Некоторые понятия теории вероятностей и математической статистики

Случайное событие

- **Случа́йное собы́тие** — подмножество множества исходов *случайного эксперимента*; при многократном повторении случайного эксперимента частота наступления события служит оценкой его вероятности.
- Случайное событие, которое никогда не реализуется в результате случайного эксперимента, называется невозможным и обозначается символом Ø. Случайное событие, которое всегда реализуется в результате случайного эксперимента, называется достоверным и обозначается символом ω.

Вероятность случайного события

- **Теория вероятностей** – математическая наука, которая по вероятностям одних событий позволяет оценивать вероятности других событий, связанных с первыми.
- Подтверждением того, что понятие «вероятность события» не имеет определения, является тот факт, что в теории вероятностей существует несколько подходов к объяснению этого понятия:
- **<u>Классическое определение вероятности</u>** случайного события**.**
- Вероятность события равна отношению числа благоприятных событию исходов опыта к общему числу исходов опыта.
- $$P(A) = \frac{m}{n}$$

, где

- m - число благоприятных исходов опыта;
- n - общее число исходов опыта.

Вероятность случайного события

- Исход опыта называется **благоприятным** для события , если при этом исходе опыта появилось событие . Например, если событие - появление карты красной масти, то появление туза бубей – исход, благоприятный событию.
-
- **<u>Примеры.</u>**
- 1) Вероятность выпадения 5 очков на грани кубика равна 1/6, поскольку кубик может упасть любой из 6 граней кверху, а 5 очков находятся только на одной грани.
- 2) Вероятность выпадения герба при однократном бросании монеты – 1/2, поскольку монета может упасть гербом или решкой – два исхода опыта, а герб изображен лишь на одной стороне монеты.
- 3) Если в урне 12 шаров, из которых 5 – черные, то вероятность вынуть черный шар – 5/12, поскольку всего исходов опыта – 12, а благоприятных из них - 5
- **<u>Замечание.</u>** Классическое определение вероятности применимо при двух условиях:
- 1) все исходы опыта должны быть равновероятными;
- 2) опыт должен иметь конечное число исходов.

Случайная величина

- **Случайная величина** — это величина, которая принимает в результате опыта одно из множества значений, причём появление того или иного значения этой величины до её измерения нельзя точно предсказать.
- Случайные величины могут принимать дискретные, непрерывные и дискретно-непрерывные значения. Соответственно случайные величины классифицируют на дискретные, непрерывные и дискретно-непрерывные (смешанные).
- На схеме испытаний может быть определена как отдельная случайная величина (одномерная/скалярная), так и целая система одномерных взаимосвязанных случайных величин (многомерная/векторная).
- Пример смешанной случайной величины — время ожидания при переходе через автомобильную дорогу в городе на нерегулируемом перекрёстке.
- В бесконечных схемах (дискретных или непрерывных) уже изначально элементарные исходы удобно описывать количественно. Например, номера градаций типов несчастных случаев при анализе ДТП; время безотказной работы прибора при контроле качества и т. п.

Случайная величина

- Числовые значения, описывающие результаты опытов, могут характеризовать не обязательно отдельные элементарные исходы в схеме испытаний, но и соответствовать каким-то более сложным событиям.
- С одной стороны, с одной схемой испытаний и с отдельными событиями в ней одновременно может быть связано сразу несколько числовых величин, которые требуется анализировать совместно.
- Например, координаты (абсцисса, ордината) какого-то разрыва снаряда при стрельбе по наземной цели; метрические размеры (длина, ширина и т. д.) детали при контроле качества; результаты медобследования (температура, давление, пульс и пр.) при диагностике больного; данные переписи населения (по возрасту, полу, достатку и пр.).
- Поскольку значения числовых характеристик схем испытания соответствуют в схеме некоторым случайным событиям (с их определёнными вероятностями), то и сами эти значения являются случайными (с теми же вероятностями). Поэтому такие числовые характеристики и принято называть случайными величинами. При этом расклад вероятностей по значениям случайной величины называется законом распределения случайной величины.

