1 Лекция 1

Зовут меня Валерий Николаевич Тутубалин. Из учебников рекомендую Б. Н. Севастьянов, Тутубалин, Тюрин, Макаров.

Говорят про вероятность наводнений, землетрясений — что-то физическое. Есть вероятность обвала рубля — чистая математика.

Какие события называются случайными. Сейчас будет ещё не определение, а мотивировка. Итак, случайности

(1) Воспроизводимость, массовость. Событие A может произойти большое число раз. Опыт можно сделать много раз, в теории бесконечное число раз.

(2) Статистическая устойчивость. Что это такое. Давайте нарисуем схему. Обозначим n число опытов, μ_A — число наступлений A. Величину $\frac{\mu_A}{n}$ называют частотой (экономисты называют частостью). На втором рисунке показана динамика частоты. В какой момент она стабилизируется? Есть некоторая *непредсказуемость*.

Величина частоты есть что-то вроде вероятности, то есть $\frac{\mu_A}{n} \approx \mathbb{P}(A)$.

Так математики, конечно, говорить не могут. Давайте попробуем ввести основные понятия.

Рассмотрим множество $\Omega = \{\omega\}$ — множество элементарных событий. Это множество возможных исходов опытов, которыми мы планируем заниматься. Полный список исходов. Мы сейчас предположим, что Ω не более чем счётно. Случай, когда Ω более чем счётно, относится к теории Колмогорова, которой мы в основном и будем заниматься. Но сейчас мы позанимаемся не более чем счётными. Отображение $\omega \to P(\omega) \geqslant 0$ — вероятность элементарного исхода. Если сложить все вероятности, получим (складываем частоты) $\sum \omega \in \Omega P(\omega) = 1$.

Аксиома 1.1. $\sum\limits_{\omega\in\Omega} \mathbb{P}(\omega) = 1.$

Определение 1.1. *Событием A называется любое подмножество Ω. Его верятностью называют* $\mathbb{P}(A) = \sum\limits_{\omega\in A} \mathbb{P}(\omega)$.

Рассмотрим пример: студенты на экзамене. Всего N билетов, столько же и студентов. n билетов счастливые для всех. Вероятность того, что первый счастливый: $\frac{n}{N}$. А то, что второй счастливый? Надо посмотреть как-то иначе.

Пусть $|\Omega| = N(\Omega) < \infty$. Пусть также события все одинаковые. Тогда $\mathbb{P}(\omega) = \frac{1}{N(\Omega)}$. Тогда $\mathbb{P}(A) = \frac{N(A)}{N(\Omega)}$.

Давайте оформим этот вывод как заклинание.

Что же считать элементарным событием в случае студентов? Пусть студенты будут занумерованы $1, 2, \ldots, N$, билеты будут подписаны $i_1, i_2, \ldots, i_N$. Обозначим

$$\begin{pmatrix} 1, & 2, & \ldots & , N \\ i_1 & i_2 & \ldots & i_N \end{pmatrix} = \omega$$

Всякая ω — подстановка (или перестановка). В качестве A_k будем брать события

$$A_k = \{k\text{-й счастливый}\} = \{\omega\colon i_k = 1, \ldots, n\}.$$

В данном случае $|\Omega| = N!$, $|A_k| = n(N-1)!$, $\mathbb{P}(A_k) = \frac{n(N-1)!}{N!} = \frac{n}{N}$.

То обстоятельство, что в данном случае $\mathbb{P}(\omega) = \frac{1}{N!}$, есть ли это математический феномен? Доказываемо ли это? Не смогли мы этого доказать. Вдруг лучшие билеты внизу стопки? Таким образом, такое предположение равновероятности элементарных событий — это экспериментальный факт. Эта информация берётся из частот.

Пусть $|\Omega| = N!$, а N большое, например, $N = 25$. Все частоты будут равны почти нулю. Это уже не экспиреметальный факт. Если билеты на столе раскладываются нечестно, то равномерного распределения не будет, но мы не знаем, как раскладывали. Мы сталкиваемся с тем, что выводы основаны непонятно на чём.

Пусть есть Ω, вероятности $\mathbb{P}(\omega)$, $A \subset \Omega$, $\mathbb{P}(A) = \sum \omega \in AP(\omega)$. Вот все наши величины.

Определение 1.2. *Противоположное событие $\overline{A}$ состоит из таких ω, что не входят в A, то есть это дополнение множества A:*

$$\overline{A} = \{\omega\colon \omega \not\subset A\}/$$

Заметим, что $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$. Ну это уже математическая теорема.

Пусть есть два события A и B. С двумя событиями можно много всего сделать.

Определение 1.3. $A \cup B = \{\omega\colon \omega \in A$ *или* $\omega \in B\}$ *— сумма событий A и B.*

Определение 1.4. $A \cap B = AB = \{\omega\colon \omega \in A$ *и* $\omega \in B\}$ *— пересечение событий.*

Что происходит с вероятностями $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$. Если события удовлетворяют $AB = \varnothing$, то $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Утверждение 1.1. *Пусть $A_1, \ldots, A_n, \ldots$ таковы, что $A_iA_j = \varnothing$ при $i \neq j$. Тогда*

$$\mathbb{P}(A_1 + A_2 + \cdots + A_n + \ldots) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Это будет верно так как $\omega \in A_i \Leftrightarrow \omega \notin A_j$ при $i \neq j$.

У Колмогорова это утверждение будет основной аксиомой.

Бывают ещё случаи, когда списков много. Пусть снова N студентов, N билетов. $p_1, \ldots, p_N$ должны в одном списке сходятся к $\frac{n}{N}$. Но с какой скоростью?

Займёмся другими вопросами. Пусть A и B — два способа изготовления цемента. Проводим опыт. Берём два куска цемента, засовываем в тиски. Кто не сломается, то победил. Пусть A классический способ изготовления цемента, B классический плюс дополнительный.

Будем проводить $n = 10$ попарных сравнений. Сосоставим событиям числа $A \sim 0$, $B \sim 1$. Тогда в наших обозначениях $\omega = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, $\varepsilon_i = 0, 1$. Чтобы B было лучше, нужно выпадание $\mu(1) = 7, 8, 9, 10$ побед B. Будем считать, что 6 — это ещё плохо, неубедительно. Будем говорить $A = B$, если все $\mathbb{P}(\omega)$ одинаковы и равны 2^{-10}. Если $A = B$, то $4 \leqslant \mu(1) \leqslant 6$ (никто не одержал 7 и более побед), это правильное решение.

$\mu(1) = \sum\limits_{i=1}^{10} \varepsilon_i$, $\mathbb{P}\{\omega\colon 4 \leqslant \mu \leqslant 6\} = \mathbb{P}\{\mu = 4\} + \mathbb{P}\{\mu = 5\} + \mathbb{P}\{\mu = 6\}$. При этом $\mathbb{P}\{\mu = k\} = C_{10}^k$ и $\mu(1) = \frac{672}{1024} \approx \frac{2}{3}$. Очень большая вероятность ошибочного решения в выборе между A и B.

А данное решающее правило, то есть $n = 10$ и $\mu \geqslant 7$ с какой вероятностью будет выдавать правильное решение?

Пусть вероятность $\mathbb{P}\{\text{победы } B\} \geqslant 0.7$, то есть $\mathbb{P}\{\varepsilon_1 = 1\} = 0.7$. А можем ли мы считать испытания независимыми (вскоре дадим определение независимости)?

1.1 Условная вероятность

Вероятность того, что случится событие B при условии того, что случилось событие A, обозначается черех $\mathbb{P}(B \mid A) = \mathbb{P}_A(B)$. Чтобы такую вероятность посчитать, надо взять события, в которых наступило A и вычислить частоту события B.

$$\frac{\mu(AB)}{\mu(A)} = \frac{\mu(AB)/n}{\mu(A)/n} \approx \frac{\mathbb{P}(AB)}{\mathbb{P}(A)}.$$

Такая вот мотивация к математическому определению условной вероятности.

Определение 1.5. $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)}$ *(*$\mathbb{P}(A) > 0$*). Или* $\mathbb{P}(AB) = \mathbb{P}(A) \cdot \mathbb{P}(A \mid |B)$ *(теорема об умножении вероятностей).*

Рассмотрим первых двух студентов: первый и второй. Может быть они оба счастливы (1 сч, 2 сч). Если первый уже вытащил счастливый билет, то мы знаем, сколько второму осталось счастливых билетов и сколько осталось билетов вообще. И условную вероятность посчитать проще чем вероятность пересечения. Используем это.

$$\mathbb{P}(1\text{сч } 2 \text{ сч}) = \mathbb{P}(1 \text{ сч})\mathbb{P}(2 \text{ сч} \mid 1 \text{ сч}) = \frac{n}{N}\frac{n-1}{N-1}.$$

Вероятность, что первый несчастен, а второй счасливый

$$\mathbb{P}(1 \text{ несч } 2 \text{ сч} = \frac{N-n}{N}\frac{n}{N-1}.$$

При этом $\mathbb{P}(2 \text{ сч}) = \frac{n(n-1)+n(N-n)}{N(N-1)} = \frac{n}{N}$.

Пусть в двух карманах 120, в левом вдвое больше, чем в правом.

$$x + y = 120, x = 2y.$$

В левом $2x$, в правом x. $2x + x = 120$. Началась какая-то математика.

2 Лекция 3

Чему мы научились за первые две лекции. Есть множество элементарных событий $\Omega = \{\omega\}$, на нём задана неотрицательная функция $\mathbb{P}(\omega)$, считаем $\mathbb{P}(A) = \sum\limits_{\omega \in A} \mathbb{P}(\omega)$. Тогда $\mathbb{P}(A+B) = \mathbb{P}(A) + \mathbb{P}(B)$, для непересекающихся событий.

Мы разбирали пример, когда есть множество людей. Вероятность наличия болезни $\mathbb{P}(\text{Б})$, $\mathbb{P}(\text{успешного лечения}) = 0.95$ — найти теоретически невозможно, только эмпирическими испытаниями. Было N больных, n неудач, оказалось, что $n/N = 0.95$. Дальнейшие рассуждения опирались на то, что вероятность выздороветь одинаковая. Но может быть и не так. Может быть так, что для 95% больных $\mathbb{P}(\text{успеха}) = 1$, для остальных — ноль. Отличить эти ситуации невозможно. Что мы получим при втором варианте? Те, кто выздоровел не будет жаловаться, и за счёт этого будет складываться соответствующее общественное мнение.

Надо изучать вопрос, насколько мы можем ошибиться, вычисляя предполагаемое значение, по отношению к истинному значению.

2.1 Случайные величины

Случайной величиной называется любая функция $\xi = \xi(\omega)$, определённая на Ω. Она может принимать действительные значения или комплексные и даже кватернионные, не важно. Напишем значения, которые принимает эта функция, строку, а под этими значениями вероятность того, что она их достигнет (вероятность прообраза). Такая таблица

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

где $p_i = \mathbb{P}(\xi = a_i) = \mathbb{P}\big\{\omega\colon \xi(\omega) = a_i\big\} = \sum\limits_{\omega\colon \xi(\omega)=a_i} \mathbb{P}(\omega)$, называется распределеним вероятности случайной величины ξ.

Хороша, когда есть Ω и все $\xi(\omega)$. Но часто бывает так, что $\xi(\omega)$ трудно задать на всех ω. Можно перейти от такой таблицы к некоторым параметрам, характеризующим случайную величину.

Одна из таких величин называется *математическим ожиданием* $\mathbb{E}$ от слова expectation. Определяется $\mathbb{E}$ следующим образом

$$\mathbb{E}\xi = \sum_{\omega\in\Omega} \xi(\omega)\mathbb{P}(\omega)$$

для тех случайных величин, для которых сумма абсолютно сходится. Имеются следующие свойства

$$\mathbb{E}(c\xi) = c\mathbb{E}\xi, \quad \mathbb{E}(\xi + \eta) = \mathbb{E}\xi + \mathbb{E}\eta.$$

Таким образом, $\mathbb{E}$ — это некоторый линейный функционал.

Лемма 2.1. *Пусть есть $\xi(\omega)$ — некоторая функция на Ω. Пусть также есть $\eta = f(\xi)$. Тогда можно образовать суперпозицию $\eta(\omega) = f\big(\xi(\omega)\big)$. Лемма утверждает, что*

$$\xi f(\xi) = \sum_a f(a_i)p_i = \sum_{a_i} f(a_i)\mathbb{P}\{\xi = a_i\}.$$

В частности, если $f(x) = x$, то $f(\xi) = \xi$ и $\mathbb{E}\xi = \sum\limits_{a_i} a_i p_i$.

Доказательство. Имеем

$$\mathbb{E}f = \sum_{\omega\in\Omega} f\big(\xi(\omega)\big)\mathbb{P}(\omega) = \sum_{\{a_i\}} \Bigg(\sum_{\omega\colon \xi(\omega)=a_i} f(a_i)\mathbb{P}(\omega) \Bigg) =$$

В каждом слагаемом внутренней суммы a_i фиксированно, значит, можно вынести.

$$= \sum_{a_i} f(a_i)\mathbb{P}\{\xi = a_i\}.$$

Что и требовалось доказать. ■

А не было ли что-то похожего в книгах по механике? Если я в точку a_i на прямой помещу массу p_i, то $\sum a_i p_i$ есть центр массы. Но раз есть центр массы, то есть и момент инерции. И он есть и имеет важнейшую роль в теории вероятности.

2.2 Дисперсия

Ужас в том, что хоть мат. ожидание по английски expectation, но дисперсия — variance.

Определение 2.1. *Дисперсией случайной величины ξ называется*

$$\mathbb{D}\xi(\xi - \mathbb{E}\xi)^2 = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \sum_{a_i}(a_i - E\xi)^2 p_i.$$

В механике такая величина называется моментом инерции. А такую величину в механике видали?

$$\sum_{a_i}(a_i - E\xi)^3 p_i$$

Я не видал.

2.3 Независимые случайные величины

У нас есть определение независимых событий. Там фигурировало понятие совместного наступления событий. Но случайные величины не наступают совместно и вообще не наступают. Нужно формулировать как-то по-другому.

Пусть A, B — числовые множества, ξ, η — случайные величины. Выделим таким события

$$\{\xi \in A\} = \big\{\omega \colon \xi(\omega) \in A\big\}, \quad \{\eta \in A\} = \big\{\omega \colon \eta(\omega) \in B\big\}.$$

Теперь важнейшее определение.

Определение 2.2. *Если для любых числовых множеств A и B события $\{\xi \in A\}$ и $\{\eta \in B\}$ независимы, то есть (запятая означает «и», то есть пересечение)*

$$\mathbb{P}\{\xi \in A, \eta \in B\} = \mathbb{P}\{\xi \in A\}\mathbb{P}\{\eta \in B\}.$$

Откуда же таким взяться. Если взять $\Omega = \Omega^{(1)} \times \Omega^{(2)} = \{\omega^{(1)}, \omega^{(2)}\}$ и такие случайные величины, которые формально зависят от всего $\omega = (\omega^{(1)}, \omega^{(2)})$, но по факту каждая только от одной компонетны, то есть $\xi = \xi(\omega^{(1)})$, $\eta = \eta(\omega^{(2)})$, то можно получить независимость. Пусть $\xi(\omega) = a_i$, $\eta(\omega) = b_j$. Вполне достаточно, чтобы такие события были независимы. Действительно,

$$\{\xi \in A\} = \bigcup_{a_i \in A} \{\xi = a_i\}, \quad \{\eta \in B\} = \bigcup_{b_j \in B} \{\eta = b_j\}.$$

И тогда вероятность $\mathbb{P}\{\xi \in A, \eta \in B\} = \sum\limits_{\substack{a_i \in A \\ b_j \in B}} \mathbb{P}\{\xi = a_i, \eta = b_j\}$. Абсолютная сходимость, которая есть из неотрицательности $\mathbb{P}$, гарантирует, что сумма произведений есть произведение сумм.

