

Теория вероятности

Дима Трушин

Семинар 3

Случайные величины

Понятие случайной величины Предположим вы изучаете какой-то случайный объект. Чтобы описать такой объект, вы должны описать множество всех его состояний Ω и вероятностную меру на этом множестве P . Однако, может так случиться, что множество Ω непонятно как устроенным, ибо вы же не знаете, что внутри вашего черного ящика. Например, если работаете с элементарными частицами. Да фиг его знает, как там что у них устроено и в каких состояниях может быть частица! Но даже если вы его знаете, может случиться другая беда, оно на столько огромное, что непонятно как с ним работать. Например, если вы хотите взять случайного человека, то все возможные исходы – это все возможные люди на земле. Кто хочет работать с множеством Ω порядка 7,5 миллиарда элементов?

Однако, часто для решения конкретной задачи нам и не надо знать все Ω целиком. Вместо этого надо лишь знать что-то, что важно для задачи. Тут на помощь приходят случайные величины.

Представьте, что у вас есть некоторый черный ящик. Вместо того, чтобы задавать ему вопрос: «В каком ты состоянии?» мы можем задавать более конкретные вопросы: «какая у тебя температура?», «какой у тебя цвет?», «какое у тебя давление?» и т.д. Про это можно думать так, что к нашему черному ящику мы прикрепляем измерительный прибор, который на выходе выдает некоторое число. И когда мы задаем вопрос черному ящику, мы просто смотрим на показания нашего прибора. В результате, мы получаем информацию не в виде полного описания в каком состоянии находится наш черный ящик, а всего лишь число, показывающее значение конкретной величины.

Если наш случайный объект описывался парой (Ω, P) , то случайная величина – это правило, которое каждому элементарному исходу $\omega \in \Omega$ ставит в соответствие некоторое вещественное число. То есть, случайная величина – это отображение $\xi: \Omega \rightarrow \mathbb{R}$.

Примеры

1. Злосчастный кубик описывается парой (Ω, P) , где $\Omega = \{1, 2, 3, 4, 5, 6\}$ и $P(k) = 1/6$. Давайте рассмотрим случайную величину $\xi: \Omega \rightarrow \mathbb{R}$ по правилу $\xi(k) = k \pmod{2}$, то есть остаток от деления на 2. Тогда ξ четные переводит в ноль, а нечетные переводит в 1. Таким образом ξ моделирует нам бросание монетки $\Omega = \{0, 1\}$. С другой стороны, 0 может выпасть на событиях $\{2, 4, 6\}$, то есть вероятность выпадения 0 для ξ равна $1/2$. То есть мы получили правильную монетку с равновероятным выпадением орла и решки.
2. Пусть теперь у нас рассматривается независимое бросание n правильных монеток. Тогда $\Omega = \{(a_1, \dots, a_n) \mid a_i \in \{0, 1\}\}$ состоит из всех последовательностей из нулей и единиц длины n . Давайте рассмотрим случайную величину $\xi: \Omega \rightarrow \mathbb{R}$ равную количеству выпавших решек, то есть $\xi(a_1, \dots, a_n) = a_1 + \dots + a_n$. Тогда ξ принимает значения от 0 до n и вероятность того, что $\xi = k$ равна $C_n^k / 2^n$. Это частный случай биномиального распределения. В данном случае на ξ можно смотреть как на черный ящик с пространством $\Omega = \{0, \dots, n\}$, при этом $P(k) = C_n^k / 2^n$.