Числовые характеристики случайной величины

Математическое ожидание случайной величины

Математическое ожидание - число, вокруг которого сосредоточены значения случайной величины. Математическое ожидание случайной величины ξ обозначается $\boldsymbol{M}\xi$.

Математическое ожидание дискретной случайной величины ξ , имеющей распределение

x_1	x_2	...	x_n
p_1	p_2	...	p_n

называется величина $M\xi = \sum_{i=1}^{n} p_i x_i$, если число значений случайной величины конечно.

Если число значений случайной величины счетно, то $M\xi = \sum_{i=1}^{\infty} p_i x_i$. При этом, если ряд в правой части равенства расходится, то говорят, что случайная величина ξ не имеет математического ожидания.

Математическое ожидание непрерывной случайной величины с плотностью вероятностей $p_\xi(x)$ вычисляется по формуле $M\xi = \int_{-\infty}^{\infty} x p_\xi(x) dx$. При этом, если интеграл в правой части равенства расходится, то говорят, что случайная величина ξ не имеет математического ожидания.

Числовые характеристики случайной величины

Дисперсия случайной величины характеризует меру разброса случайной величины около ее математического ожидания.
Если случайная величина ξ имеет математическое ожидание $\boldsymbol{M}\xi$, то *дисперсией* случайной величины ξ называется величина $\boldsymbol{D}\xi = \boldsymbol{M}(\xi - \boldsymbol{M}\xi)^2$.
Легко показать, что $\boldsymbol{D}\xi = \boldsymbol{M}(\xi - \boldsymbol{M}\xi)^2 = \boldsymbol{M}\xi^2 - \boldsymbol{M}(\xi)^2$.
Эта универсальная формула одинаково хорошо применима как для дискретных случайных величин, так и для непрерывных. Величина $\boldsymbol{M}\xi^2$ >для дискретных и непрерывных случайных величин соответственно вычисляется по формулам

$$M\xi^2 = \sum_{i=1}^{n} p_i x_i^2, \quad M\xi^2 = \int_{-\infty}^{\infty} x^2 p_\xi(x)dx.$$

Для определения меры разброса значений случайной величины часто используется *среднеквадратичное отклонение* σ_ξ, связанное с дисперсией соотношением $\sigma_\xi = \sqrt{D\xi}$.

Основные свойства дисперсии:

- дисперсия любой случайной величины неотрицательна, $\boldsymbol{D}\xi \geq 0$;
- дисперсия константы равна нулю, $\boldsymbol{D}c=0$;
- для произвольной константы $\boldsymbol{D}(c\xi) = c^2\boldsymbol{D}(\xi)$;
- дисперсия суммы двух *независимых* случайных величин равна сумме их дисперсий: $\boldsymbol{D}(\xi \pm \eta) = \boldsymbol{D}(\xi) + \boldsymbol{D}(\eta)$.

Числовые характеристики случайной величины

- **Медиана** — это такое значение признака, которое разделяет ранжированный ряд распределения на две равные части — со значениями признака меньше медианы и со значениями признака больше медианы. Для нахождения медианы, нужно отыскать значение признака, которое находится на середине упорядоченного ряда.
- Предположим, что в одной комнате оказалось 19 бедняков и один миллиардер. Каждый кладёт на стол деньги — бедняки из кармана, а миллиардер — из чемодана. По $5 кладёт каждый бедняк, а миллиардер — $1 млрд ($10^9$). В сумме получается $1 000 000 095. Если мы разделим деньги равными долями на 20 человек, то получим $50 000 004,75. Это будет среднее арифметическое значение суммы наличных, которая была у всех 20 человек в этой комнате.
- Медиана в этом случае будет равна $5 (полусумма десятого и одиннадцатого, *срединных* значений ранжированного ряда). Можно интерпретировать это следующим образом. Разделив нашу компанию на две равные группы по 10 человек, мы можем утверждать, что в первой группе каждый положил на стол не больше $5, во второй же не меньше $5. В общем случае можно сказать, что медиана это то, сколько принёс с собой *средний* человек. Наоборот, среднее арифметическое — неподходящая характеристика, так как оно значительно превышает сумму наличных, имеющуюся у среднего человека.