Пусть случайные величины ξ и η независимы. Построим новые случайные величины $f(\xi)$ и $g(\eta)$. Пусть $f(\xi \in A$ (эф от кси попало в A), это значит, что $\xi \in f^{-1}(A)$. Не факт, что новые случайные величины будет независимы.

Давайте выпишем: события A и B независимы, если $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

Лемма 2.2. *Пусть ξ, η независимы, а $\mathbb{E}\xi$ и $\mathbb{E}\eta$ существуют. Тогда*

$$\mathbb{E}(\xi\eta) = \mathbb{E}\xi\eta = \mathbb{E}\xi \cdot \mathbb{E}\eta.$$

Доказательство. Вспомним, что математическое ожидание можно написать следующим образом

$$\mathbb{E}f(\xi) = \sum_{a_i} f(a_i)\mathbb{P}\{\xi = a_i\}.$$

При замене a_i на векторы, пары чисел, что угодно, ничего в доказательстве меняться не будет.

Возьмём функцию $f(a, b) = ab$. Тогда

$$\mathbb{E}\xi\eta = \mathbb{E}f(\xi, \eta) = \sum_{(a_i, b_j)} a_i b_j \mathbb{P}\{\xi = a_i, \eta = b_j\} = \sum_{(a_i, b_j)} \mathbb{P}\{\xi = a_i\}\mathbb{P}\{\eta = b_j\} = \sum_{a_i} a_i \mathbb{P}\{\xi = a_i\} \sum_{b_j} b_j \mathbb{P}\{\eta = b_j\}.$$

Математик пришёл бы в ужас, что мы не наложили каких-то ограничений на величины. Но мы ответим: наложили. Мы сказали, что $\mathbb{E}\xi$ и $\mathbb{E}\eta$ существуют, а поэтому суммы сходятся абсолютно. ∎

А что будет с дисперсией любых двух случайных величин, не обязательно независимых.

$$\mathbb{D}(\xi + \eta) = \mathbb{E}\Big\{\big((\xi + \eta) - \mathbb{E}(\xi + \eta)\big)^2\Big\} = \mathbb{E}\big((\xi - \mathbb{E}\xi) + (\eta - \mathbb{E}\eta)\big)^2 =$$
$$= \mathbb{E}\big\{(\xi - E\xi)^2 + (\eta - \mathbb{E}\eta)^2 + 2(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)\big\} = \mathbb{D}\xi + \mathbb{D}\eta + 2\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta).$$

Последнее слагаемое нам не нравится, но никак от него не избавиться. Поэтому придётся ввести новое понятие.

Определение 2.3. *Величина*

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)$$

называется ковариацией.

Теперь можем записать дисперсию суммы

$$\mathbb{D}(\xi_1 + \dots + \xi_n) = \sum_i \mathbb{D}\xi_i + \sum_{i \neq j} \operatorname{cov}(\xi_i, \xi_j).$$

А если ξ и η независимы, то

$$\operatorname{cov}(\xi, \eta) = \mathbb{E}(\xi - \mathbb{E}\xi) \cdot \mathbb{E}(\eta - \mathbb{E}\eta) = 0.$$

Для независимых случайных величин, у которых существует дисперсия, дисперсия суммы есть сумма дисперсий.

2.4 Неравенство Чебышёва

Понятно, что дисперсия как-то измеряет отклонение случайной величины от математического ожидания. Чебышёв обнаружил полезное неравенство для тех случаев, когда дисперсия мала.

$$\mathbb{P}\big\{|\xi - \mathbb{E}\xi| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2}.$$

Давайте его докажем.

Доказательство. Обычно так доказывают.

$$\mathbb{D}\xi = \sum_{a_i} (a_i - \mathbb{E}\xi)^2 \mathbb{P}\{\xi = a_i\} \geqslant$$

мы эту сумму уменьшим, если складывать не по всем a_i, а по некоторым

$$\geqslant \sum_{a_i\,:\,|a_i - \mathbb{E}\xi| \geqslant \varepsilon} (a_i - \mathbb{E}\xi)^2 \mathbb{P}\{\xi = a_i\} \geqslant \sum_{a_i\,:\,|a_i - \mathbb{E}\xi| \geqslant \varepsilon} \varepsilon^2 \mathbb{P}\{\xi = a_i\} = \varepsilon^2 \mathbb{P}\big\{|\xi = \mathbb{E}\xi| \geqslant \varepsilon\big\}.$$

$\blacksquare$

2.5 Закон больших чисел в форме Чебышёва

Пусть имеется набор попарно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, причём пусть $\mathbb{D}\xi_i < C$. Тогда рассмотрим

$$\mathbb{P}\left\{\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \frac{\mathbb{E}\xi_1 + \mathbb{E}\xi_2 + \cdots + \mathbb{E}\xi_n}{n}\right| \geqslant \varepsilon\right\} \xrightarrow[n\to\infty]{} 0.$$

По неравенству Чебышёва

$$\mathbb{P}\big\{|\xi_n - \mathbb{E}\xi_n| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{D}\xi_n}{\varepsilon^2}$$

достаточно доказать, что $\mathbb{D}\xi_n \xrightarrow[n\to\infty]{} 0$.

$$\mathbb{D}\left\{\frac{\xi_1 + \cdots + \xi_n}{n}\right\} \leqslant \frac{1}{n^2} Cn = \frac{C}{n} \xrightarrow[n\to\infty]{} 0.$$

Пусть мы хотим узнать велицину a. Мы знаем $x_i = a + \delta_i$ — показания приборов. Пусть также $\delta_1, \delta_2, \ldots, \delta_n$ — независимые одинаково распределённые случайные величины, причём $\mathbb{E}\delta_i = 0$. Пусть $\sigma^2 = \mathbb{D}\delta_i$. Определим величину

$$\overline{x} = \frac{1}{n} \sum x_i = a + \frac{1}{n} \sum_{i=1}^{n} \delta_i.$$

Второе слагаемое имеет дисперсию $\frac{\sigma^2}{n}$.

Поскольку $\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2$, то можно оценить порядок отклонения $|\xi - \mathbb{E}\xi| \sim \sqrt{\mathbb{D}\xi}$. А если $\mathbb{E}\xi_i = a$, то модуль разности

$$\left|\frac{\xi_1 + \cdots + \xi_n}{n} - a\right|$$

мал, если n велико. Откуда взять σ^2? А вот

$$S^2 = \frac{1}{n} \sum (x_i - \overline{x})^2 \approx \sigma^2.$$

Часто также используют

$$s^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2.$$

Так как оценить отклонение. А вот так

$$\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}.$$

3 Лекция 4

Имеем случайную величину $\xi = \xi(\omega)$, заданную законом распределения

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

Здесь $p_i = \mathbb{P}\{\xi = a_i\} = \sum\limits_{\omega\colon \xi(\omega)=a_i} \mathbb{P}(\omega)$.

Ввели понятия математического ожидания

$$\mathbb{E}\xi = \sum_{a_i} a_i p_i,$$

дисперсии

$$\mathbb{D}\xi = \sum_{a_i} (a_i - \mathbb{E}\xi)^2 p_i.$$

Поскольку математическое ожидание определяется несколько иначе, а именно

$$\mathbb{E}\xi = \sum_{\omega\in\Omega} \xi(\omega)\mathbb{P}(\omega),$$

имеем аддитивность

$$\mathbb{E}(\xi+\eta) = \mathbb{E}\xi + \mathbb{E}\eta, \quad \mathbb{D}(\xi+\eta) = \mathbb{D}\xi + \mathbb{D}\eta + 2\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta).$$

Для независимых случайных величин дисперсия аддитивна и $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$.

Также обсудили неравенство Чебышёва

$$\mathbb{P}\big\{|\xi - \mathbb{E}\xi| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2}.$$

Определим $\sigma = \sqrt{\mathbb{D}\xi}$ стандартное отклонение. Пусть $\varepsilon = 3\sigma$. Тогда

$$\mathbb{P}\big\{|\xi - \mathbb{E}\xi| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2} = \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}.$$

Ещё доказали закон больших чисел, Пусть величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и $D\xi_i < C$. Тогда

$$\mathbb{D}\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{D}\xi_i \leqslant \frac{Cn}{n^2} = \frac{C}{n} \xrightarrow[n\to\infty]{} 0.$$

Определение 3.1. *Последовательность случайных величин η_k сходится к нулю по вероятности, если*

$$\forall\ \varepsilon > 0 \ \ \mathbb{P}\big\{|\eta_n| > \varepsilon\big\} \xrightarrow[n\to\infty]{} 0.$$

Также определяется $\eta_n \xrightarrow[n\to\infty]{} \zeta$, если $\eta_n - \zeta \xrightarrow[n\to\infty]{} 0$ по вероятности.

Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимые одинаково распределённые случайные величины, причём $\mathbb{E}\xi_n = a$. Тогда

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - a \xrightarrow[n\to\infty]{\text{(вер.)}} 0.$$

Будем обозначать $\frac{\xi_1 + \dots + \xi_n}{n} = \overline{\xi}$. Тогда

$$\mathbb{E}f(\xi_i) \approx \frac{f(\xi_1) + \dots + f(\xi_n)}{n}.$$

Здесь примерное равенство означает, что вероятность любой ошибки стремится к нулю.

$$\sigma^2 = \mathbb{D}\xi_i = \mathbb{E}(\xi_i - a)^2 \approx \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2.$$

Пусть a — измеряемая величина. Делаем n измерений $x_1, x_2, \dots, x_n$. Тогда

$$a \approx \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Как тогда оценить $|\overline{x} - a|$?

Делаем предположение, что $x_i = a + \delta_i$, причём δ_i — независимые одинаково распределённые случайные величины, для которых $\mathbb{E}\delta_i = 0$, то есть нет систематической ошибки. Удалять систематические ошибки эксперимента должен физик без всякой математики. А вот то, что это случайные величины, комментировать сложнее. Это же конкретные числа, вовсе не случайные. Здесь мы возносимся ввысь и говорим, что будем рассматривать x_i как реализации n независимых случайных величин. Если это случайность, то должен быть ансамбль опытов. $x_1, \ldots, x_n$ — это не n опытов, а один большой опыт. Зделали бы ещё опыт, были бы числа $y_1, \ldots, y_n$, а ещё потом $z_1, \ldots, z_n$. Этих чисел y_i и z_i у нас нет, но мы можем их выдумать. Если Бога нет, то следует его выдумать.

Наблюдаем $\overline{x} = a + \frac{1}{n}\sum\limits_{i=1}^{n} \delta_i$. Значит, $\overline{\delta} \xrightarrow[n\to\infty]{\text{(вер.)}} 0$. Закон больших чисел даёт надежду на то, что $\overline{\delta}$ будет чем-то вроде нуля с большой вероятностью, то есть в большинстве случаев. То есть среди ансамблей x_i, y_i, z_i, … мало случаев, когда $\overline{\delta}$ сильно отличается от нуля. Но почему именно наш ансамбль иксов оказался хорошим? Надо оценить, насколько большой может быть всё-таки $\overline{\delta}$.

Обозначим $\mathbb{D}\delta_i = \sigma^2$, $\sqrt{\mathbb{D}\delta_i} = \sigma$. Пусть каким-то чудом σ нам известна. Тогда

$$\mathbb{D}\overline{\delta} = \mathbb{D}\left(\frac{\delta_1 + \cdots + \delta_n}{n}\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Дисперсия среднего арифметического в n раз меньше дисперсии слагаемых. Стандартное отклонение — в $\sqrt{n}$ раз. Тогда

$$\mathbb{P}\left\{|\overline{\delta}| > \frac{3\sigma}{\sqrt{n}}\right\} \leqslant \frac{\mathbb{D}\overline{\delta}}{9\frac{\sigma^2}{n}} = \frac{1}{9}.$$

Это если σ нам известна. Считаем $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$. Но наука в лице Гаусса сказал, что лучше использовать

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \approx \sigma^2.$$

Таким образом, сделав измерения $x_1, \ldots, x_n$, вычислим $\overline{x}$ и $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$. Тогда

$$|\overline{x} - a| \leqslant \frac{3s}{\sqrt{n}}$$

На это можно надеяться.

Закон больших чисел мы в прошлый раз доказывали. Но то, что мы делаем сегодня, это, конечно, не математическая теорема.

Пусть есть $\overline{x}$ и $\overline{y}$. Имеем $\mathbb{E}\overline{x} = a = \mathbb{E}\overline{y}$. Значит, (допустим, что всё-таки дисперсии разные)

$$\mathbb{E}(\overline{x} - \overline{y}) = 0, \quad \mathbb{D}(\overline{x} - \overline{y}) = \mathbb{D}\overline{x} + \mathbb{D}\overline{y} = \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n} \approx \frac{s_x^2}{m} + \frac{s_y^2}{n}.$$

Тогда давайте подберём ε, чтобы следующая вероятность была оценена сверху числом $\frac{1}{9}$.

$$\mathbb{P}\big\{|\overline{x} - \overline{y}| > \varepsilon\big\} \leqslant \frac{\mathbb{D}(\overline{x} - \overline{y})}{\varepsilon^2} = \frac{1}{9}.$$

Отсюда $\varepsilon = 3\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}$. Мы можем и убрать оттуда тройку. Ввести величину $\gamma = \sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}$. Если $|\overline{x} - \overline{y}| \leqslant \gamma$, хорошо. Это ещё один способ взглянуть на наши измерения.

Чтобы закончить введение в теорию вероятности, надо сделать ещё немного. Обозначение для распределения веротности

$$\begin{pmatrix} a_1 & a_2 & \ldots \\ p_1 & p_2 & \ldots \end{pmatrix}$$

годится не всегда.

3.1 Предельная теорема Пуассона

Мы с вами изучали испытания Бернулли. Пусть есть n испытаний, μ — успехов, p — вероятность успеха. Тогда

$$\mathbb{P}\{\mu = m\} = C_n^m p^m q^{n-m}, \quad q = 1 - p.$$

На компьютере, конечно, это считать можно. Но вообще не особо наглядно получается задавать эту случайную величину таблицей.

Запишем схему серий: число испытаний, вероятность успеха, количество успехов

$$\begin{array}{ccc} n=1 & p_1 & \mu_1 \\ n=2 & p_2 & \mu_2 \\ \vdots & \vdots & \vdots \\ n & p_n & \mu_n \end{array}$$

Пуассон утверждает, что удобно устремить $np_n \to \lambda > 0$, то есть $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$.

$$\mathbb{P}\{\mu = m\} = \mathbb{C}_n^m p_n^m (1-p_n)^{n-m} = \frac{n(n-1)\dots(n-m+1)}{m!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^n \left(1 - \frac{\lambda}{n} - o\left(\frac{1}{n}\right)\right)^{n-m} =$$
$$= \frac{n(n-1)\dots(n-m+1)}{m!}\left(\frac{\lambda^m}{n^m} + \dots\right)e^{-\lambda} = \frac{\lambda^m}{m!}e^{-\lambda}.$$

Зададим случайную величину $\mu = 0, 1, 2, \dots$, $\mathbb{P}\{\mu = m\} = \frac{\lambda^m}{m!}e^{-\lambda}$, она называется законом Пуассона. Легко посчитать

$$\mathbb{E}\mu\lambda, \quad \mathbb{D}\mu = \lambda.$$

Нам нужен закон Пуассона, чтобы какие-то вероятности относительно легче считались. Давайте применим закон Пуассона.