Теперь когда мы измеряем только показания прибора (нашу случайную величину), мы можем смотреть на черный ящик с прибором, как на новый случайный объект, состояния которого – это все действительные числа \mathbb{R} . Возникает вопрос, а как же на этом черном ящике устроена вероятность? По определению вероятность нового ящика устроена так:

$$P_\xi(A) = P(\xi \in A) = P(\{\omega \mid \xi(\omega) \in A\}) = P(\xi^{-1}(A)), \text{ для любого } A \subseteq \mathbb{R}$$

То есть для каждого подмножества $A \subseteq \mathbb{R}$ мы смотрим на событие, что наша случайная величина выдала ответ из множества A . Чтобы посчитать это значение, мы смотрим все элементарные исходы ω , при которых

$\xi(\omega) \in A$, сваливаем их в одно событие и измеряем его вероятность. Это и будет то, что мы только что определили.¹

Теперь, когда нас интересуют лишь показания прибора, вместо изучения (Ω, P) мы будем изучать новую пару (\mathbb{R}, P_ξ) . И нам совершенно не важно, каким было оригинальное пространство Ω и мера P . Если ответ на нашу задачу зависит только от значений ξ , то все, что нам надо знать, – это мера P_ξ на прямой. Теперь возникает резонный вопрос: а как вообще задавать вероятностные меры на прямой и какими они бывают? Я предлагаю разобраться с этим вопросом далее.

Вероятностные меры на прямой Что значит задать вероятность на прямой? Это значит, мы должны для каждого подмножества $A \subseteq \mathbb{R}$ задать число $P(A)$ так, чтобы выполнялись аксиомы для вероятности:

1. Для любого $A \in \mathbb{R}$ верно, что $0 \leq P(A) \leq 1$.
2. $P(\emptyset) = 0$, $P(\mathbb{R}) = 1$.
3. Для любой последовательности попарно непересекающихся подмножеств $A_1, A_2, \dots, A_n, \dots \subseteq \mathbb{R}$ (то есть $A_i \cap A_j = \emptyset$) выполнено $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Мягко говоря непонятно, как вообще перебрать все подмножества и приписать им нужное число, так еще и отвратительно сложное условие под третьим пунктом намекает, что мы будем мучиться. Оказывается, что достаточно научиться задавать вероятность того, что случайная величина попала в полуинтервал $(a, b]$, то есть вероятность события $P(a < \xi \leq b)$.^{2, 3} А далее есть общая процедура, которая говорит, как распространить нашу вероятность на любое подмножество прямой.

Мы не будем себя мучить этой абстрактной общей процедурой, а поясним на паре примеров, как отсюда вытащить меры отрезка и интервала. Пусть мы хотим посчитать $P([a, b])$. Тогда надо выбрать последовательность чисел $a_n < a$ таких, что $a_n \rightarrow a$ при $n \rightarrow \infty$. Тогда $P([a, b]) = \lim_{n \rightarrow \infty} P((a_n, b])$, то есть мы пересечем полуинтервалы с левыми концами чуть левее точки a . Аналогично, если мы хотим посчитать $P((a, b))$, выберем последовательность чисел $b_n < b$ такую, что $b_n \rightarrow b$ при $n \rightarrow \infty$. Тогда $P((a, b)) = \lim_{n \rightarrow \infty} P((a, b_n])$, то есть мы объединим полуинтервалы с правыми концами чуть левее b .

Функция распределения Чтобы научиться задавать вероятности $P(\xi \in (a, b])$ достаточно научиться измерять следующую величину

$$F_\xi(x) = P(\xi \leq x) = P(\xi \in (-\infty, x])$$

Тогда функция $F_\xi(x)$ называется функцией распределения случайной величины ξ . Действительно, в этом случае

$$P(\xi \in (a, b]) = P(\xi \in (-\infty, b]) - P(\xi \in (-\infty, a]) = F_\xi(b) - F_\xi(a)$$

То есть, чтобы задать вероятностную меру на прямой, нам достаточно задать одну единственную функцию $F_\xi(x)$ и она однозначно определит некоторую вероятность. Причем формула выше дает явный вид для вероятности попадания в полуинтервал, а в предыдущем разделе я описал как считается попадание в открытый и замкнутый интервалы.