Числовые характеристики случайной величины

- **Мода** — значение во множестве наблюдений, которое встречается наиболее часто. Случайная величина может не иметь моды. Иногда в совокупности встречается более чем одна мода *(например: 2, 6, 6, 6, 8, 9, 9, 9, 10; мода = 6 и 9)*. В этом случае можно сказать, что совокупность мультимодальна. Из структурных средних величин только мода обладает таким уникальным свойством. Как правило мультимодальность указывает на то, что набор данных не подчиняется нормальному распределению.
- Мода как средняя величина употребляется чаще для данных, имеющих нечисловую природу. Среди перечисленных цветов автомобилей — *белый, черный, синий металлик, белый, синий металлик, белый* — мода будет равна *белому* цвету. При экспертной оценке с её помощью определяют наиболее популярные типы продукта, что учитывается при прогнозе продаж или планировании их производства.

Нормальное распределение случайной величины

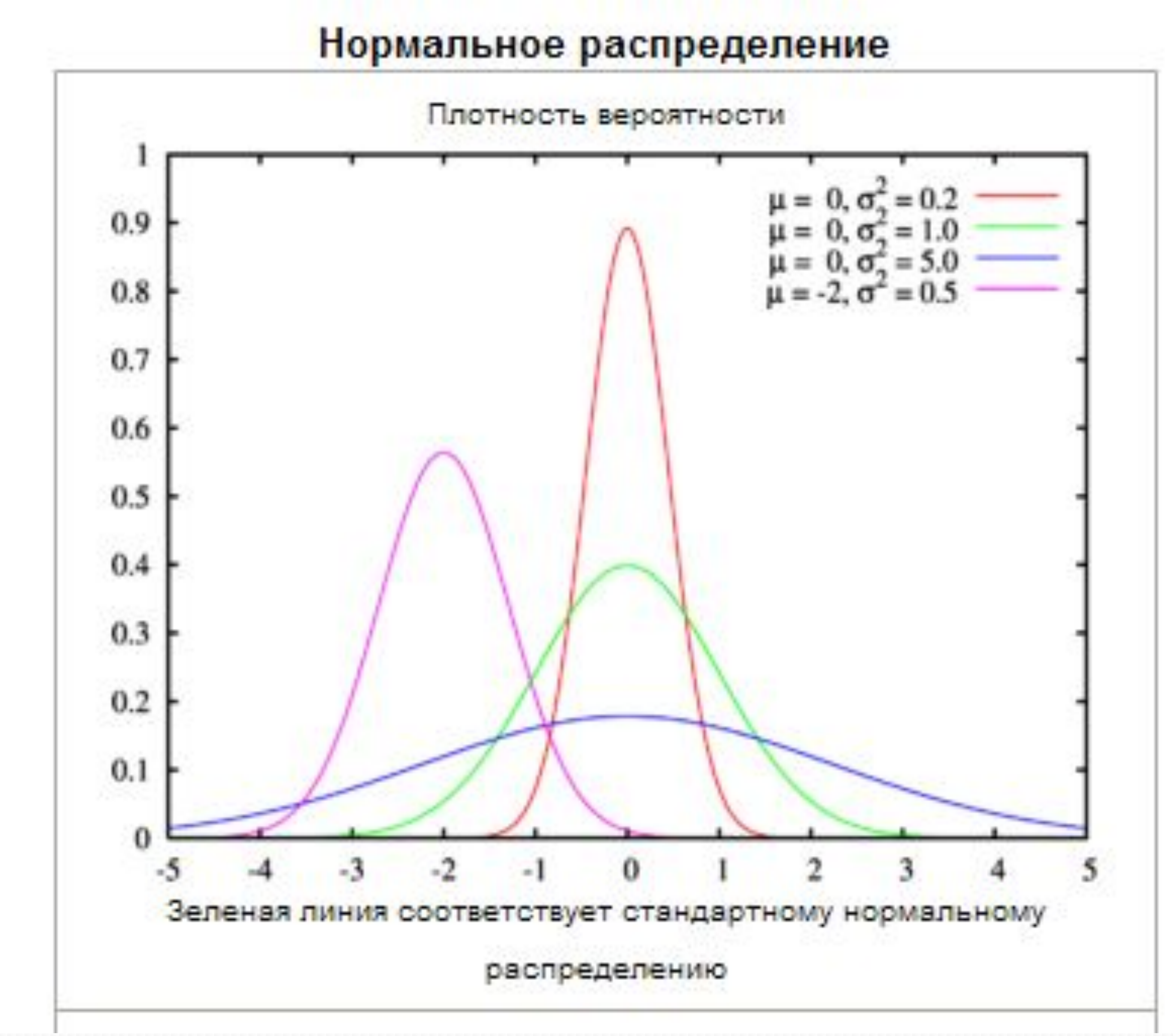
Нормальное распределение, также называемое **гауссовым распределением**, **гауссианой** или **распределением** Гаусса — распределение вероятностей, которое задается функцией плотности распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