3.2 Проверка статистических гипотез

Статистическая гипотеза — любое положение, позволяющее считать какие-то вероятности.

Имеется множество Ω событий. Опыт заканчивается результатом $\omega \in \Omega$. То, чем опыт кончается, нас не волнует. Пусть $X = \{x\}$ — множество наблюдаемых вещей. Есть какое-то отображение $\Omega \to X$. Гипотеза задаёт правило, по которому считаются вероятности вида

$$\mathbb{P}\{x \in A \subset X \mid H\}.$$

Как проверить, что эта гипотеза верна. Выберем $S \subset X$, для которого $\mathbb{P}\{x \in S \mid H\} \leqslant \alpha$. Число α называют в этом контексте уровнем значимости. Возможны случаи

(1) $x \in S$. Тогда гипотеза H отклоняется на уровне значимости α.

(2) $x \notin S$. Тогда гипотеза H не отклоняется на уровне значимости α.

Рассмотрим пример. На 1 000 булочек уходит 10 000 изюминок. Производитель так объявил, это его реклама. Но работники фабрики могут иметь способ воровать изюминки. Нужно найти способ понять, воруют или нет. Обобзначим H_a для $0 \leqslant a \leqslant 1$ гипотезу «доля a украдена».

Возьмём булочку, обозначим μ — число изюменок в ней. Пусть у нас 10 000 испытаний, то есть перебираем изюминки. Вероятность, что данная изюминка попала в данную булочку $p = \frac{1}{1\,000}$. Числа большие, это случай пуасона: $\lambda = np = 10(1-a)$.

Будем брать булочки. Пусть x — число изюминок в будлочке. Пусть $S = \{x = 0\}$. Если H_0 верна, то вероятность, распределённая по Пуассону,

$$\mathbb{P}\{x = 0\} = \approx 0.5 \cdot 10^{-4}.$$

Пусть всякий человек, купивший булочку без изюма, пишет жалобу. А принимающий такую жалобу тут же закрывает фикрму. Вероятность, что жалобу напишет конкретный человек, 10^{-4}. А вероятность, что напишут хотя бы одну жалобу, 0.05 (ну тысяча булочек). Но булки-то пекутся каждый день. В году вероятность, что не будет ни одной жалобы $(1-0.05)^{365} = 0$ (ну это в общем нуль). Булочки печь будет уже некому.

4 Лекция 5

Мы сегодня заканчиваем введение в теорию вероятностей. Главная цель введения: показать, что вероятность это не совсем то, что подсказывает нам интуиция. На примере медицинских исследований: что может сказать врач больному о лекастве? Что среди всех, кто принимал эти лекарства 96% выздоровели. Но это не значит, что данный больной будет излечен с вероятностью 95 %. Мы установили, что

$$\mathbb{P}\left\{|\overline{x}-a| \leqslant \frac{3s}{\sqrt{n}}\right\} \geqslant \frac{8}{9}, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

Позанимаемся логикой. Пусть все A суть B, все B суть C. Отсюд вывод: все A суть C. Теперь немного это изменим. Пусть все A суть B и некоторые B суть C. Вытекает ли отсюда, что некоторые A суть C? Да нет же. (Например $A \subset B$ и $A \cap C = \varnothing$.)

Пусть если гипотеза H верна, то S невозможно. Пусть также S в опыте наступило. Отсюда какой вывод? Вывод: H неверна. Я прошу вас заметить, что если мы попадаем в вероятностную ситуацию, такой логикой мы пользоваться не будем. Напишем следующим образом: пусть если H неверна, то S маловероятна; пусть также S в опыте наступило. Отсюда хочется сказать: H маловероятно. Но почему? На самом деле так не получится. Мы поэтому вводим понятие значимости статистической гипотезы.

Пусть есть $\Omega = \{\omega\}$. Но сами ω не наблюдаются. Но есть отображение $\Omega \to X = \{x\}$, причём иксы уже наблюдаются. Кроме того, пусть мы знаем $\mathbb{P}\{x \in A \subset X \mid H\}$ (черта заменяет слова: «если верна H»). Формально, уровень значимости α — это число из $[0,1]$. Мы это число задаём заранее, то есть до эксперимента. Чаще всего в качестве α выбирают одно из трёх: $0.05, 0.01, 0.001$. Для этого α находим множество $S \subset X$, для которого

$$\mathbb{P}\{x \in S \mid H\} \leqslant \alpha.$$

При эксперименте у нас либо x попало в S, либо нет. Если $x \in S$, то говорят, что гипотеза H отклоняется на уровне значимости α. Если же $\alpha \notin S$, то H не отклоняется.

α — ограничение сверху вероятности зря обидеть нашу гипотезу.

На примере булочек. Было у нас 1 000 булочек, 10 000 изюминок. x изюминок в эксперименте. Мы рассуждали в терминах $S = \{x = 0\}$. Тогда если $x \in S$, то покупатель идёт жаловаться. Пусть $\lambda_0 = 10$, $0 \leqslant a \leqslant 1$, $\lambda_a = 10(1-a)$. Вероятность обвинения

$$\mathbb{P}\{x \in S \mid H_0\} \approx 0.5 \cdot 10^{-4}.$$

Гипотезу мы сейчас обозначим. Пусть A_k — событие типа k-й недоволен. Тогда

$$\mathbb{P}(A_1 + A_2 + \dots + A_{1\,000}) \approx \sum \mathbb{P}(A_i).$$

Вероятность того, что k-й недоволен есть 0.05. Тогда вероятность не получить жалобу за год $(0.95)^{365} \approx 0$.

Так вот обозначаем H_0 — основная гипотеза (ничего не воруют). У нас целое семейство гипотез H_a, $0 \leqslant a \leqslant 1$. Естественно жаловаться когда изюминок не ноль, но слишком мало, то есть $S = \{x \leqslant k\}$. Вероятность ограничена уровнем значимости $\mathbb{P}\{x \in S \mid H_0\} \leqslant \alpha$. $\mathbb{P}\{\mu = x\} = \frac{\lambda^x}{x!}e^{-\lambda}$.1

Выпишем таблицу вероятности обвинения неповинного человека, который на самом деле ничего не воровал.

Доля a	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.15$
	$S = \{x \leqslant 3\}$	$S \leqslant x \leqslant 13\}$	$S = \{x \leqslant 6\}$
0.0	0.020	0.00050	0.13
0.1	0.021	0.0013	
0.2	0.042	0.0031	
0.5	0.27	0.041	0.76
0.7	0.65	0.20	
1	1	1	

Здесь у нас выписана некая функция, обозначим её $\beta(a)$. Назовём функцией мощности.

Поставим камеры видеонаблюдения. Пусть они показали, что 80 % честных людей. Из 0.8 честных 0.13 по нашему критерию обвинения на основе изюминок будут объявлены виновными.

Если будем слишком решительно поступать на основе статистики, можем прийти вот к такому абсурду. На этом введение в теорию вероятностей закончено.

5 Аксиматика Колмогорова

Рассмотрим мысленный эксперимент. На отрезок $[0,1]$ случайно бросается точка. Как ввести пространство событий и как ввести вероятность? Проще склеить отрезок в окружность, вбить в центр окружности гвоздить и прицепить туда стрелку. Где стрелка остановилась, то число на отрезке и выпало.

Так что же будет в роли ω. Для любого физического опыта ω можно считать рациональным числом. Но что такое $\mathbb{P}(\omega)$? Они должны быть все одинаковы, если чисто случайно бросаются точки. Если $\mathbb{P}(\omega) > 0$, то $\sum\limits_{\omega \in \mathbb{Q}} \mathbb{P}(\omega) = \infty$; а если $\mathbb{P}(\omega) = 0$, $\sum \mathbb{P}(\omega) = 0$. Всё это нехорошо.

Тогда решаемся сделать следующее. Пусть $\{\omega\}$ — все вещественные числа на отрезке, а вероятность определим $\mathbb{P}\{\omega \in [\alpha,\beta]\} = |\alpha-\beta|$ как длину отрезка. Теперь обратите внимание, что здесь потребуем.

У нас будет Ω — множество элементарных событий. Есть какие-то подмножества $A_i \subset \Omega$, обладающие вероятностью, то есть на которых определена $\mathbb{P}$. Так вот мы потребуем счётной аддитивности

$$\mathbb{P}(A_1 + A_2 + \dots) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Мы в качестве $\mathbb{P}$ на отрезке будем брать меру Лебега. Есть класс подмножеств $\mathfrak{B}$ этого отрезка, на котором мера Лебега определена.

В общем случае $\Omega = \{\omega\}$ — какое-то множество (в весеннем семетре, это будут, например, функции вещественной переменной), $\mathfrak{B}$ — сигма-алгебра подмножеств Ω. Что за сигма-алгебра:

(1) $\varnothing, \Omega \in \mathfrak{B}$;

(2) если $A \in \mathfrak{B}$, то $\overline{A} = \Omega \setminus A \in \mathfrak{B}$;

(3) если $A_1, A_2, \dots, A_n, \dots \in \mathfrak{B}$, то и $\bigcup_i A_i, \bigcap i A_i \in \mathfrak{B}$.

И есть у нас третий объект: счётно аддитивная функция на $\mathfrak{B}$: для непересекающихся A_i имеем $\mathbb{P}(A_1+A_2+\dots) = \sum \mathbb{P}(A_i)$.

Определение 5.1. *Тройка $(\Omega, \mathfrak{B}, \mathbb{P})$ называется вероятностным пространством.*

Элементы $\mathfrak{B}$ называют событиями.

Вот у нас $[\alpha,\beta] \in \mathfrak{B}$. Давайте введём понятие наименьшей σ-алгебры. Возьмём все σ-алгебры, содержащии все отрезки. Хотя бы одна такая есть: все подмножества. Наименьшей назовём пересечения всех таких σ-алгебр. Наименьшая σ-алгебра, содержащая отрезки, называется σ-алгеброй борелевских подмножеств.

Что входит в борелеские множества? Одноточечные множества $\{\omega\}$ входят, канторовское множество входит. Пример неборелевского множества вообще сложно привести, хотя и можно. Раз входят все точки, то вероятность точки есть нуль. $\mathbb{P}\{\text{рациональна}\} = 0$, $\mathbb{P}\{\text{иррациональна}\} = 1$. Физику это понравится? Это недостаток такой аксиоматики.

Теперь обсудим, зачем нам счётная аддитивность? Если в дискретной теории мат ожидания считаются через ряд, а у нас тут будут некие интегралы. Для интеграла нужна счётная аддитивность меры. Чтобы считать интегралы, надо обсудить, какие будем брать функции.

Определение 5.2. *Случайной величиной называется любая измеримая функция $\xi(\omega)\colon \Omega \to \mathbb{R}$.*

Что это значит. На Ω есть какая-то σ-алгебра $\mathfrak{B}$, на числовой прямой $\mathbb{R}$ есть борелевская σ-алгебра. Функция измерима, если для любого B борелевского $\xi^{-1}(B) = \{\omega\colon \xi(\omega) \in B\} \in \mathfrak{B}$.

Давайте покажем, как можно построить интеграл Лебега. Пусть $\xi(\omega)$ — измеримая функция, принимающая не более чем счётное число значений. Тогда её можно записать в виде

$$\xi(\omega) = \sum_{i=1}^{\infty} a_i J_{A_i}(\omega), \qquad A_iA_j = \varnothing \Leftarrow i \neq j, \qquad A_1 + A_2 + \dots = \Omega.$$

Тогда

$$\mathbb{E}\xi = \int_{\Omega} \xi(\omega)\,\mathbb{P}(d\omega) = \sum_{\{A_i\}} a_i \mathbb{P}(A_i).$$

Это то, что мы уже писали. Формула имеет смысл, если ряд сходится. Единственное, что здесь может смущать, это независимость мат ожидания от выбора разбиения $\{A_i\}$, с помощью которого мы представляем $\xi(\omega)$. (Мы же можем разбить одно A_i на несколько частей.)

Пусть $\xi(\omega) = \sum\limits_{\{A_i\}} a_i J_{A_i}(\omega) = \sum\limits_{\{B_i\}} b_i J_{B_i}(\omega)$. Тогда введём множества $D_{ij} = A_iB_j$. Некоторые из них могут быть пустыми, неважно. Тогда

$$\xi(\omega) = \sum_{D_{ij}} d_{ij}\mathbb{P}(A_iB_j) = \sum_{\{A_i\}} a_i \sum_{\{B_j\}} \mathbb{P}(D_{ij}) = \sum_{\{A_i\}} a_i \mathbb{P}(A_i).$$

Ну хорошо. Тогда возникает мысль, что интеграл обладает свойством линейности, а именно $\mathbb{E}(\xi+\eta) = \mathbb{E}\xi + \mathbb{E}\eta$. Для счётнозначных случайных величин это легко уже можем проверить. А также $\mathbb{E}c\xi = c\mathbb{E}\xi$.

Теперь нужно с помощью предельного перехода определить мат ожидание для произвольной случайной величины, конечно, измеримой.

Если $\bigl|\xi(\omega)\bigr| \leqslant \varepsilon$, то и $|\mathbb{E}\xi| \leqslant \varepsilon$. Это очевидно, так как сумма вероятностей элементов разбиения равна единице. Если $\xi_n \underset{n\to\infty}{\rightrightarrows} \xi$, причём ξ измерима, то можно сделать предельный переход под интегралов. Давайте посмотрим

$$|\mathbb{E}\xi_n - \mathbb{E}\xi_m| = \bigl|\mathbb{E}(\xi_n - \xi_m)\bigr| \xrightarrow[n\to\infty]{} 0.$$

Пусть $\xi(\omega)$ измерима. Определим $\xi_n(omega) = \frac{k}{n}$, подобрав для каждого n k так, чтобы $\frac{k}{n} \leqslant \xi(\omega) \leqslant \frac{k+1}{n}$. Тогда $|\xi - \xi_n| \leqslant \frac{1}{n}$ и

$$\mathbb{E}\xi = \int\limits_{\Omega} \xi(\omega)\mathbb{P}(d\omega) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mathbb{P}\left\{\frac{k}{n} \leqslant \xi(\omega) < \frac{k+1}{n}\right\}.$$

Не забудем следущее: предел функций $\xi_n(\omega)$ должен быть измерим.

Если мат ожидания вычислять таким предельным переходом, мы далеко не уйдём. На следующей лекции расскажу, как их считать.

6 Лекция 6

Мы дожили до такого момента, когда нам придётся заниматься математикой. Вспомнили, что такое интеграл Лебега и определили мат. ожидание

$$\mathbb{E}\xi = \int\limits_{\Omega} \xi(\omega)\mathbb{P}(d\omega) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mathbb{P}\left\{\omega\colon \frac{k}{n} \leqslant \xi(\omega) < \frac{k+1}{n}\right\}.$$

Но считать такой предел почти невозможно, нужно перейти к каким-то другим понятиям.