Теперь весь вопрос в том, а любую ли функцию можно взять в качестве $F_\xi(x)$ и если не любую, то какие на нее должны быть условия? Конечно же совсем любая функция не подойдет иначе не будут выполняться аксиомы на вероятность. Но следующий список аксиом является полным описанием функций распределения.

Утверждение. Функция $F: \mathbb{R} \rightarrow \mathbb{R}$ является функцией распределения, то есть задает некоторую вероятностную меру, тогда и только тогда, когда она удовлетворяет следующим свойствам:

1. F является неубывающей функцией, то есть, если $x \leq y$, то $F(x) \leq F(y)$.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ и $\lim_{x \rightarrow \infty} F(x) = 1$.
3. Функция F непрерывна справа, то есть для любой точки $x \in \mathbb{R}$ верно

$$F(x+) = \lim_{t \rightarrow 0, t > 0} F(x+t) = F(x)$$

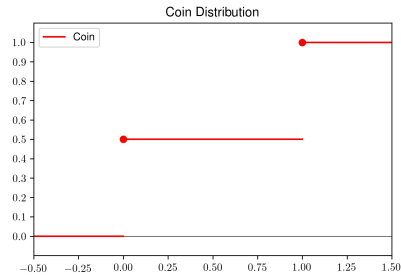
¹Ровно по этому же принципу были определены вероятностные меры в примерах выше.

²Если у вас дежавю с итераторами в стандартных библиотеках языков программирования, то это не случайно. И там и тут ренджи измеряются полуинтервалами.

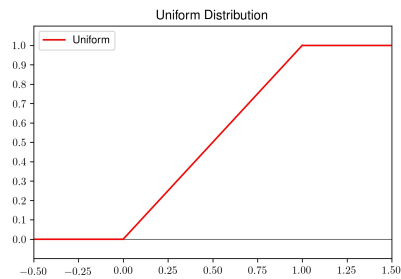
³Необходимость работы с полуинтервалами связана вот с чем. Может так оказаться, что вероятность попасть в границу интервала является ненулевым числом. Чтобы не было путаниц с правилами и вычислениями, оказывается, что полуинтервалы ведут себя сильно лучше, чем интервалы (открытые или замкнутые).

Примеры функций распределения

1. Бросание монетки.



2. Равномерное распределение на отрезке $[0, 1]$.



Какие бывают вероятностные меры на прямой Функция распределения – это замечательно. С помощью нее можно задать абсолютно любую вероятность на прямой. Но оказывается, что если умерить аппетиты и стараться описать не все возможные вероятностные меры, можно дать еще более клевое описание. Так получилось, что в математике есть три класса вероятностных мер:

1. Дискретные.
2. Непрерывные.
3. Экзотика.

Говоря по честному, второй класс называется «Абсолютно непрерывные», а «Непрерывными» называется второй и третий класс вместе. Именно такая терминология принята в математике. Однако, так сложилось, что экзотические распределения в жизни не встречаются, это лишь плот больной фантазии математиков (хотя тут я могу и парочку примеров привести). Поэтому в инженерной и прикладной литературе третий класс вовсе не упоминают, а второй для простоты называют непрерывными. Я решил, что прикладная терминология нас вполне устроит, тем более что придется говорить меньше запутанных слов.

Дискретные случайные величины Пусть я хочу задать дискретную случайную величину ξ на прямой \mathbb{R} . Я в начале должен зафиксировать некоторые данные:

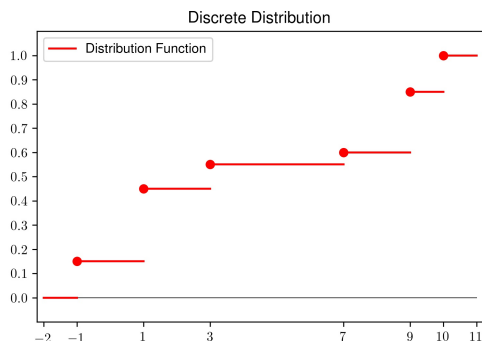
1. Последовательность чисел $a_n \in \mathbb{R}$. Это может быть конечная или бесконечная последовательность любых чисел на прямой. Это будут выделенные значения, которые разрешено принимать нашей случайной величине. Эти значения называются еще атомами.
2. Последовательность чисел $p_n \in \mathbb{R}$. Этих чисел должно быть столько же, сколько и a_n . Это будут вероятности, с которыми наша случайная величина будет принимать выделенные значения. Потому они должны удовлетворять следующим свойствам:

- (a) $0 < p_i \leq 1$.
- (b) $\sum_{i=1}^{\infty} p_i = 1$.

Тогда для произвольного $A \subseteq \mathbb{R}$ вероятность задается так

$$P_{\xi}(A) = \sum_{a_i \in A} p_i$$

то есть мы смотрим какие атомы попали в множество A и складываем соответствующие этим атомам вероятности p_i . В частности $P(\xi = a_i) = p_i$. И если число a не равно никакому из a_i , то вероятность $P(\xi = a) = 0$. Функция распределения $F_{\xi}(x)$ в этом случае будет ступенчатой



В качестве примера можно рассмотреть бросание кубика. Случайная величина ξ – число на грани. Атомами будут числа от 1 до 6, а приписанные им вероятности все равны $1/6$ если кубик сбалансированный. Еще один пример – бросание монетки. Если орел, то прибор показывает 0, а если решка, то – 1. Если монетка сбалансированная, то приписанные вероятности будут $1/2$ в обоих случаях.

Примеры

1. Распределение Бернулли. Это распределение для произвольной (необязательно правильной) монетки. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 \\ p & q \end{cases}$$

при этом $p + q = 1$ и $p, q \geq 0$.

2. Биномиальное распределение $B(n, p)$. Это распределение для числа выпавших решек в n независимых бросаниях (необязательно правильной) монетки. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 & \dots & k & \dots & n \\ q^n & C_n^1 p q^{n-1} & \dots & C_n^k p^k q^{n-k} & \dots & p^n \end{cases}$$

при этом $p + q = 1$ и $p, q \geq 0$.

3. Распределение Пуассона $P(\lambda)$, при этом $\lambda > 0$. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 & \dots & k & \dots \\ e^{-\lambda} & e^{-\lambda} \lambda & \dots & \frac{e^{-\lambda} \lambda^k}{k!} & \dots \end{cases}$$

4. Геометрическое распределение. Это распределение отвечает за вероятность появления первого успеха при бросании неправильной монетки, а именно, эта величина равна номеру первого успеха.

$$\xi \sim \begin{cases} 1 & 2 & \dots & k & \dots \\ p & qp & \dots & q^{k-1}p & \dots \end{cases}$$

Второй вариант геометрического распределения: количество неудач до первого успеха, то есть $\xi - 1$.

Непрерывные случайные величины Пусть я хочу задать непрерывную случайную величину ξ на прямой \mathbb{R} . Я в начале должен зафиксировать некоторые данные:

1. Функция $p: \mathbb{R} \rightarrow \mathbb{R}$ по правилу $x \mapsto p(x)$. Эта функция будет называться плотностью случайной величины ξ и должна удовлетворять следующим свойствам:

- (a) $0 \leq p(x)$ для всех $x \in \mathbb{R}$.
- (b) $\int_{-\infty}^{\infty} p(x) dx = 1$.

