Наша цель — доказать, что если взять одинаково распределённые независимые случайные величины с конечной дисперсией, то зная только среднее и дисперсию можно довольно точно оценивать вероятность принятия средним из этих величин значений в заданном интервале.

Напомним, как устроены распределения вероятностей на множестве исходов $\mathbb{R}$ при нулевых вероятностях каждого их несчётного набора интересующих нас исходов.

Вспомним базовые свойства вероятности, которые нас интересуют.

Определение 1. Вероятность любого события неотрицательна. Вероятность достоверного события равна 1. Если событие является объединением не более, чем счётного числа непересекающихся событий, вероятности которых известны, то его вероятность определена и равна сумме их вероятностей.

Как мы помним, можно полностью задать это распределение неубывающей функцией $F(a) : F(a) = P(\{x \mid x < a\})$ (функцией распределения).

Теорема 1. При этом будет выполнено $P(\{x \mid x \leqslant a\}) = \lim\limits_{b \to a+} F(b)$.

Доказательство 1. Действительно,

$$\mathbb{R} \setminus \{x \mid x \leqslant a\} = \{x \mid x > a\} = \{x \mid x \geqslant a+1\} \cup [a + \frac{1}{2}; a+1) \cup \dots$$

. Это счётное объединение непересекающихся множеств, каждое из которых выражается через множества вида $\{x \mid x < c\}$. Частичные суммы ряда — это некоторая монотонная последовательность значений вида $P(\{x \mid x \geqslant c\})$, которые дополнительны до $F(c)$. Из-за монотонности такая последовательность значений позволяет найти и предел функции.

Далее через F удобно выражается вероятность попадания в любой промежуток.

Вероятность равенства случайной величины конкретному числу a равна $F(a) - \lim\limits_{b \to a-} F(b)$.

Функция распределения при конечном числе элементарных событий будет ступенчатой. Но для счётного числа элементарных событий она уже может иметь весьма сложный вид.

Собственно, теперь нам не важно, что сами исходы являются числами — мы построили способ задать вероятности в эксперименте, где элементарным событиям (в несчётном количестве) сопоставлены точки числовой прямой.

В частности, у нас может быть несколько независимых случайных величин на вероятностном пространстве с огромным количеством элементарных событий, на котором задана вероятность (то есть числовая функция, определённая на некоторых событиях, для которой выполнены базовые свойства). Тогда можно рассмотреть функцию распределения для одной из этих случайных величин.

Определение 2. Если $F(x) = \int\limits_{-\infty}^{x} f(x)dx$, то функция f называется *плотностью вероятности*.

Теорема 2. При сложении независимых случайных величин с плотностями вероятности f_1 и f_2 получается случайная величина с плотностью вероятности $x \mapsto \int\limits_{-\infty}^{\infty} f_1(t)f_2(x-t)dt$.

Доказательство 2. Проинтегрируем от $-\infty$ до некоторого y. У нас должна получится вероятность того, что сумма меньше y. С другой стороны, у нас получится интеграл по полуплоскости с границей под $45°$, и интеграл подынтегрального выражения по произвольному прямоугольнику со сторонами параллельно осям равен вероятности того, что каждое из слагаемых попало в интервал, соответствующий стороне этого прямоугольника. Такими прямоугольниками легко покрыть чуть больше нашей полуплоскости и чуть больше её дополнения со сколь угодно малым пересечением (для непрерывных f_1, f_2 это очевидно из-за уменьшения площади по которой мы интегрируем и стремления их к 0 при росте аргумента). Но тогда такие покрытия приближают наше событие (для непрерывных f_1, f_2 можно не заботиться про границу; в принципе, рассматривать надо прямоугольник без правой и верхней границы), а значения интеграла по ним приближают значение нашего интеграла. Отсюда получаем, что интеграл предполагаемой плотности вероятности задаёт нужную функцию распределения, что и требовалось.

Как теперь искать математическое ожидание?

Если распределению F вероятностей для x соответствует плотность f, и есть величина $\xi(x)$, то $M\xi = \int\limits_{-\infty}^{\infty} \xi(x)f(x)dx (= \int\limits_{-\infty}^{\infty} \xi(x)dF(x))$. При правильном понимании интеграла с $dF(x)$ та сумма просто окажется частным случаем написанного интеграла. Правильное понимание — что мера отрезка равна разности значений F на его концах.