Пусть есть случайная величина $\xi = \xi(\omega)$ $(\xi^{-1}(B) \in \mathfrak{B})$, есть $f(x)$ измеримая по Борелю (то есть аргументы $x \in \mathbb{R}$ действительные, значения $f(x) \in \mathbb{R}$ тоже). Придумать действительную неизмеримую по Борелю функцию достаточно сложно. Кто забыл, $f^{-1}(B)$ борелевское, если B борелевское, — выполнение этого означает измеримость. В наших предположениях $f(\xi) = f\big(\xi(\omega)\big) = \eta(\omega)$. Тогда

$$\Big\{\omega\colon f\big(\xi(\omega)\big) \in B\Big\} = \big\{\omega\colon \xi(\omega) \in f^{-1}(B)\big\} \in \mathfrak{B}.$$

И теорема вот какая:

$$\mathbb{E}f(\xi) = \int\limits_{-\infty}^{+\infty} f(x)\,\mu_\xi(dx), \qquad \mu_\xi(B) = \mathbb{P}\big\{\omega\colon \xi(\omega) \in B\big\} = \mathbb{P}\big\{\xi^{-1}(B)\big\}.$$

Новая мера $\mu_\xi(B)$ тоже σ-аддитивна. Она носит название распределения вероятности случайной величины ξ.

Очень редко задаются в приложениях зависимости $\xi(\omega)$, а вот распределения вероятностей встречаются довольно часто. Так вот что утверждается

$$\mathbb{E}f(\xi) = \int\limits_{\mathbb{R}^1} f(x)\mu_\xi(dx).$$

Обратите внимание на логику. μ_ξ — это вероятностная мера. Интеграл Лебега для неё заново определять нам не нужно. Математическое ожидание функции от случайной величины равно интегралу от этой функции по распределению вероятности случайной величины.

Многие студенты в этой формуле потают, пишут, например, $f(\omega)$, хотя f — функция вещественного аргумента и в этом суть.

Такой интеграл тоже никто вычислять не умеет. Но это наш первый шаг. В некоторых случаях его вычислить всё-таки можно.

Доказательство. Доказательство вот к чему сводится.

$$\mathbb{E}f(\xi) = \int\limits_{\Omega} f\big(\xi(\omega)\big)\mathbb{P}(d\omega) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mathbb{P}\left\{\omega\colon \frac{l}{n} \leqslant \xi(\omega) < \frac{k+1}{n}\right\} =$$

$$= \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mathbb{P}\left\{f\big(\xi(\omega)\big) \in \left[\frac{k}{n}, \frac{k+1}{n}\right)\right\} = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mathbb{P}\left\{\xi(\omega) \in f^{-1}\left[\frac{k}{n}, \frac{k+1}{n}\right)\right\} =$$

$$\lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n}\mu_\xi\left\{x\colon f(x) \in \left[\frac{k}{n}, \frac{k+1}{n}\right)\right\} = \int\limits_{\mathbb{R}^1} f(x)\,\mu_\xi(dx).$$

■

Случайнвя величина $\xi = \xi(\omega)$, её распределение вероятностей $\mu_\xi(B) = \mathbb{P}\{\xi \in B\} = \mathbb{P}\big\{\xi^{-1}(b)\big\}$. Для того, чтобы понять как считать интеграл по распределению вероятности, введём понятие плотности. На прямой есть замечательная мера dx, которая является мерой длины, для объединения отрезков это знали ещё в античную эпоху. Она продолжается на сложные множества. Давайте попробуем выписать меру μ выраженную через dx.

Определение 6.1. *Пусть $\mu(B) = \int\limits_B \rho_\xi(x)\,dx = \mathbb{P}\{\xi \in B\}$. Тогда $\rho_\xi(x)$ называется плотностью распределения вероятностей.*

Вероятность того, что ξ попала в B, можно переписать как интеграл по всему объемлющему множеству, используя индикатор $J_B(x)$.

$$\mu_\xi(B) = \mathbb{P}\{\xi \in B\} = \int\limits_{-\infty}^{+\infty} J_B(x)\rho_\xi(x)\,dx = \int\limits_{-\infty}^{+\infty} J_B(x)\,\mu_\xi(dx).$$

Мы видим, что если в качестве функции f подставить индикатор $J_B(x)$, то интеграл по $\rho(x)dx$ и интеграл по мере $\mu_\xi(dx)$ совпадают. Таким образом мы можем и для простых функций (линейных комбинаций индикаторов)

$$\int\limits_{-\infty}^{+\infty} f(x)\rho_\xi(x)\,dx = \int\limits_{-\infty}^{+\infty} f(x)\,\mu_\xi(dx).$$

Раз для простых, то и для измеримых существование плотности понятно. Совсем плохих функций в приложениях не бывает.

Всё изложено в предположении, что все интегралы и суммы сходятся в смысле абсолютной сходимости. Лебеговские интегралы считаются существующими, если сходятся в абсолютном значении.

Возьмём частный случай $f(x) = x$.

$$\mathbb{E}f(\xi) = \mathbb{E}\xi = \int\limits_{\mathbb{R}} x\rho_\xi(x)\,dx.$$

Очень понятная формула. Вопрос: а почему же надо таскать с собой все эти σ-алгебры, почему нельзя определить мат. ожидание сразу через плотность? Так поступить математик абсолютно не может вот по какой причине. Давайте случайную величину $f(\xi)$ обозначим через η. У случайной величины тоже есть плотность $\rho_\eta(x)$. Тогда

$$\mathbb{E}f(\xi) = \int\limits_{\mathbb{R}} f(x)\,\mu_\xi(dx) = \int\limits_{R} f(x)\rho_\xi(x)\,dx.$$

С другой стороны

$$\mathbb{E}\eta = \int\limits_{\mathbb{R}} x\rho_\eta(x)\,dx.$$

Нам надо доказать такое правило обращения с интегралами

$$\int\limits_{\mathbb{R}} f(x)\,\rho_\xi(x)\,dx = \int\limits_{\mathbb{R}} x\,\rho_{f(\xi)}\,dx.$$

Нет такого правила в математическом анализе, но оно верно. Для доказательство этого правила нам и нужен формализм Колмогорова.

Чего нет в математических учебниках: плостность распределения — величина размерная. Математики обычно считают всё по умолчанию безразмерным, на практике вы можете действовать и не так. Однако всегда можно от размерности освободиться в самом начале.

Пусть $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$. А функция $f(x) \in \mathbb{R}^1$, но $x = (x_1, \ldots, x_n)$. Изменится ли что-нибудь в нашей формуле?

$$\mathbb{E}f(\xi) = \int\limits_{\mathbb{R}^n} f(x)\,\mu_\xi(dx) = \int\limits_{\mathbb{R}^1} f(x)\,\rho_\xi(x)\,dx.$$

Плотность теперь надо понимать не по длине, а по объёму.

Пусть ещё $\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2 = 1$, то есть $(\xi_1, \ldots, \xi_n) \in S^{n-1}$. Но ведь на сфере есть замечательная мера. И снова ничего не меняется в наших формулах. Это большое преимущество.

Если $\rho_\xi(x)$ кусочно-гладкая, имеем

$$\rho_\xi(x) = \lim_{O(x)\to\{x\}} \frac{\mathbb{P}\big\{\xi \in O(x)\big\}}{V\big(O(x)\big)}.$$

Математические науки как устроены: есть несколько основных формул. Из них выводится множество других более узких. В теории вероятностей таких основных формул всего две.

Пусть имеются две области X и Y в $\mathbb{R}^n$. Пусть имеется гладкое взаимнооднозначное отображение $y = f(x)$, $f = (f_1, f_2, \dots, f_n)$. Пусть случайная величина ξ принимается значения в X. Можно с помощью f определить случайную величину $\eta = f(\xi)$, $\xi = (\xi_1, \dots, \xi_n)$, $\eta = (\eta_1, \dots, \eta_n)$. Наша задача состоит в том, чтобы выразить плотность $\rho_\eta(y)$ случайной величины η через плотность случайной величины $\rho_\xi(x)$.

Сейчас мы будем её искать.

$$\rho_\eta(y) = \lim_{O(y)\to\{y\}} \frac{\mathbb{P}\{\eta \in O(y)\}}{V\{O(y)\}} = \lim_{O(y)\to\{y\}} \frac{\mathbb{P}\{f(\xi) \in O(y)\}}{V\{O(y)\}} =$$

$$= \lim_{O(y)\to\{y\}} \frac{\mathbb{P}\big\{\xi \in f^{-1}\big[O(y)\big]\big\}}{V\{O(y)\}} =$$

сделаем деликатную вещь: прообраз $O(y)$ есть некая другая окрестность $O\big(f^{-1}(y)\big)$

$$= \lim_{O(y)\to\{y\}} \frac{\mathbb{P}\big\{\xi \in O[f^{-1}(y)]\big\}}{V\{O[f^{-1}(y)]\}\}} \frac{V\big\{O[f^{-1}(y)]\big\}}{V\big(O(y)\big)} =$$

$$= \rho_\xi\big(f^{-1}(y)\big)\big|(\mathfrak{J}f^{-1})(y)\big|$$

Можем ли мы хоть для чего-нибудь использовать эти все знания. Зачем мы всем этим занимаемся. Все эти плотности, вероясности, мат ожидания, не могут быть совсем уж произвольными вещами. Есть некий таинственный список распределений, которыми пользуются. Он меняется от эпохи к эпохе.

6.1 Ошибка округления

Есть шкала, расстояние между делениями Δ, есть стрелка, которая останавливается между делениями шкалы. Ошибка может меняться в интервале $\big[-\frac{\Delta}{2}, \Delta 2\big]$. А можем ли мы догадаться, какое распредление вероятности этой ошибки, какая плотность? Вроде бы всё равно, куда попасть, значит, постоянная плотность должна быть внутри отрезка, а вне отрезка равна нулю.

$$\rho(x) = \begin{cases} 0, & x \notin \big[-\frac{\Delta}{2}, \frac{\Delta}{2}\big]; \\ c, & x \in \big[-\frac{\Delta}{2}, \frac{\Delta}{2}\big]. \end{cases}$$

А как найти c. Надо, чтобы вероятность всей прямой была единицей, значит, $c = \frac{1}{\Delta}$. Таким образом, $\rho_\xi(x) = \frac{\Delta}{J}\big\{-\frac{\Delta}{2} \leqslant x \leqslant \frac{\Delta}{2}\big\}$.

Определим ещё одну случайную величину $\eta = a + b\xi$. Тогда образ отрезка при линейном отображении — тоже отрезок. Получим аналогичное «равномерное» распределение, но уже на некотором отрезке $[c, d]$.

Чему равно математическое ожидание ξ? $\mathbb{E}\xi = \int\limits_{-\Delta/2}^{\Delta/2} x\frac{1}{\Delta}\,dx = 0$. Умея вычислять мат ожидание, мы фактически обладаем понятием дисперсии

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \int\limits_{-\infty}^{+\infty} (x - \mathbb{E}\xi)^2 \rho_\xi(x)\,dx = \frac{\Delta}{12}.$$

Аналогично, $\mathbb{D}\eta = \frac{(d-c)^2}{12}$.

Раз знаем дисперсю, хотим перетащить на эту теорию закон больших чисел. Для этого надо ввести понятие независимых случайных величин. Это позже. Давайте ещё пример посмотрим.

6.2 Момент отказа

Пусть компьютер работает при $t = 0$ и есть момент $\tau > 0$, в который он откажет. Введём случайую величину $Q_\tau(t) = \mathbb{P}\{\tau > t\}$. Рассмотрим приращение

$$Q_\tau(t + \Delta t) = Q_\tau(t)\big(1 - \lambda(t)\Delta t + o(\Delta t)\big).$$

Составляем дифференциальное уравнение

$$\Delta Q_\tau(t) = -\lambda(t) Q_\tau(t)\Delta t + o(\Delta t); \qquad \frac{}{dt} Q_\tau(t) = -\lambda(t) Q_\tau(t).$$

Отсюда получается, что
$$Q_\tau(t) = e^{-\int\limits_0^t \lambda(s)\,ds}.$$
Пусть $\lambda(t) = \mathrm{const}$, то есть от t никак не зависит. Тогда $Q_\tau(t) = e^{-\lambda t}$. Хотим поняти, какова плотность этого распределения
$$e^{-\lambda t} = \int\limits_t^{+\infty} \rho_\tau(s)\,ds.$$
Таким образом,
$$\rho_\tau(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geqslant 0; \\ 0 & t < 0. \end{cases}$$
Такое распределение называется показательным.

Несложными вычислениями можно показать, что $\mathbb{E}\tau = \frac{1}{\lambda}$, $\mathbb{D}\tau = \frac{1}{\lambda^2}$.

6.3 Нормально распределение

Оно используется для оценки ошибок общего характера. Разберём случай, обозначаемый $\xi \sim N(0,1)$. Определим плотность
$$\rho_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$
С помощью этого распределения мы уточним неравенство Чебышёва, но чуть позже.

Общее нормальное распределение получается из стандартного $N(0,1)$ линейной заменой
$$\eta = \sigma\xi + a, \quad f(x) = \sigma x + a = y; \quad f^{-1}(y) = \frac{y-a}{\sigma} = x.$$
Тогда
$$\rho_\eta(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-a)^2}{2\sigma^2}}.$$
Вычисления показывают, что $\mathbb{E}\xi = 0$, $\mathbb{E}\eta = a$, $\mathbb{D}\xi = 1$, $\mathbb{D}\eta = \sigma^2$. Пишут $\eta \sim N(a, \sigma^2)$.

Мечта ещё со времён Лапласа, чтобы все ошибки описывались нормальным законом с некоторыми a и σ^2. Но не обязательно вообще ощибка описывается с помощью распределения.

7 Лекция 7

В чём состоят основы теории вероятности. Постараюсь в одном месте все их написать. Как и полагается в аксиоматике Колмогорова множество элементарных событий Ω, на нём есть σ-алгебра $\mathfrak{B}$, и есть вероятность (мера) $\mathbb{P}$. Кроме того есть функции $\xi = \xi(\omega) \in \mathbb{R}^1$, но не абы какие, но такие, что если $B \subset \mathbb{R}^1$ (борелевское), то $\xi^{-1} = \big\{\omega\colon \xi(\omega) \in B\big\} \in \mathfrak{B}$. При этом
$$\mathbb{P}\big\{\xi^{-1}(B)\big\} = \mathbb{P}\{\xi \in B\} = \mu_\xi(B)$$
называется вероятностной мерой или распределением ξ.

Можно сказать, что Колмогоров сказал больше, чем нужно в приложениях. Но если нам задана только вероятностная мера μ на прямой, то ничего не стоит создать и Ω, и $\mathfrak{B}$, $\mathbb{P}$ и ξ, для которых $\mu = \mu_\xi$. Возьмём $\Omega = \mathbb{R}$, $\mathfrak{B} = \{\text{борелевские}\}$, $\mathbb{P} = \mu$, $\xi(\omega) = \omega$. Таким образом, Колмогоров не запросил ничего лишнего.