Тогда для произвольного подмножества $A \subseteq \mathbb{R}$ вероятность задается так

$$P_{\xi}(A) = \int_A p(x) dx$$

Если, например, $A = [a, b]$, полуинтервал $A = (a, b]$ или интервал $A = (a, b)$, то

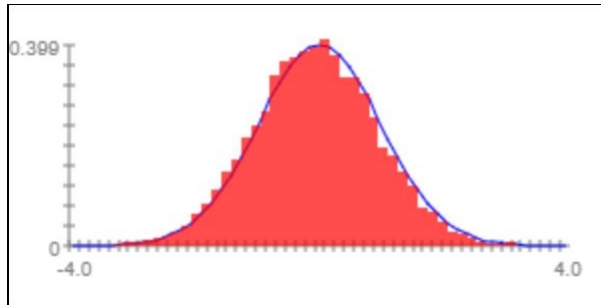
$$P_{\xi}([a, b]) = P_{\xi}((a, b]) = P_{\xi}((a, b)) = \int_a^b p(x) dx$$

То есть интеграл от a до b задает вероятность, что случайная величина выдала ответ в интервале от a до b . Еще обратите внимание, что для любого числа $a \in \mathbb{R}$ выполнено

$$P(\xi = a) = \int_a^a p(x) dx = 0$$

То есть вероятность попасть в каждую конкретную точку равна нулю, но при этом сама функция вероятности не нулевая. Еще стоит обратить внимание на связь между плотностью и функцией распределения. Если вы знаете функцию распределения $F_{\xi}(x)$, то плотность будет вычисляться как производная $p(x) = F_{\xi}(x)'$.

Частотное понимание плотности Здесь я опишу, как можно думать про плотность случайной величины. Пусть задана непрерывная случайная величина ξ с плотностью $p: \mathbb{R} \rightarrow \mathbb{R}_+$, то есть $\int_{\mathbb{R}} p(x) dx = 1$. Давайте разобьем всю прямую на интервалы равной длины, скажем длины δ . Будем думать про эти интервалы как про корзины. Теперь поставим n экспериментов с нашей случайной величиной ξ , то есть спросим ее состояние n раз. Она нам выдаст n чисел, скажем, a_1, \dots, a_n . Тогда разложим полученные ответы по корзинам. Теперь частота попадания в корзину получается равна количеству ответов в корзине деленному на n – общее количество ответов. Получится ступенчатая функция, которая при малом δ и большом n будет очень похожа на плотность. Ниже на картинке синим изображен график плотности, а красным ступеньки из частот для корзин. Сами корзины отмечены штрихами вдоль горизонтальной оси.



Функция $p(x)$ не просто так называется плотностью. Если думать про меру P не как про вероятность, а как про массу, то величина $P(A)$ показывает какая масса у куса прямой A . Тогда $p(x)$ имеет физический смысл той самой плотности к которой мы привыкли в физике. То есть она показывает как много вещества находится в данной точке пространства, а масса будет интегралом от плотности. Потому плотность вероятности показывает как бы сколько «правдоподобности» находится в данной точке, а вероятность будет интегралом от этой «правдоподобности».

Примеры

1. Равномерное распределение на отрезке $[a, b]$. Плотность имеет вид

$$p(x) = \frac{1}{b-a} \chi_{[a,b]}(x), \quad \text{где} \quad \chi_{[a,b]}(x) = \begin{cases} 1, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

2. Экспоненциальное распределение. Его плотность задается в виде

$$p(x) = \lambda e^{-\lambda x} \theta(x), \quad \text{где} \quad \theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

При этом $\lambda > 0$.

3. Нормальное распределение $N(a, \sigma)$. Его плотность задается в виде

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Здесь $a \in \mathbb{R}$ любое, а $\sigma > 0$.⁴

Характеристики случайных величин

Пусть у нас ξ – некоторая случайная величина на некотором вероятностном пространстве (Ω, P) . Формально, ξ – это некоторая функция на Ω . А что значит, что эта величина на самом деле не случайная? Это значит, что ξ принимает одно и то же значение для всех элементарных исходов, то есть ξ является константой. Таким образом мы можем рассматривать обычные числа как случайные величины, которые правда оказываются уже не такими уж и случайными. Так как с числами работать проще, чем с непонятными случайными величинами. Действительно, чтобы задать число – надо задать число⁵, а чтобы задать случайную величину нам надо задавать какую-то там функцию распределения с кучей всяких свойств, неприятно. Потому очень полезно иметь на виду некоторые числовые характеристики для случайных величин.