Определение 3. Важным примером распределения с гладкой функцией распределения является нормальное распределение. Если его ожидание равно 0, а дисперсия равна 1, то его плотность равна $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Чтобы проверить, что его интеграл и правда единица, проще всего посмотреть на совместное распределение двух независимых нормальных случайных величин. Получится распределение, зависящее только от расстояния от начала координат, которое можно проинтегрировать с помощью перехода в полярные координаты.

Теорема 3. (Центральная предельная теорема, одна из ослабленных формулировок) Если есть счётное число независимых одинаково распределённых ограниченных случайных величин ξ_n с математическим ожиданием 0 и дисперсией 1, то вероятность попадания $\rho_n = \frac{\xi_1+\ldots+\xi_n}{\sqrt{n}}$ в любой заданный интервал стремится при $n \to \infty$ к вероятности попадания в этот интервал случайной величины, имеющей нормальное распределение.

Для доказательства этой теоремы определим вспомогательное понятие — характеристическую функцию распределения.

Определение 4. Для случайной величины ξ *характеристической функцией* называется $\chi_\xi(t) = Me^{it\xi}$.

Теорема 4. Характеристическая функция среднего n независимых ограниченных случайных величин, одинаково распределённых с ξ, имеющей математическое ожидание 0 и дисперсию 1, сходится при $n \to \infty$ к $e^{-\frac{t^2}{2}}$. При этом сходимость равномерная по t на каждом отрезке.

Доказательство 3. Из-за независимости ξ_k, и, следовательно, e^{ξ_k} выполняется:

$$\chi_{\frac{it(\xi_1+\ldots+\xi_n)}{\sqrt{n}}}(t) = Me^{\frac{it(\xi_1+\ldots+\xi_n)}{\sqrt{n}}} = \prod_{k=1}^{n} Me^{i\xi_k\frac{t}{\sqrt{n}}} = \prod_{k=1}^{n} \chi_{\xi_k}(\frac{t}{\sqrt{n}})$$

Мы получаем, что надо доказать $\chi_{\xi}^n(\frac{t}{\sqrt{n}}) \to e^{-\frac{t^2}{2}}$. Но в силу разложения экспоненты, а также известных $M\xi = 0$ и $M\xi^2 = D\xi = 1$ выполняется

$$(Me^{\frac{it\xi}{\sqrt{n}}})^n = (M(1 + it\frac{\xi}{\sqrt{n}} - \frac{1}{2n}t^2\xi^2 + O(n^{-\frac{3}{2}})))^n = (1 + 0 + \frac{1}{2n}t^2D\xi + O(n^{-\frac{3}{2}}))^n \to e^{-\frac{t^2}{2}}$$

В чём польза от характеристической функции? Характеристическая функция — ожидание мнимой экспоненты от $t\xi$, то есть математическое ожидание синуса и косинуса от значения ξ. Но через синусы и косинусы можно приблизить любую непрерывную функцию; а через ожидания кусочно-линейных функций, быстро возрастающих от нуля до единицы и через некоторое время быстро спадающих обратно, можно выразить ожидание индикатора промежутка, то есть вероятность попадания в него. Но это и есть то, что требовалось.

Самый простой способ показать, что именно нормальное распределение обладает указанной характеристической функцией, такой: предел не зависит от исходного распределения, а усреднение нормального распределения опять даёт нормальное распределение — тогда для этого распределения мы знаем предел.

На самом деле, одинаковость распределения не обязательна, как и ограниченность (достаточно существования среднего и дисперсии), но это требует более аккуратного рассмотрения.

В чём польза от предельной теоремы? Она позволяет сказать, что если у нас есть шум, то есть результат сложения большого количества независимых воздействий, то любые его отклонения от нормального распределения — повод для анализа. Если всё хорошо, то получится нормальное распределение. Если нормальное распределение не получилось — это повод искать систематическую погрешность. Например, если при измерении уровня жидкости студентами у распределения ответов два максимума, может оказаться, что одни измерили уровень нижней части мениска, образующегося при смачивании, а другие — верхней.

Это объясняет популярность метода наименьших квадратов — при его использовании максимизируется плотность вероятности около реально полученного значения при нормальном распределении ошибок относительно подбираемого идеала.

С помощью центральной предельной теоремы можно получить и оценку на биномиальные коэффициенты, близкие к середине. Если рассмотреть эксперимент с доской Гальтона (высыпание крупы или шариков на доску с вбитыми гвоздями), то одна из его разновидностей — это сложение независимых сдвигов на +1 и -1 с вероятностями по $\frac{1}{2}$. Тогда распределение результатов описывается и центральной предельной теоремой, и биномиальными коэффициентами. Эти описания должны быть согласованы.