Дальше, что такое математическое ожидание.
$$\mathbb{E}\xi = \int\limits_\Omega \xi(\omega)\mathbb{P}(d\omega).$$
Оказывается, что если есть какая-то функция $f(x)$, то функция $\eta = f(\xi) = f\big(\xi(\omega)\big)$ тоже будет являться случайной величиной и можно вычислить мат ожидание
$$\mathbb{E}f(\xi) = \int\limits_{\mathbb{R}}^{1} f(x)\,\mu_\xi(dx).$$

При это доказать это без Ω и ω не получится. Доказывают равенства интегральных сумм для интеграла

$$\mathbb{E}f(\xi) = \int\limits_{\Omega} f\big(\xi(\omega)\big)\mathbb{P}(d\omega).$$

Однако такие формулы невозможно использовать для расчётов. Интегральные суммы считать — дело неблагодарное. Но вводится функция плотности.

Пусть существует такая функция от x, что

$$\int\limits_{B} p_\xi(x)\,dx = \mu_\xi(B).$$

Ну и легко видеть, что наконец-то для случая, когда существует плотность

$$\mathbb{E}\big(f(\xi)\big) = \int\limits_{\mathbb{R}} f(x)\,\rho_\xi(x)\,dx.$$

А это уже то, что вполне возможно как-то посчитать.

Далее если мы знаем, что такое мат ожидание, то и знаем, что такое дисперсия.

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

Любимое обозначения для математического ожидания $a = \mathbb{E}\xi$. Тогда

$$\mathbb{D}\xi = \mathbb{E}(\xi - a)^2 = \int\limits_{\mathbb{R}} (x-a)^2 p_\xi(x)\,dx.$$

А это мы умеем вычислять.

Но если не существует плотности, тогда в самом общем случае

$$\mathbb{D}\xi = \int\limits_{\mathbb{R}} (x-a)^2 \mu_\xi(dx).$$

Глядя на эти равенства, мы видим, что неравенство Чебышёва тоже будет иметь место. В чём оно заключается
Утверждение 7.1. $\mathbb{P}\big\{|\xi - \mathbb{E}\xi| \geqslant \varepsilon\big\} \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2}$.
Доказательство. Доказывается это вот как

$$\mathbb{D}\xi \geqslant \int\limits_{|x-a|\geqslant\varepsilon} (x-a)^2\,\mu_\xi(dx) \geqslant \varepsilon^2 \int\limits_{|x-a|\geqslant\varepsilon} \mu_\xi(dx) = \varepsilon^2 \mathbb{P}\big\{|\xi - a| \geqslant \varepsilon\big\}.$$

■ Ну раз у нас есть неравенство Чебышёва, то где-то недалеко закон больших чисел. Но для этого нам нужно ввести понятие независимых случайных величин.

7.1 Закон больших чисел

Возьмём две случайных величины ξ, η. Если для любых двух борелевских множеств $A, B \subset \mathbb{R}$ события

$$\big\{\omega\colon \xi(\omega) \in A\big\} = \{\xi \in A\}, \{\eta \in B\}$$

независимы, то есть

$$\mathbb{P}\{\xi \in A, \xi \in B\} = \mathbb{P}\{\xi \in A\}\mathbb{P}\{\eta \in B\}.$$

Для конъюнкции двух событий используют знак умножения, пересечения или запятую. $\mathbb{P}\{\xi \in A, \xi \in B\} = \mathbb{P}\big\{(\xi, \eta) \in A \times B\big\}$. Здесь возникает мера на плоскости

$$\mathbb{P}\big\{(\xi,\eta) \in A \times B\big\} = \mu_{\xi\eta}(A \times B) = \mu_\xi(A)\mu_\eta(B).$$

Здесь $\mu_{\xi\eta} = \mu_\xi \times \mu_\eta$.

Для чего я это рассказываю. У нас была теорема о математическом ожидании произведения случайных величин.

Пусть есть две независимые случайной величины ξ, η, пусть $\mathbb{E}|\xi| < \infty$, $E|\eta| < \infty$. Тогда

$$\mathbb{E}(\xi\eta) = \mathbb{E}\xi\eta = \mathbb{E}\xi\mathbb{E}\eta.$$

Почему. Вот есть ξ. Когда строим интегральные суммы, мы заменяем ξ на ξ_n имеющую не более чем счётное число значений, для которой

$$|\xi - \xi_n| < \frac{1}{n}.$$

И заметьте, что ξ_n можно воспринимать, как функцию от ξ. Это нужно для того, чтобы говорить о независимости ξ_n и η_n. Если случайные величины независимы, то и функции от этой пары случайных величин независимы. И наконец, во всех точках Ω есть сходимость $\xi_n\eta_n \to \xi\eta$ и можно пользоваться теоремой Лебега о предельном переходе под знаком интеграла.

Кто не помнит теорему Лебега, можно написать

$$\xi\eta - \xi_n\eta_n = \xi_\eta - \xi_n\eta + \xi_n\eta - \xi_n\eta_n.$$

Расставляя скобки мы видим, что возникает оценка.

Отсюда получаем сразу, что для независимых ξ, η и $\mathbb{D}(\xi+\eta) = \mathbb{D}\xi + \mathbb{D}\eta$. А отсюда мы сразу получаем закон больших чисел в том же виде. И таким образом, всё что мы знали про дискретные случайные величины, мы перетащили на общих случай.

7.2 Двигаемся дальше

Пусть плотность случайной величины ξ зависит от ещё каких-то параметров $p_\xi(x; \alpha, \beta, \gamma)$.

Кстати сказать, вместо $\mathbb{R}$ можно везде поставить $\mathbb{R}^n$, никакого отличия.

Нам нужно будет много случайных величин, бесконечное число. И в терминах плотности нужно получать плотности нескольких случайных величин.

Пусть ξ, η независимы. И пусть $p_\xi(x), p_\eta(x)$ существуют. Тогда на плоскости существует «совместная» плотность $p_{\xi,\eta}(x,y) = p_\xi(x) \cdot p_\eta(y)$. Надо это доказать

Доказательство. Мы знаем, что $\mu_{\xi,\eta}(A \times B) = \mu_\xi(A) \cdot \mu_\eta(B)$. Когда существует плотность, имеем

$$\int\limits_A p_\xi(x)\,dx \int\limits_B p_\eta(y)\,dy = \iint\limits_{A\times B} p_\xi(x) p_\eta(y)\,dx\,dy.$$

Но мера счётно-аддитивная однозначно определяется своими значениями на все прямоугольниках. ∎

Не следует забывать, что кроме плотности есть ещё и функция распределения случайной величины. Возьмём функция $F_\xi(x) := \mathbb{P}\{\xi < x\}$ (иногда ставят $\xi \leqslant x$). Конечно же

$$F_\xi(x) = \mathbb{P}\bigl\{\omega \colon \xi(\omega) < x\bigr\}.$$

Если $x_1 < x_2$, то $F_\xi(x_1) \leqslant F_\xi(x_2$. Ну и если нужно посчитать вероятность попасть в полуинтервал

$$\mathbb{P}\{a \leqslant \xi < b\} = \mathbb{P}\{\xi < b\} - \mathbb{P}\{\xi < a\} = F_\xi(b) - F_\xi(a).$$

В случае существования плотности можно написать

$$\mathbb{P}\{a \leqslant \xi < b\} = \int\limits_a^b p_\xi(x)\,dx = F_\xi(b) - F_\xi(a).$$

Таким образом, F_ξ является первообразно от $p_\xi(x)$.

7.3 Распределение многомерной случайной величины

Пусть $\xi = \{\xi_1, \xi_2\}$. По определение функцией распределения двумерной случайной величины называют

$$F_{\xi_1,\xi_2}(x_1, x_2) = \mathbb{P}\{\xi_1 < x_1, \xi_2 < x_2\}.$$

Но на плоскости не одна система координат. Например, я могу оси повернуть. Как изменится функция распределения? Абсолютно непонятно.

Попытаемся снарядом попасть в некоего врага. Военные считают, что если при постоянном прицеле снаряды не будут попадать в одну и ту же точку. Есть некое рассеивание. Считают, что плотность вероятности имеет линии уровня в виде эллипсов. Двумерную функцию распределения не использовали в докомпьютерную эпоху. Сложно было считать.

Легко убедиться, что для независимых ξ, η выполняется

$$F_{\xi,\eta}(x,y) F_\xi(x) F_\eta(y).$$

Для других такого нет.

Пусть $\xi_1, \xi_2, \dots, \xi_n$ независимы. Тогда имеется совместная плотность $\rho(x_1, x_2, \dots, x_n)$. Можно взять $\eta = f(\xi_1, \xi_2, \dots, \xi_n)$. Как получить из многомерной плотности набор каких-то случайных величин. Ответ вот какой для двумерного случая.

Пусть ξ, η случайные одномерные. Тогда

$$p_\xi(x) = \int\limits_{\mathbb{R}} p_{\xi,\eta}(x, y)\, dy.$$

Понижение размерности производится путём интегрирования по всем остальным значениям.
Доказательство. Что это значит. Это значит, что вероятность

$$\mathbb{P}(\xi \in B) = \int\limits_B p_\xi(x)\, dx.$$

Мы докажем, что это так. $\mathbb{P}\{\xi \in B\} = \mathbb{P}\{\xi \in B, \eta \in \mathbb{R}\} = \int\limits_B \int\limits_{\mathbb{R}} p_{\xi\eta}(x, y)\, dy\, dx = \int\limits_B p_\xi(x)\, dx.$ ■

7.4 Суммы случайных величин

В основном рассматривают суммы независимых случайных величин. Мы хотим оценивать, насколько $\overline{x} = \frac{1}{n}\Sigma x_i$ отличается от истинного значения a. Поэтому случайные величины важны. И мы ими сейчас займёмся.

Пусть ξ_1, ξ_2 независимы и имеют плотности. Тогда

$$p_{\xi_1,\xi_2}(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2).$$

Положим $\eta_1 = \xi_1 + \xi_2$, $\eta_2 = \xi_2$ (чтобы было обратное преобразование). Тогда сделаем на плоскости замену $y_1 = x_1 + x_2$ и $y_2 = x_2$.

Надо вспомнить формулу. Если $y = f(x)$, где x, y — векторы, но отображение гладкое и взаимнооднозначное: $x = f^{-1}(y)$. Пусть ξ — случайный вектор с плотностью $p_\xi(x)$ (относительно неориентированного объёма). Положим $\eta = f^{-1}(\xi)$. Тогда

$$p_\eta(y) = p_\xi\big(f^{-1}(y)\big)\big|Jf^{-1}(y)\big|.$$

В нашем случае обратное отображение $x_1 = y_1 - y_1$, $x_2 = y_2$. Понятно, что якобиан такого отображения есть единица.

$$p_{\eta_1,\eta_2}(y_1, y_2) = p_{\xi_1}(y_1 - y_2) p_{\xi_2}(y_2).$$

Если хотим только первое, то

$$p_{\eta_1}(y_1) = p_{\xi_1+\xi_2}(y_1) = \int\limits_{-\infty}^{+\infty} p_{\xi_1}(y_1 - y_2) p_{\xi_2}(y_2)\, dy_2.$$

Таким образом плотность суммы есть свёртка плотностей.

Лобачевский посчитал, как сложить n случайных величин.

Есть такое чудо. Пусть есть независимые случайные величины $\xi_1, \dots, \xi_n$. Сумму обозначим через $S_n = \xi_1 + \dots + \xi_n$. Чем больше слагаемых вы возьмёте, то больше будет область, где где значения не нулевые. Но раз она большая, то и плотность должна быть маленькая. Но Лаплас посоветовал взять нормированную сумму s_n. Вот, что это такое

$$s_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}}.$$

Заметим, что $\mathbb{E}s_n = 0$ и $\mathbb{D}s_n = 1$. Суть состоит в том, что s_n сходится к предельному нормальному закону с плотностью, которую я уже выписывал

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Для нормированной суммы с ростом n будет более выделяться горб в нуле. Плотность уже не будет размазываться.

Самое удивительно, что можно даже увидеть такую сходимость для дискретных законов, например, если $\xi_i \in \{0, 1\}$.

Пусть есть последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ случайных величин и ещё случайная величина ξ.

Определение 7.1. *Говорят, что $\xi_n \to \xi$ в слабом смысле, если*

$$\forall\ f \in C_0^\infty \quad f(\xi_n) \to \mathbb{E}f(\xi).$$

Скоро мы увидим, что слабая сходимость случайных величин гарантирует сходимость вероятности попасть в заданный интервал. Именно в этом смысле будет доказывать сближение нормированных сумм случайных величин к нормальному закону.

8 Центральная предельная теорема

Самая интересная теорема теории вероятностей. Она говорит, что сумма случайных величин при каких-то условиях имеет нормальное распределение, то есть

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Похоже на колокол. Обозначают это распределение $N(0,1)$. У него $\mathbb{E}\xi = 0$, $\mathbb{D}\xi = 1$, $p_\xi = \varphi$

Если же $\eta = a + \sigma\xi$, то $\mathbb{E}\eta = a$, $\sqrt{\mathbb{D}\xi} = \sigma$ — стандартное отклонение. Плотность тоже можно выразить

$$p_\eta(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{(y-a)^2}{2\sigma^2}\right\}.$$

Обозначим меру ν, которой отвечает функция плотности $\varphi(x)$. То есть

$$\nu(B) = \int\limits_B \varphi(x)\,dx.$$

Пусть $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. Обозначим $s_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}}$. Тогда $\mu_{s_n}(B) = \mathbb{P}\{s_n \in B\}$. Будем доказывать, что μ_{s_n} сходится (в каком-то смысле) к ν.

Что такое слабая сходимость

Определение 8.1. $\frac{\xi_1+\xi_2+\dots}{\xi}{}_n n \xrightarrow[n\to\infty]{} 0$ *по вероятности, если*

$$\forall\ \varepsilon > 0 \quad \mathbb{P}\left\{\frac{\xi_1 + \dots + \xi_n}{n} > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0.$$

Пусть $\varphi(x) = \varphi$ — гладкая (C^2) и финитная функция. Обощначим $\mu(\varphi) = \int\limits_{-\infty}^{\infty} \varphi(x)\mu(dx) = \mathbb{E}\varphi(\xi)$.

Определение 8.2. *Последовательность мер μ_n называется сходящейся слабо к мере μ. Если*

$$\forall\ \varphi \quad \mu_n(\varphi) \xrightarrow[n\to\infty]{} \mu(\varphi).$$

Мы будем доказывать как раз такую сходимость.

Пусть $J = [a,b]$, $\mu_n \xrightarrow[n\to\infty]{} \nu$ слабо, ν — нормальное, то $\mu_n(J) \xrightarrow[n\to\infty]{} \nu(J)$. Это можно увидеть, взяв индикатор этого интервала $I_{[a,b]}$, этот индикатор приблизив гладкими функциями сверху и снизу: $\varphi_+(x) \geqslant J_{[a,b]}(x) \geqslant \varphi_-(x)$. (Приближения легче строить так, чтобы функции отличались от индикаторов на маленьком отрезке и с ростом n этот отрезок уменьшился).

Мы выясняли, что если у случайных величин есть плотности, то плотность суммы есть свёртка плотностей. Со свёрток удобно работать через преобразование Фурье. Преобразование Фурье от функции плотности называется характиристической функцией случайной величины.