Математическое ожидание Философия этой характеристики следующая. Да, мы знаем, что наша случайная величина не является константой и выдает разный ответ. Но уж очень нам хочется считать ее константой. Тогда возникает вопрос, а какое число является наилучшей заменой для нашей случайной величины? То есть: какое число в среднем будет отвечать так же, как и наша случайная величина? На самом деле в математическом анализе давно известен механизм замены функции числом – замена функции ее интегралом. Раз $\xi: \Omega \rightarrow \mathbb{R}$ – это функция на Ω , то мы хотим определить математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \int_{\Omega} \xi dP = \int_{\mathbb{R}} x dP_{\xi}$$

Вот только придать смысл мы этой формуле строго не сможем. Потому что в отличие от матанализа, мы не понимаем, что такое интегрировать по произвольному множеству Ω вместо отрезка или прямой, да еще и вместо dx стоит какая-то дрянь в виде dP или dP_{ξ} . Но эта формула должна восприниматься интуитивно очень ясно. Мы просто в каком-то смысле усредняем значения $\xi(\omega)$ в зависимости от вероятности их появления. К сожалению в общем виде я вам не смогу объяснить, что такое математическое ожидание. Вместо этого я расскажу, как оно считается в дискретном и непрерывном случаях. Этого будет достаточно для абсолютно всех приложений. Еще полезно понимать, что математическое ожидание вообще говоря может не существовать для данной случайной величины.

⁴Ниже я расскажу про него чуть более подробно. Это один из самых важных классов распределения в основном благодаря Центральной предельной теореме, которая говорит, что это распределение в некотором смысле универсально.

⁵Спасибо, Кэп.

Математическое ожидание в дискретном случае Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases}$$

Тогда определим ее математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \sum_{i=1}^{\infty} a_i p_i = \sum_{i=1}^{\infty} a_i P(\xi = a_i)$$

Если подумать, то написанное здесь говорит следующее. Мы должны каждое значение a_i взять с коэффициентом равным вероятности (частоте появления ответа в черном ящике) для этого значения. А потом все сложить. То есть мы получим какой ответ наш черный ящик дает в целом.

Математическое ожидание в непрерывном случае Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда определим ее математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \int_{-\infty}^{\infty} x p(x) dx$$

В данном случае физический смысл такой же как и в дискретном случае. Мы усредняем все возможные значения x , которые принимает наша случайная величина, учитывая каждый x пропорционально плотности $p(x)$. Чем больше плотность в точке x , тем больше вклад внесет x в финальное значение. Опять же, математическое ожидание показывает какой ответ в среднем нам будет давать наш черный ящик.

Сравнение формул для математического ожидания Давайте вспомним непонятную формулу:

$$\mathbb{E}(\xi) = \int_{\mathbb{R}} x dP_{\xi}$$

В случае дискретной случайной величины, надо заменить знак интеграла суммой. Переменная x в этом интеграле пробегает все значения на прямой, то есть ее надо заменить на значения случайной величины, то есть на a_i . А dP_{ξ} должна отвечать за вероятность принять значение a_i . Делая замену

$$\int_{\mathbb{R}} \mapsto \sum_{i=1}^{\infty} \quad x \mapsto a_i \quad dP_{\xi} \mapsto p_i$$

мы получим формулу для дискретного случая. В непрерывном случае, знак интеграла оставляем нетронутым, переменную x оставляем нетронутой, а вот символ dP_{ξ} меняем на $p(x) dx$, то есть замена вида

$$\int_{\mathbb{R}} \mapsto \int_{-\infty}^{\infty} \quad x \mapsto x \quad dP_{\xi} \mapsto p(x) dx$$

При такой подстановке, мы получим формулу для непрерывного случая. На самом деле есть общая математическая теория, которая позволяет все формулы считать одним способом, но нам придется немного помучиться от нашего незнания. Однако, не забывайте, что от знания мучаешься еще больше! Так что это приемлемое страдание.