где параметр μ — среднее значение (математическое ожидание) случайной величины и указывает координату максимума кривой плотности распределения, а σ^2 — дисперсия.

Нормальное распределение играет важнейшую роль во многих областях знаний, особенно в статистической физике. Физическая величина, подверженная влиянию значительного числа независимых факторов, способных вносить с равной погрешностью положительные и отрицательные отклонения, вне зависимости от природы этих случайных факторов, часто подчиняется нормальному распределению, поэтому из всех распределений в природе чаще всего встречается нормальное (отсюда и произошло одно из названий этого распределения вероятностей).

распределение случайной величины



КОРРЕЛЯЦИЯ

- **КОРРЕЛЯЦИЯ** [correlation] — величина, характеризующая взаимную зависимость двух *случайных величин*, X и Y, безразлично, определяется ли она некоторой причинной связью или просто случайным совпадением (**ложной К.**). Для того чтобы определить эту зависимость, рассмотрим новую случайную величину — произведение отклонения значений x от его среднего Mx и отклонения y от своего среднего My. Можно вычислить среднее значение новой случайной величины:
- $r_{xy} = M\{(x - Mx)(y - My)\}$.
- Это среднее получило название **корреляционной функции**, или **ковариации.** На ее основе (делением на корень из произведения *дисперсий* σ^2_x, σ^2_y, т. е. на произведение *стандартных отклонений*) строится **коэффициент К.**:

$$R_{xy} = \frac{r_{xy}}{\sigma_x \sigma_y}.$$

- При нелинейной зависимости аналогичный показатель носит название **индекса К.**
- Если x и y независимы, то $R_{xy} = 0$. Если же x и y зависимы, то обычно $R_{xy} \neq 0$. Причем в тех случаях, когда зависимость полная, то либо $R_{xy} = 1$ (x и y растут или уменьшаются одновременно), либо $R_{xy} = -1$ (при увеличении одной из них другая уменьшается). Следовательно, коэффициент К. может изменяться от -1 до $+1$.

Корреляция

- Впервые в научный оборот термин «корреляция» ввёл французский палеонтолог Жорж Кювье в XVIII веке. Он разработал «закон корреляции» частей и органов живых существ, с помощью которого можно восстановить облик ископаемого животного, имея в распоряжении лишь часть его останков. В статистике слово «корреляция» первым стал использовать английский биолог и статистик Фрэнсис Гальтон в конце XIX века.