Определение 8.3. *Пусть есть случайная величина ξ, её распределение μ_ξ, её плотность $p_\xi(x)$. Тогда обозначают*

$$f_\xi(t) = \mathbb{E}e^{it\xi} = \int\limits_{-\infty}^{\infty} e^{itx}\mu_\xi(dx) = \int\limits_{-\infty}^{\infty} e^{itx}p_\xi(x)\,dx$$

Это и называют характеристических функций.

Утверждение 8.1. *Пусть ξ и η независимы. Тогда $f_{\xi+\eta}(t) = \mathbb{E}e^{it(\xi+\eta)} = \mathbb{E}(e^{it\xi}e^{it\eta}) = f_\xi(t)f_\eta(t)$.*

Пусть есть гладкая финитная функция $\varphi(t)$. Тогда преобразование Фурье будем обозначать волнами.

$$\widetilde{\varphi}(t) = \int\limits_{-\infty}^{\infty} e^{itx}\varphi(x)\,dx.$$

Этот интеграл совсем не страшный, так как φ финитная функция. С другой стороны

$$\varphi = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-itx}\widetilde{\varphi}(t)\,dt.$$

А этот интеграл ужасный? Можем сказать, что $\left|\widetilde{\varphi}(t)\right| \leqslant \frac{\text{const}}{1+t^2}$. (Убывает на бесконечности быстрее второй степени.) Ничего ужасного здесь тоже нет.

Можем теперь записать

$$\mu(\varphi) = \int\limits_{-\infty}^{\infty} \varphi(x)\mu(dx) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\Bigg(\int\limits_{-\infty}^{\infty} e^{-itx}\widetilde{\varphi}(t)\,dt\Bigg)\mu(dx).$$

Есть асболютная сходимость интегралов, так как для $\widetilde{\varphi}$ у нас есть оценка, а мера μ конечная (мера всего единица). Значит, можно менять порядок интегрирования.

$$\mu(\varphi) = \frac{1}{\pi}\int\limits_{-\infty}^{\infty}\Big(e^{-itx}\mu(dx)\Big)\widetilde{\varphi}(t)\,dt = \int\limits_{-\infty}^{\infty}\overline{f_\mu(t)}\widetilde{\varphi}(t)\,dt.$$

Если вы помните теорию обобщённых функций, знаете, что

$$(F,\varphi) = \int\limits_{-\infty}^{\infty}\overline{F(x)}\varphi(x) = F(\varphi).$$

Как со скалярным (эрмитовым) произведением, над одним из множителей ставят сопряжение.

В нашем случае

$$\mu(\varphi) = (\mu,\varphi) = \frac{1}{2\pi}(f_\mu,\widetilde{\varphi}).$$

Вот нам дано, что $f_n(t) \xrightarrow[n\to\infty]{} f(t)$. Нужно показать слабую сходимость мер. Ну а что тут доказывать теперь

$$\varphi(\varphi) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\overline{f_\mu(t)}\varphi(t)\,dt.$$

Предельный переход под знаком интеграла сделаем по теореме Лебега. (Мы же имеем фиксированную φ).

Таким образом мы доказали, что

Утверждение 8.2. *Пусть $f_n \xrightarrow[n\to\infty]{} f(t)$. Тогда $\mu_n \xrightarrow[n\to\infty]{} \mu$ слабо.*

Можно наконец объявить теорему.

Теорема 8.1. *Пусть $\xi_1,\ldots,\xi_n,\ldots$ независимы и одинаково распределённые случайные величины, у которых существуют математические ожидания $\mathbb{E}\xi_k = a$ и дисперсии $\mathbb{E}\xi_k = \sigma^2$. Тогда обозначим*

$$S_n = \xi_1+\cdots+\xi_n,\quad \mathbb{E}S_n = an,\quad \mathbb{D}S_n = n\sigma^2.$$

Введё нормированную сумму

$$s_n = \frac{S_n - na}{\sigma\sqrt{n}}.$$

Утверждается, что $\mu_{s_n} \xrightarrow[n\to\infty]{} \nu$ слабо.

Доказательство. Достаточно показать сходимость характеристических функций. Напишем суммы

$$S_n - na = \sum_{k=1}^{n}\xi_k - na = \sum_{k=1}^{n}(\xi_k - a).$$

Теперь нормированные суммы

$$\frac{S_n-na}{\sigma} = \sum_{k=1}^{n}\underbrace{\frac{\xi_k-a}{\sigma}}_{\eta_k} = \sum_{k=1}^{n}\eta_k.$$

Видно, что $\mathbb{E}\eta_k = 0$, $\mathbb{D}\eta_k = 1$. Рассмотрим характеристическую функцию

$$f_k(t) = \mathbb{E}e^{it\eta_k} = \int\limits_{-\infty}^{\infty} e^{itk}\mu_k(dx).$$

Продифференцируем это выражение

$$f_k'(t) = \int\limits_{-\infty}^{\infty} ixe^{itx}\mu_k(dx).$$

Если получившийся интеграл сходится равномерно, то слава богу. Ну так и будет. Остюда мы получим

$$f_k'(0) = i\mathbb{E}\eta_k = 0, \quad f_k''(0) = -1.$$

Обозначим характеристическую функцию случайной величины η_k через $g_k(t)$. Тогда

$$f_{s_n}(t) = \left[g\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Давайте по формуле Тейлора это вычислим.

$$f_{s_n}(t) = \left[1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{1}{n}\right)\right]^n \xrightarrow[n\to\infty]{} e^{-t^2/2}.$$

И вижу с удовольствием, что теорема доказана. ■

Это конечно хорошо. Но у нас количество испытаний всегда конечно. Давайте посмотрим, что будет для испытаний Бернулли.

Пусть имеется n испытаний, p вероятность удачи, $q = 1-p$ вероятность неудачи. Исследуется число успехов $\mu = \varepsilon_1 + \dots + \varepsilon_n$. Посчитаем

$$\mathbb{E}\varepsilon_i = p, \quad D\varepsilon_i = \mathbb{E}\varepsilon_i^2 - (\mathbb{E}\varepsilon)^2 = p - p^2 = pq.$$

Отсюда

$$\mathbb{E}\mu = np, \quad \mathbb{D}\mu = npq.$$

И неплохо было бы это даже запомнить. Положим $\mu^* = \frac{\mu - np}{\sqrt{npq}}$. Мы хотим, чтобы

$$\mathbb{P}\{a \leqslant \mu^* \leqslant b\} = \nu\big\{[a,b]\big\} = \Phi(b) - \Phi(a).$$

Когда говорят о нормальном распредлении, состовляются таблицы Лапласа, в которой записываются значения

$$\Phi(x) = \int\limits_{-\infty}^{x} \varphi(y)\,dy$$

функции распределения $N(0,1)$.

Например, пользуется спросом значение следующей вероятности

$$\mathbb{P}\big\{|\mu^*| \leqslant 1.96\big\} = 0.95.$$

Ну и раскручивая назад

$$\mathbb{P}\big\{|\mu - np| \leqslant 2\sqrt{npq}\big\} \approx 0.95.$$

А из неравенства Чебышёва следовало, что

$$\mathbb{P}\big\{|\mu - np| \leqslant 3\sqrt{npq}\big\} \geqslant \frac{8}{9}.$$

То есть мы уточнили неравенство Чебышёва.

Если монеты бросались $n = 100$ раз, а $\mu = 70$. Предположим, что $p = \frac{1}{2}$. Тогда $npq = 25$, $\sqrt{npq} = 5$ и $\mathbb{E}\mu = 50$. Скажем «мошенничать изволите».

Пирсон бросал монету $n = 24\,000$ раз. При этом получилось $\mu = 12\,012$. В этом случае $\sqrt{npq} = \sqrt{6000} \approx 78$. Отсюда

$$\mu^* = \frac{12}{78} = 0.155 = \xi \sim N(0,1).$$

Получилось так у Пирсона. Чему равна вероятность того, что

$$\mathbb{P}\big\{|\xi| \leqslant 0.155\big\} = \int\limits_{-0.155}^{0.155} \frac{1}{\sqrt{2\pi}} \underbrace{e^{-x^2/2}}_{\approx 1}\, dx \approx \frac{1}{8}.$$

Мог ли Пирсон получить такой результат? Дело в том, что Пирсон сначала 6 000 раз, оказалась слишком большая разница. Решил ещё побросать. Бросать монету легко, а сбиться со счёта ещё легче.

8.1 Наблюдение физических величин

Физик наблюдает величину a, но получается $a + \delta_i$, δ_i — ошибка. Тогда

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = a + \frac{1}{n}\sum_{i=1}^{n}\delta_i.$$

Если мы наделим ошибки δ_i душой, которая заключается в том, что они станут независимыми случайными величинами, $\mathbb{E}\delta_i = 0$, $\mathbb{D}\delta_i = \mathbb{E}\delta_i^2 = \sigma^2$. Осюда

$$\mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i\right) = \frac{\sigma^2}{n}.$$

Тогда $\sqrt{\mathbb{D}(1/n\Sigma\delta_i)} = \frac{\sigma}{\sqrt{n}}$. Всё решает один единственный параметр σ. Возникает вопрос, при каком n следующая вероятность

$$\mathbb{P}\big\{|\overline{x} - a| \leqslant 2s/\sqrt{n}\big\} = 0.95.$$

9 После цпт

Эта теорема состояла в следующем. Если есть независимые одинаково распределённые случайные величины $\xi_1, \dots, \xi_n, \dots$ с характеристиками $\mathbb{E}\xi_k = a$, $\mathbb{D}\xi_k = \sigma^2$, можно ввеси $S_n = \xi_1 + \dots + \xi_n$ и $s_n = \frac{S_n - na}{\sigma\sqrt{n}}$. Так вот s_n слабо сходится к нормальному распределению, то есть к распределению ν, имеющему плотность

$$p_\nu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

При этом $\mathbb{P}\big\{|\nu| \leqslant 1\big\} = 0.84$, $\mathbb{P}\big\{|\nu| \leqslant 1.96\big\} = 0.95$, $\mathbb{P}\big\{|\nu| \leqslant 2.57\big\} = 0.99$, $\mathbb{P}\big\{|\nu| > 3\big\} = 0.003$.

Ну а тогда ведь $S_n = na + \sigma\sqrt{n}s_n$. Конечно, $\mathbb{E}S_n = na$ и $|na| \gg \sigma\sqrt{n}$ при $a \neq 0$.

Значит, если вам предлагается сыграть в азартную игру. Делаете ставки $\xi_1, \dots, \xi_n, \dots$, вам первым делом стоит поинтересоваться знаком a. Если $a > 0$, стоит играть, но так никогда не бывает. А если $a < 0$, то надо уходить. Выиграть можно только играя в не совсем случайную игру.

Что мы имеем, когда хотим написать так называемый доверительный интервал. Имеем данные $x_1, \dots, x_n$. В высшей степени тёмное дело объяснить, почему мы считаем, что эти иксы есть реализация независимых одинаково распределённых случайных величин. Но за неимением лучшего мы именно так и считаем. Тогда $x_n = a + \delta_i$. В высших сферах есть генератор случайных чисел и один ансабль реализаций генератора мы имеем в выборке. n называется объёмом выборки (оно же число наблюдений). Выводы не обладают абсолютной надёжностью. Но вот принято путать обозначения в учебниках. С одной стороны $x_1, x_2, \dots, x_n$ — результаты измерений, а с другой стороны обозначают теми же буквами случайные величины, которые якобы лежат за тем, что мы имеем. Пишут

$$\mathbb{D}x_i = \mathbb{D}\delta_i = \sigma^2.$$

Поскольку все одинаковые, можно обозначить за σ^2.

Ну вот не хватает букв в математике, обозначают разные вещи одной буквой. Иногда всё-таки случайные величины обозначают $X_1, \dots, X_n$.

Как много раз я объяснял, оценкой дисперсии является $\sigma^2 \approx s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$.

Давайте поймём, откуда взялась минус единица. Пусть $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$. Тут сейчас снова возникнет путаница в обозначениях реализации и случайной величины. Будем считать мат ожидание, уже подставив вместо наблюдений случайные величины. Можно ещё соорудить случайную величину

$$\eta \sim \begin{pmatrix} x_1 & x_2 & \dots x_n & \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}.$$

Тогда $\mathbb{E}S^2 = \mathbb{D}\eta$. Ну давайте считать, подставив x_i как случайные величины. Давайте для простоту считать, что $a = 0$, всё равно они сокращаются.

$$\mathbb{E}S^2 = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]^2\right) = \frac{1}{n}(n\sigma^2) - \frac{1}{n^2}\mathbb{E}\sum_{i,j=1}^{n}x_ix_j =$$
$$= \sigma^2 - \frac{1}{n^2}\mathbb{E}\sum_{i=1}^{n}x_i^2 = \sigma^2 - \frac{\sigma^2}{n} = \left(1 - \frac{1}{n}\right)\sigma^2.$$

А нам бы хотелось иметь мат ожидание оценки σ^2. Поэтому считаем $\mathbb{E}s^2 = \sigma^2$. Это пример того, как изучаются характеристики наблюдений, возводя наблюдения в ранг случайных величин. Если наблюдений много, хотя бы несколько десятков, разница S^2 и s^2 ничтожна:

$$S^2 = \underbrace{\frac{1}{n}\sum x_i^2}_{\to \mathbb{E}x_i^2} - \underbrace{(\overline{x})^2}\to (\mathbb{E}x_i)^2.$$

Сходимостм слабые.

Обозначают $\overline{x} = \hat{a}$. Имеем $\mathbb{D}\overline{x} = \frac{\sigma^2}{n}$. Соорудим центрированную и нормаированную величину. Она тогда близка к стандартному нормальному, а значит

$$\mathbb{P}\left\{\left|\frac{\overline{x} - a}{\frac{\sigma}{\sqrt{n}}}\right| \leqslant 1.96\right\} = 0.95.$$

При этом мы подразумеваем, что оценили дисперсию $\sigma \approx s$. Тогда

$$\mathbb{P}\left\{|\overline{x} - a|\frac{\sqrt{n}}{s} \leqslant 1.96\right\} \approx 0.95.$$

Чуть-чуть огрубим

$$|\overline{x} - a| \leqslant 2\frac{s}{\sqrt{n}}.$$

Может быть наша выборка оказалась среди тех пяти процентов, которые плохие. 0.95 называется доверительной вероятностью. Доверительный интервал определяется последним интервалом.

Из неравенства Чебышёво получалась оценка более грубая.

9.1 Эмпирическая функция распределения

Почему мы вообще наделяем наблюдения сущностью случайных величин. Есть один приём.

Если у вас есть много чисел, хотя бы штук 50. Вас будет тошнить, если вы на них посмотрите. Но существует приём, с помощью которого вы можете эту тошноту преодолеть.

Пусть $x_1, \ldots, x_n$ — опытные данные. Для случайной величины

$$\begin{pmatrix} x_1 & x_2 & \ldots x_n & \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}.$$

строится функция распределения $\mathcal{F}_n(x) = \frac{\text{число } x_i < x}{m} = \text{доля}(x_i < x)$. Для $-\infty < x < \infty$. Возникает понятие вариационного ряда

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}.$$

Левее точки $x_{(1)}$ функция ноль. Когда достиг, тоже ноль. А как только сдвинулись чуть правее $x_{(1)}$ получается ступенька. На последнем наблюдении $x_{(n)}$ выйдем на единичку.

Если вы такой график нарисуете, то чувство тошноты у вас уже пройдёт. В этом графике вы должны наблюдать что-то гладкое и хорошее. Теоретическое приближение. Хочется наблюдать, конечно, нормально распределени.