Свойства математического ожидания

1. Пусть ξ – некоторая случайная величина и $a, b \in \mathbb{R}$, тогда $\mathbb{E}(a\xi + b) = a\mathbb{E}(\xi) + b$.
2. Пусть ξ – некоторая случайная величина, тогда $\mathbb{E}(\xi - \mathbb{E}(\xi)) = 0$.
3. Пусть ξ и η – две случайные величины, тогда $\mathbb{E}(\xi + \eta) = \mathbb{E}(\xi) + \mathbb{E}(\eta)$.

Дисперсия Отлично, мы заменили случайную величину ξ самым ее лучшим числовым приближением $\mathbb{E}\xi$. А на сколько отличается случайная величина от своего приближения? Очень просто, разница $\xi - \mathbb{E}(\xi)$ – это новая случайная величина, которая характеризует отличие. Что, опять случайная величина? А можно число? Просто взять матожидание от $\xi - \mathbb{E}(\xi)$ – не самая лучшая идея, ибо получится ноль, что означает, что в среднем ξ и $\mathbb{E}(\xi)$ не отличаются. А вот чтобы ответить, на сколько они в среднем отличаются надо вместо разности померить расстояние между ξ и $\mathbb{E}(\xi)$. Есть разные способы это сделать, но самый популярный – квадрат разности. А потом взять математическое ожидание.

Если ξ – некоторая случайная величина, то ее дисперсией называется следующее выражение

$$\mathbb{D}(\xi) = \mathbb{E}((\xi - \mathbb{E}(\xi))^2) = \mathbb{E}(\xi^2) - (\mathbb{E}(\xi))^2$$

Величина $(\xi - \mathbb{E}(\xi))^2$ показывает квадратичное отклонение матожидания от ξ . А значит, дисперсия показывает среднее квадратичное отклонение от матожидания. То есть это число, которое характеризует, на сколько в среднем ξ отличается от своего наилучшего приближения неслучайным числом. Проблема с дисперсией только в одном. Если ξ измерялась в мерах, то дисперсия будет измеряться в метрах в квадрате. Чтобы получить более разумную величину, определим стандартное отклонение следующим образом

$$\sigma(\xi) = \sqrt{\mathbb{D}(\xi)}$$

То есть это корень из среднеквадратичного отклонения. Вот эту величину можно рассматривать как чуть более адекватную меру отклонения от математического ожидания. Ее преимущество перед дисперсией в том, что она измеряется в тех же единицах, что и исходная случайная величина.

Дисперсия в случае дискретного распределения Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{Bmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{Bmatrix}$$

Тогда дисперсия считается так

$$\mathbb{D}(\xi) = \sum_{i=1}^{\infty} (a_i - m)^2 p_i, \text{ где } m = \mathbb{E}\xi = \sum_{i=1}^{\infty} a_i p_i$$

Дисперсия в случае непрерывного распределения Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда Дисперсия считается по формуле

$$\mathbb{D}(\xi) = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx, \text{ где } m = \mathbb{E}(\xi) = \int_{-\infty}^{\infty} x p(x) dx$$

Свойства дисперсии Если ξ – некоторая случайная величина, а $a, b \in \mathbb{R}$, тогда

$$\mathbb{D}(a\xi + b) = a^2 \mathbb{D}(\xi)$$

Моменты Если ξ – некоторая случайная величина, то можно определить ее k -ый момент следующим образом

$$\mathbb{M}_k(\xi) = \mathbb{E}(\xi^k)$$

То есть первый момент – это матожидание. Если матожидание нулевой, то второй момент – это дисперсия.

