может находиться неизвестное значение коэффициента проницаемости образца цементного камня. В исследованиях прикладного характера доверительная вероятность принимается обычно $P=0,95$. Соответственно, уровень значимости $\alpha = 1 - P = 0.05$. В случае нормально распределённой случайной величины это соответствует правилу двух сигма (3.24).

Вычислим доверительное отклонение коэффициента проницаемости для уровня значимости $\alpha = 0.05$. По таблице (Приложение 1) найдём значение двустороннего критерия Стьюдента для числа степеней свободы $\nu = n - 1 = 6 - 1 = 5$. Табличный двусторонний критерий Стьюдента $t_5^{0,05} = 2,57$

=2,57. Вычислим по формуле (3.27) доверительное отклонение

$$s_k^{0,05} = \frac{0,5108 \cdot 10^{-15}}{\sqrt{6}} \cdot 2,57 = 0,536 \cdot 10^{-15}. \text{Теперь результаты расчётов более}$$

определённые, с доверительной вероятностью $P=0,95$ математическое ожидание коэффициента физической проницаемости исследованного образца цементного камня находится в интервале:

$$0,2464 \cdot 10^{-13} - 0,536 \cdot 10^{-15} \leq M_k \leq 0,2464 \cdot 10^{-13} + 0,536 \cdot 10^{-15},$$

округляя, получим:

$$(0,246 - 0,005) \cdot 10^{-13} \leq M_k \leq (0,246 + 0,005) \cdot 10^{-13}.$$

Теперь можно обсуждать вопрос о «выпадающих точках», т.е. о значениях k , находящихся за пределами доверительных границ и с вероятностью $P=0,95$ являющихся результатами грубых ошибок экспериментатора.

Выпадающими точками являются точки под номерами 1 и 4. Удалим их и произведём расчёт с оставшимися четырьмя значениями: $\bar{k} = 0,24675 \cdot 10^{-13}$, $S_{on}^2 = 0,1275 \cdot 10^{-30}$, $S_k = 0,357 \cdot 10^{-15}$, коэффициент вариации $V_x = 0,014468$. По коэффициенту вариации заметим, что относительная ошибка эксперимента снизилась с 2,07% до 1,45%. Этот факт демонстрирует целесообразность применения процедуры отсева выпадающих точек, для получения более точного результата оценки статистических характеристик. Далее можно записать доверительный интервал и в случае большого количества исходных данных повторить процедуру отсева.