Мы обозначали $\nu \sim N(0, 1)$. А общий вид $\mu = a + \sigma\nu$ для $\sigma > 0$. Вводится функция Лапласа, которая не выражается в элементарных функциях

$$\Phi(x) = \mathbb{P}\{\mu < x\} = \mathbb{P}\left\{\nu < \frac{x - a}{\sigma}\right\} = \Phi\left(\frac{x - a}{\sigma}\right).$$

Соответственно $\Phi^{-1}\big(\mathbb{P}\{\mu < x\}\big) = \frac{x-a}{\sigma}$.

В Excel проще работать столбиками. Стоблец с x_i сортируете, получаете столбец с $x_{(i)}$. Дальше

$$\begin{matrix} x_1 & x_{(1)} & \frac{1}{2n} & \Phi^{-1}\left(\frac{1}{2n}\right) \\ x_2 & x_{(2)} & \frac{1}{2n}+\frac{1}{n} & \Phi^{-1}\left(\frac{1}{2n}+\frac{1}{n}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_n & x_{(n)} & 1-\frac{1}{2n} & \Phi^{-1}\left(1-\frac{1}{2n}\right) \end{matrix}$$

Как понять, похожа эмпирическая функция на прямую или не похожа. Колмогоров предложил $\sup \big|F_n(x) - F(x)\big|$. Здесь $F(x) = \mathbb{P}\{\xi < x\}$ — теоретическая функция распределения. Теперь делаем монотонное преобразование $\eta = g(\xi)$, то есть g монотонно возрастает. Так же преобразуем $y_i = g(x_i)$, То есть будем наблюдать вместо x значение $g(x)$. Расстояния по вертикали между значениями функций распределения не произойдёт. Если у меня есть две монотонных непрерывных, то я могу подобрать монотонную, которая переводит одну в другую. Будем брать функцию распределение статистики Колмогорова

$$\mathbb{P}\Big\{\sqrt{n}\big|F_n(x) - F(x)\big| < y\Big\} \xrightarrow[n\to\infty]{} K(y).$$

Функция $K(y)$ называется функцией Колмогорова. Сходмость здесь достаточно быстрая, с нескольких десятков наблюдений уже можно пользоваться.

уровни	0.15	0.10	0.05	0.02	0.01
значения	1.14	1.23	1.36	1.52	1.63

То есть $1 - K(1.14) = 0.15$, $1 - K(1.36) = 0.05$.

В результате из функции $F_n(x)$ вычитаем $\Phi\left(\frac{x-\overline{x}}{s}\right)$. Значения этой статистики как-нибудь обозначим

$$\sup_x \left| F_n(x) - \Phi\left(\frac{x-\overline{x}}{s}\right)\right| = \hat{D}_n.$$

Если взять для оценки нормальности распределения вот такую статистику (умножим её на зависящий от n коэффициент)

$\hat{D}_n(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$

уровни	0.15	0.10	0.05	0.02	0.01
значения	0.775	0.819	0.895	0.955	1.035

Если параметры распределения a, σ заранее известны, то нормальность оцениваем Функцией Колмогорова. Если оцениваем по выборке, используем вторую функцию. Это называется критерием Лиллифорса.

На века приёмы построения доверительных интервалов не годятся. А вот когда вы измеряете одно наблюдение в понедельник, вмороне во вторник, есть шансы на адекватность.

Пусть есть две серии наблюдений $x_1, \ldots x_m$, $y_1, \ldots, y_n$. Вы хотите, чтобы результаты согласовывались. Мы вычисляем $\overline{x}$, s_x, $\overline{y}$, s_y. Рассмотрим разницу $\overline{x} - \overline{y}$. Хотим, чтобы $\mathbb{E}(\overline{x} - \overline{y}) = 0$, $\mathbb{D}(\overline{x} - \overline{y}) = \mathbb{D}\overline{x} + \mathbb{D}\overline{y} = \frac{s_x^2}{m} + \frac{s_y^2}{n}$. Мы считаем, что

$$\frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \sim N(0,1).$$

Можно с помощью эмпирических функций взять супремум, когда m и n велики (хотя бы по несколько десятков)

$$\sup \big|F_m(x) - F_n(x)\big| \sqrt{\frac{mn}{m+n}}$$

Для больших n, m эта штука имеет распределение Колмогорова. Мы здесь работаем в предположении, что наши статистические функции равны.

10 Лекция 10

Пусть $\xi \sim N(0,1)$. То есть $\mathbb{E}\xi = 0$, $\sqrt{\mathbb{D}\xi} = 1$. Пусть также $\eta = \sigma\xi + a$, где $\sigma \neq 0$. Тогда $\mathbb{E}\eta = a$, $\mathbb{D}\eta = \sigma^2$. Пусть мы хотим вычислить

$$\underbrace{\int_0^1 \ldots \int_0^1}_{n} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = J.$$

Можем применить вероятностный подход. Зададим последовательность случайных векторов $\xi^{(i)} = \left(\xi_1^{(i)}, \xi_2^{(i)}, \ldots, \xi_n^{(i)}\right)$.

Тогда

$$\mathbb{E}f(\xi^{(1)}) = \int\limits_0^1 \ldots \int\limits_0^1 f(x_1, x_2, \ldots, x_n)\cancelto{1}{p_{\xi^{(1)}}(x_1, \ldots, x_n)}\, dx_1 \ldots dx_n = J.$$

При этом $s^2 = \frac{1}{n-1}\sum(x_i - \overline{x})^2$. И величина $\frac{(\overline{f(\xi)} - J)}{\sqrt{n}} \sim N(0,1)$. Это будет чем-то похожим на нормальное распределение.

Обозначим номер последнего наблюдения за N. Какое надо взять число наблюдений, чтобы получить хорошую точность?

Пусть есть два набора наблюдений $x_1, \ldots, x_m$, $y_1, \ldots, y_n$. Насколько выводы по этим наблюдениям согласованы? Как известно, надо составить эмпирическую функцию распределения.

$$\sup_x \left|F_m(x) - F_n(x)\right| \sqrt{\frac{mn}{m+n}}$$

будет статистикой, называемой функцией Колмогорова. Перебрать все выборки невозможно. Даже пять или десять выборок сложно обработать. С появлением Компьютера любое распределение можно приблежённо расчитать с помощью метода Монте-Карло.

Займёмся примерами из медицины. Дело сведётся к испытаниям Бернулли. У нас будет n испытаний, p вероятность успеха в одном испытании, μ — общее число успехов. Тогда $\mathbb{E}\mu = np$, $D\mu = npq$, где $q = 1 - p$. Займёмся клиническими испытаниями по следующей схеме. Больные будут появляеться не по одиночке, а парами. Есть два метода лечения A и B. Методом бросания монетки решается, какой из больных получит лечение методом A. И хотим выяснить, какой из методов лучше. Если предполагаем, что методы эквивалентны, то $p = \frac{1}{2}$ — вероятность того, что на одной паре B окажется лучше A. Пусть μ — число побед B в n испытаниях. Пусть $n = 100$, $E\mu = 50$, $\sqrt{\mathbb{D}\mu} = 5$. В реальности μ может отклониться от мат ожидания на один корень из дисперсии легко, на два корня с трудом, на три едва ли. Таким бразом $\mathbb{E}\mu + 2\sqrt{\mathbb{D}\mu} = 60$. $\mathbb{P}\{\mu \geqslant 60\} \approx 0.02$. Это уровень значимости.

Пусть теперь $p = 0.6$ и $\mathbb{E}\mu = 60$. Тогда вероятность $\frac{1}{2}$ того, что $\mu \geqslant 60$. Это расчёт мощности.

10.1 Метод наименьших квадратов

Наши наблюдения имеют вид $x_i = a_i + \delta_i$, причём $\delta_i \sim N(0, \sigma)$. При этом σ иногда известна, а иногда оценивается. δ_i — НОРСВ соответственно. Вектор $a = (a_1, \ldots, a_n) \in L$ — некоторое линейное подпространство $\mathbb{R}^n$. Можно даже предположить, что $a \in M = L + b$, где L и b известны. Но вычитая, b из всех a_i, сведём этот случай к $M = L$. При этом $\mathbb{D}\delta_i = \frac{\sigma^2}{w_i}$, где w_i — вес, то есть число измерений, в результате которых получилось число x_i. Рассмотрим $y_i = x_i\sqrt{w_i}$, $b_i = a_i\sqrt{w_i}$. Таким образом задача сводится к равным весам.

Пример. Измеряем углы a_1, a_2, a_3 треугольника. Тогда $a_1 + a_2 + a_3 = \pi$. При этом наши измерения $x_i = a_i + \delta_i$ уже этим свойством не обладают. Рассмотрим $y_i = a_i - a_i^0 + \delta_i$, где a_i^0 — какая-то грубая оценка углов треугольника, но дающая $a_1^0 + a_2^0 + a_3^0 = \pi$. Обозначим $b_i = a_i - a_i^0$. Тогда $\Sigma b_i = 0$. Свели к линейному подространству.

Или пусть мы измеряем грузы a_1, a_2. При этом наблюдения $x_i = a_i + \delta_i$ для $i = 1, 2$. Но ещё делаем наблюдение $x_3 = a_1 + a_2 + \delta_3$. Тут тоже $x_3 \neq x_1 + _2$.

В общем. Есть величины $a_1, \ldots, a_n$. Если какие-то уравнения $F_1(a_1, \ldots, a_n) = 0, \mathbb{F}_2(a_1, \ldots, a_n) = 0, \ldots$. Вводим $a_1 = a_1^0 + \Delta a_1, \ldots a_n = a_n^0 + \Delta a_n$. Неизвестными являются только поправки Δa_n. В точке $(a_1^0, \ldots, a_n^0)$ пишем линейные приближения наших уравнений F.

Для $n = 3$ и $\dim L = 2$ получается такая картинка. Где-то в плоскости лежит a. Но дьявол вносит ошибки в мои наблюдения и x в L не лежит. Поэтому можно спроецировать x на L. $\hat{a} = \arg\min\limits_{a \in L} \|x - a\|^2$.

То, что мы предположитли о наблюдениях, как реализациях случайных величин, запишем в следующем виде

$$p_i(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - a_i)^2}{2\sigma^2}\right\}, \quad p_1(x_1)\ldots p_n(x_n) = \left(\frac{2}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i - a_i)^2\right\} \to \max_{a \in L}$$

Нахождение такого максимума (эта функция называется функцией правдоподобия) равносильно минимизации $\sum(x_i - a_i)^2 \to \min\limits_{a \in L}$.

Вопрос остаётся, насколько велико отличие $\hat{a} - a$. Запишем, что $x = a + \delta$. Тогда $\hat{a} = a + \mathrm{proj}_L\delta$. Таким образом $\hat{a} - a = \mathrm{proj}_L\delta$. Плотность δ_i имеет вид

$$p_{\delta_i}(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_i^2}{2\sigma^2}}.$$

Значит, совместная плотность есть произведение этих плотностей

$$p_\delta(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2\right\}$$

При ортогональном преобразовании сумма квадратов компонент вектора не меняется. Пусть u — ортогональное преобразование. Выберем его так, чтоб $uL = \langle e_1, \ldots, e_k\rangle \ni \delta$. Тогда $\mathrm{proj}_L \delta = (\delta_1, \delta_2, \ldots, \delta_k, 0, \ldots, 0)$. Если бы мы знали δ_i, то оценили бы $\sigma^2 = \mathbb{E}\sigma_i^2$. Нас интересует ортогональная составляющая

$$\mathrm{proj}_{L^\perp} x = \mathrm{proj}_{L^\perp}\delta = (0, 0, \ldots, 0, \delta_{k+1}, \ldots, \delta_n).$$

При этом, учитывая $\delta_i = \sigma\xi_i$, где $\xi_i \sim N(0,1)$, имеем $\|\mathrm{proj}_L x\|^2 = \sum\limits_{j=k+1}^{n} \delta_j^2 = \sigma^2 \sum\limits_{j=k+1}^{n} \xi_j^2$.

Нам надо оценить величину σ. Пусть x — наблюдения сейчас. Проекция $\mathrm{proj}_L x = \hat{a}$, мы её умеем вычислять, особенно если у нас компьютер. А ортогональная составляющая $\mathrm{proj}_{L^\perp} x = x - \hat{a} = \Delta$ — вектор кажущихся ошибок. Размерность обозначим $\dim L = k$. Тогда $\dim L^\perp = n - k$.

$$\frac{1}{n-k}\|\delta\|^2 = \sigma^2 \frac{1}{n-k}\sum_{j=k+1}^{n} \xi_j^2.$$

Число $n - k = f$ называется числом степеней свободы. Этот термин из теории распределения газа. Чем больше степеней свободы, тем больше

$$\frac{1}{n-k}\sum_{j=k+1}^{n} \xi_j^2$$

похож на единицу.

В девятнадцатом веке всё считали на бумаге. И вводили следующее распределение. Пусть $\xi_i \sim N(0,1)$. Тогда $\sum\limits_{i=1}^{m} \xi_i^2 = \chi_m^2$ — распределение хи-квадрат (К. Пирсон). Для функции распределения этой величины составляли таблицы. Через эти таблицы можно оценить σ^2. Как раз обозначают оценку, как обычно

$$s^2 = \frac{1}{n-k}\|\Delta\|^2.$$

Пусть есть выборка $x_1, \ldots, x_n$. Пусть можно записать $x_i = a + \delta_i$, то есть $a_i = a$ все одинаковые. Давайте спроектируем. У нас есть $L = \langle(1, \ldots, 1)\rangle$ — одномерное подпространство. Спроецируем

$$\frac{\big((x_1, \ldots, x_n, (1, \ldots, 1)\big)}{n}(1, \ldots, 1) = (\overline{x}, \ldots, \overline{x}).$$

Считаем $\Delta = (x_1 - \overline{x}, \ldots, x_n - \overline{x})$, $s^2 = \frac{1}{n-1}\sum\limits_i (x_i - \overline{x})^2$. Вот мы новым методом получили ту же оценку, к которой уже привыкли.

11 Лекция 12

Из какой-то замечательной науки в $\mathbb{R}^n$ разбросаны точки. Мы хотим выяснить с помощью статистики, какие факторы должны быть этими точками выражены. Пусть t_i — теория, x_i — наблюдения.

12 Регрессионный анализ

Математику я всю рассказал. А какие ожидания возлагаются на эту науку всеми, кроме математиков.

Хотим найти линейную зависимость $y = \sum\limits_{i=1}^{n} c_i x_i + \delta$. Делаем кучу измерений

$$y_j = \sum_{i=1}^{n} c_i x_{ji} + \delta_j, \quad j = 1, 2, \ldots, N \gg n.$$

Будем считать, что j номер строчки, все y_j вместе составляют столбец. Ничего не остаётся, кроме как написать

$$Y = \sum_{i=1}^{n} c_i X_i + \delta$$

Вся математика сводится в ортогональном проектировании вектора Y на линейную оболочку системы $\{X_i\}$. На всякий случай я ещё раз выпишу нашу математику.