Моменты в дискретном случае Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{Bmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{Bmatrix}$$

Тогда моменты считаются по формуле

$$\mathbb{M}_k(\xi) = \sum_{i=1}^{\infty} a_i^k p_i$$

Моменты в непрерывном случае Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда моменты считаются по формуле

$$\mathbb{M}_k(\xi) = \int_{-\infty}^{\infty} x^k p(x) dx$$

Замечание про функции от случайной величины Пусть у вас есть случайная величина ξ и произвольная функция $f: \mathbb{R} \rightarrow \mathbb{R}$, тогда вы можете подставить случайную величину в функцию и получить новую случайную величину $f(\xi)$. Смысл этого действия – вы обработали исходные данные своего прибора, который был воткнут в черный ящик. Если вы хотите посчитать математическое ожидание от $f(\xi)$, то можно использовать следующую мнемоническую картинку

$$\mathbb{E}f(\xi) = \int_{\mathbb{R}} f(x) dP_{\xi}$$

которая в дискретном и непрерывном случае превращается в следующие формулы

$$\mathbb{E}f(\xi) = \sum_{i=1}^{\infty} f(a_i) p_i \quad \text{и} \quad \mathbb{E}f(\xi) = \int_{-\infty}^{\infty} f(x) p(x) dx$$

Эти формулы получаются из предыдущей непонятной заменами описанными в разделе про матожидание.

Нормальное или Гауссово распределение Есть в науке очень популярный вид распределения, который называется нормальным или гауссовым. Так получилось, что этот вид распределения моделирует много похожих процессов, например, аккуратность стрелка по мишени или случайный шум. Популярность гауссова распределения в основном связана с тем, что его математические свойства очень простые и изучены вдоль и поперек. А так как на глаз, его поведение адекватно отражает то, что нужно в приложениях, то почему бы его не использовать?

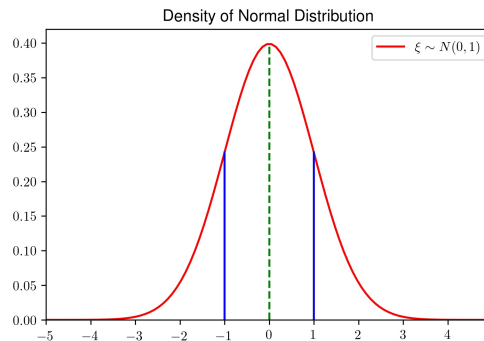
Гауссово распределение – это непрерывная случайная величина ξ на прямой \mathbb{R} задаваемая плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Когда задают гауссову случайную величину ξ с параметрами a и σ , обычно для краткости пишут $\xi \sim N(a, \sigma)$. Если посчитать числовые характеристики этого распределения, то выяснится

1. $\mathbb{E}(\xi) = a$.
2. $\mathbb{D}(\xi) = \sigma^2$ и в частности $\sigma(\xi) = \sigma$.

На следующей картинке изображен график плотности:



Пик плотности приходится на число a от которого плотность спадает в обе стороны экспоненциально. Думать про это надо так, мы с большей вероятностью отвечаем числа в окрестности точки a . У этого графика есть

две точки перегиба слева и справа от максимум, а именно, это точки $a - \sigma$ и $a + \sigma$. То есть среднее квадратичное отклонение указывает в точности на расстояние от максимума, до точек перегиба по бокам. Еще полезный факт, в интервале $[a - 3\sigma, a + 3\sigma]$ содержится 0,997 всей плотности, то есть вероятность $P(a - 3\sigma \leq \xi \leq a + 3\sigma) \geq 0,997$. На практике это означает, что вне этого интервала можно пренебречь значениями плотности и считать ее нулевой. Вот для ориентира еще несколько функций плотности для нормального распределения сосредоточенного в нуле, но разными дисперсиями.