Столбец Y называют столбцом значений объясняемой переменной. Столбцы $X_1, X_2, \ldots, X_n$ — столбцы значений объясняющих переменных. Из это столбцов запишем матрицу X.

$$Y = XC + \delta.$$

Как же нам найти оценку $\hat{C}$ для вектора C. Разность Y и его проекции должна быть ортогональна всем векторам линейной оболочки, я это запишу так

$$Y - X\hat{C} \perp X.$$

Эта запись означает, что левый вектор ортогонален всем столбцам матрицы X. Как это ещё записать

$$X^T(Y - X\hat{C}) = 0.$$

Здесь написано, что X^T имеет своими строчками ортогональные вектору $Y - X\hat{C}$ строчки. Отсюда автоматически следует, что

$$\hat{C} = (X^TX)^{-1}X^TY.$$

Если подставить вместо Y его представление $Y = XC + \delta$, получим

$$\hat{C} = C + (X^TX)^{-1}\delta.$$

Мы не можем получить само C, мы можем получить $\hat{C}$, то есть C, но с ошибкой

$$(X^TX)^{-1}\delta.$$

Бывает такое, что $y_1, y_2, \ldots, y_N$ все объясняются только одной константой. Например, закон всемирного тяготения. Если бы не было ошибок, все y_j должны были бы получиться одинаково. В этом случае пишем нашу священную величину

$$\overline{y} = \frac{1}{N}\sum_{j=1}^{N} y_j.$$

В 90-х годах несколько групп намерили значения констант Ньютона. Значения получились близкие, но доверительные интервалы не пересекаются. Можно как-то посчитать среднее, но мы не знаем, с какими весами.

Описанная схема была совершенно нереализуема, пока не было компьютеров. А теперь пожалуйста.

Пусть y — прочность стали, объясняемая переменная. x_i — химический состав. Строим регрессию y на x_i.

Или же y — медицинский показатель, x_i — дозировка лекарств. Опять регрессионный анализ.

12.1 Шанс

Рассмотрим случай, когда $x_1, \ldots, x_n$ — значения показателей. А наблюдения z принимают два значения $\{1, 0\}$ удача или неудача. Хотим $p = p(x_1, \ldots, x_n)$ — вероятность удачи. Линейную функцию искать неудобно. Вводят так называемый логшанс

$$L = \text{Logchance} = \ln\frac{1}{1-p}.$$

Тогда значения уже меняются от $-\infty$ до $+\infty$.

Тогда $L = c_0 + \sum c_i x_i$. Можно интерпретировать $x_1 = 1$. Это называется логистической регрессией.

Поделим эксперименты на две части. Одна будет обучением, а другая экзаменом. Случайное деление такого сорта никакого смысла не имеет. Экзамен будет сдаваться. Вы сделайте вот что. Возьмите два на проверку, а остальные на выборку. Частая ошибка делать такое деление случайно, в этом нет смысла.

13 Многомерное нормальное распределение

Одномерная случайная ξ_1 распределена нормально $N(0,1)$, если её плотность

$$p_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Пусть теперь $\xi = (\xi_1, \ldots, \xi_n)$. Пусть у этой системы матрица ковариации $C_\xi = E$ единичная. Тогда

$$p_\xi(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)\left\{-\frac{x_1^2 + \cdots + x_n^2}{2}\right\} = p_\xi(x) = \left(\frac{2}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{(x,x)}{2}\right\}.$$

И наконец общий случай сводится к замене $\eta = a\xi + b$. Вектор b есть математическое ожидание вектора η. А что происходит с матрицей ковариации.

Пусть вообще ξ — столбец, причём $\mathbb{E}\xi = 0$, то есть $\mathbb{E}\xi_1 = \dots = \mathbb{E}\xi_n = 0$. Посмотрим

$$\mathbb{E}\xi_i\xi_j = \operatorname{cov}(\xi_i, \xi_j).$$

И составляем из ковариаций матрицу. Как проще это сделать. Составить матрицу всевозможных компонент $\xi\xi^T$. Тогда $C_\xi = \mathbb{E}\xi\xi^T$. Пусть теперь $\eta = A\xi$, где $\det A \neq 0$. Тогда, так как A предполагается постоянной, имеем

$$C_\eta = \mathbb{E}(\eta\eta^T) = \mathbb{E}(A\xi\xi^T A^T) = A\mathbb{E}(\xi\xi^T)A^T = AC_\xi A^T.$$

Вдохновясь этим законом, вычислим, какова плотность вектора η. Я напомню такую формулу для $\eta = f(\xi)$, где $f\colon \mathbb{R}^n \to \mathbb{R}^n$ взаимнооднозначно. Она выглядит следующим образом

$$p_\eta(y) = p_\xi\big(f^{-1}(y)\big)|\text{якобиан } f^{-1} \text{ в точке } y|.$$

Для отображения $\eta = A\xi + b$ имеем соответствующее преобразование в $\mathbb{R}^n$ вида $y = Ax + b$ или $x = A^{-1}(y - b)$. Таким образом,

$$p_\eta(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\big(A^{-1}(y-b), A^{-1}(y-b)\big)}{2}\right\}\frac{1}{|\det A|}.$$

Заметим, что $(A^{-1})^T A^{-1} = (AA^T)^{-1} = C_\eta^{-1}$, а также $|\det A| = \sqrt{\det C_\eta}$. Поэтому

$$p_\eta(y) = (2\pi)^{-n/2}\frac{1}{\sqrt{\det C_\eta}}\exp\left\{-\frac{\big(C_\eta^{-1}(y-b), (y-b)\big)}{2}\right\}.$$

Поверхности уровня у неё эллипсоиды.

Пусть есть две компоненты η_1, η_2. Есть плотность $p_{\eta_1,\eta_2}(y_1, y_2)$. Хочу плотность отдной компоненты, как это сделать. Ответ

$$\int\limits_{-\infty}^{+\infty} p_{\eta_1,\eta_2}(y_1, y_2)\,dy_2 = p_{\eta_1}(y_1).$$

Как доказать, что компонента нормального вектора нормальна. Нужно воспользоваться преобразованием Фурье.

$$f_\xi(t) = \int\limits_{-\infty}^{+\infty} e^{itx}p_\xi(x)\,dx = \mathbb{E}e^{i(t,\xi)}.$$

Когда t, ξ — одномерные, запись в виде скалярного произведения (t, ξ) ни к чему, но возможна. Итак обобщение

$$f_\xi(t) = \mathbb{E}e^{i(t,\xi)} = \int\limits_{\mathbb{R}^n} e^{i(t,x)}p\xi(x)\,dx.$$

Для станлартной нормальной случайной величины

$$f_\xi(t) = \mathbb{E}e^{i(\xi,t)} = \prod_{i=1}^n e^{-\frac{t_i^2}{2}} = e^{-\frac{1}{2}(t,t)}.$$

Для общего случая нормальной случайной величины

$$f_\eta(t) = \mathbb{E}\exp\{i(t, A\xi + b)\} = e^{i(t,b)}\mathbb{E}\exp\big\{i(t, A\xi)\big\} = e^{i(t,b)}\exp\left\{-\frac{(A^T t, A^T t)}{2}\right\} = e^{i(t,b)}\exp\left\{-\frac{(C_\eta t, t)}{2}\right\}.$$

Чтобы получить характеристическую функцию только по части компонент, надо просто занулить соответствующие t_i.

Утверждение 13.1. *Если нормальные случайные величины не кореллируют, то они независимы.*

Утверждение 13.2. *Любой подвектор нормального вектора нормален*

Пусть мы исследуем p через $\ln\frac{p}{1-p}$. Есть больные, есть здоровые. Ищем шанс, как функцию каких-то параметров.

14 Теория оценок

Есть измерения $x_1, \ldots, x_n$. Предположим, что мы знаем какой характер у совместной плотности $p(x_1, \ldots, x_n; \theta)$ случайных величин, реализацией которых мы предполагаем выборку, но не знаем один параметр этой плотности. Допустим, мы предположил, что закон нормальный у одного x, тогда

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad \theta = \{a, \sigma\}.$$

А что такое плотность. $\mathbb{P}(\xi \in B) = \int\limits_B p_\xi(x)\, dx$, где $\xi = \xi(\omega) \in X$.

Пусть $X = \{0, 1, 2, \ldots\}$ и $\mathbb{P}(\xi = x) = \frac{\lambda^x}{x!} e^{-\lambda}$. Какая здесь плотность? Какая здесь мера? На множестве X можно поместить в каждой точке массу единицу. Тогда будет интаграл по этой мере-массе.

Итого

$$\int\limits_B p(x, \theta)\, dx = \mathbb{P}(B).$$

Это такое обобщение понятия плотности и вероятности.

Итак есть наблюдения $x_1, \ldots, x_n$. Мы уверены, что некая формула, зависящая от x и θ, даёт нам разумное описание тех вероятностей, которые нам дают $x_1, \ldots, x_n$.

Оценкой называют любую функцию от измерения

$$\hat{\theta} = \varphi(x_1, \ldots, x_n).$$

У нас постоянно будут возникать выражения $\log p(x, \theta)$. Это натуральный логарифм, просто следуем обозначениям учебника Себастьянова. Однако $p(x, \theta) \geqslant 0$. Равняться нулю может. Для показательного закона

$$p(x, \theta) = \theta e^{-\theta x} J_{x \geqslant 0}.$$

Надо понять, какой смысл за этим будет стоять.

Наблюдения $(x_1, \ldots, x_n) \in X$. Выделяем в множестве X такую область. Выделяем такие $x \in X$, где плотность $p(x_1, \ldots, x_n, \theta) > 0$. В дальнейшем эта область не обозначается никак. Надо экономить буквы. Будем далее вообще предполагать, что эта область не зависит от θ. Для равномерного распределения это не так

$$p(x, \theta) = \frac{1}{\theta} J_{x \in [0, \theta]}.$$

Тут область, где $p > 0$, зависит от θ. Для θ есть интуитивная оценка $\max\{x_1, \ldots, x_n\}$. Она будет точнее, чем то, что я далее буду рассказывать.

Интеграл по области $\{p(x, \theta) > 0\}$ $\int \ldots dx$ будет неопределённым. Мы будем писать без пределов интеграл, который берётся по всей области $\{p > 0\}$.

Мы сейчас выведем, точнее всего не может быть оценка. При дополнительных предположениях о гладкости плотности по оцениваемому параметру. Имеем равенство $\int p(x, \theta)\, dx = 1$. Дифференцируем его

$$\frac{d}{d\theta} \int p(x, \theta)\, dx = \int \frac{\partial p}{\partial \theta}\, dx = 0.$$

Если бы область зависела от θ, то производную мы бы так не протащили.

Поскольку оценка в конечном итоге получается случайной величиной. Отсюда можно выписать основные характеристики оценки.

14.1 Мат ожидание оценки

$g(\theta) = \mathbb{E}\hat{\theta} = \mathbb{E}\varphi(x) = \int\limits_X \varphi(x) p(x, \theta)\, dx$. Здесь до того, как писали интеграл, под x понималась случайная величина, а под интегралом это уже переменная интегрирования. Далее выделим те иксы, которые дают плотность ноль. Ничего не изменится, если их убрать. И мы договорились ничего не подписывать, когда интеграл берётся по множеству $p > 0$.

$$g(\theta) = \mathbb{E}(\hat{\theta}) = \mathbb{E}\varphi(x) = \int \varphi(x) p(x, \theta)\, dx = \int \varphi p\, dx.$$

Мы надеемся, что $g(\theta) \approx \theta$, хотя бы для большого числа наблюдений.

14.2 Дисперсия оценки

Давайте напишем, что такое производная g

$$g'(\theta) = \int \varphi \frac{\partial p}{\partial \theta}\, dx.$$

Пришло время воспользоваться понятием логарифма. Вводится величина, называемая информацией Фишера. Обозначается $J(\theta) = \mathbb{E}\left(\frac{\partial}{\partial\theta}\log p(x,\theta)\right)^2$. Выпишем, что получится

$$J(\theta) = \int \left(\frac{p'_\theta}{p}\right)^2 p\, dx$$

Если $g(\theta) = 0$, оценка называется несмещённой.

Посмотрим снова на производную

$$g'(\theta) = \int \varphi \frac{\partial p}{\partial \theta}\, dx = \int \varphi \frac{\partial \log p}{\partial \theta} p(x,\theta)\, dx.$$

Но у нас есть равенство

$$0 = \int \frac{\partial p}{\partial \theta}\, dx = \int \frac{\partial \log p}{\partial \theta} p(x,\theta)\, dx.$$

Умножим его на $g(\theta)$ и привавим к производной.

$$g'(\theta) = \int (\varphi - g) \frac{\partial \log p}{\partial \theta} p(x,\theta)\, dx.$$

Ну а это скалярное произведение в L_2 с весом $p(x,\theta)$. Значит,

$$\left[g'(\theta)\right]^2 \leqslant \int \left[\varphi(x) - g(x)\right]^2 p\, dx \cdot J(\theta).$$

Второй множитель оказался информацией Фишера. А первый множитель есть дисперсия. Получается следующее неравенство

$$\left[g'(\theta)\right]^2 \leqslant \mathbb{D}\hat\theta \cdot J(\theta).$$

Или, эквивалентно, $\mathbb{D}\hat\theta \geqslant \frac{\left[g'(\theta)\right]^2}{J(\theta)}$. Это знаменитое неравенство Рао—Крамера.

Вот мы хотим $g(\theta) \approx \theta$. Бывает так, что функции близки, а производные у них не близки, но в приложениях этого не бывает. Поэтому если $g'(\theta)$, то

$$\mathbb{D}\hat\theta \geqslant \frac{1}{J(\theta)}.$$

Вот так получается наилучшая точность для оценки.

Теперь возьмём вторую производную

$$\frac{\partial^2}{\partial\theta^2}\log p = \frac{\partial}{\partial\theta}\left(\frac{p'_\theta}{p}\right) = \frac{p''_{\theta\theta}p - (p'_\theta)^2}{p^2} = \frac{p''_{\theta\theta}}{p} - \left(\frac{p'_\theta}{p}\right)^2.$$

Однако $\int p''_{\theta\theta}\, dx = 0$, поэтому информацию Фишера ещё можем написать

$$J(\theta) = -\mathbb{E}\frac{\partial^2 \log p}{\partial\theta^2}.$$

Есть у нас наблюдения $x_1, \ldots, x_n$, запишем плотность, будто все распределены нормально

$$p(x, a, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum(x_i - a)^2}{2\sigma^2}\right\}.$$

В качестве оценок возьмём $\hat a = \frac{1}{n}\sum s_i$, $\mathbb{D}\hat a = \frac{\sigma^2}{n}$. Тогда информаци Фишера

$$J = -\mathbb{E}\left\{C - \frac{\sum(x_i - a)^2}{2\sigma^2}\right\}''_{aa} = \frac{n\sigma^2}{.}$$

А у нас есть оценка $\mathbb{D}\hat a \geqslant \frac{\sigma^2}{n}$. Ну значит предложенная оценка с помощью среднего является минимизирующей дисперсию.

Оценка с минимальной дисперсией называется **эффективный**.

15 Метод максимального правдоподобия

Полное доказательство сегодня не успею. Обсудим, что это вообще.

Пусть есть $x_1, \ldots, x_n$. Пусть они описываются плотностью $p(x, \theta)$. Подставим наблюдение в плотность, останется только функция от одного θ. Обратим это в максимум по θ.

$$p(x, \theta) \to \max_\theta.$$

Полученный $\arg\max_\theta p(x, \theta) = \hat{\theta}(x_1, \ldots, x_n)$. Это некоторая оценка.

Пусть плотность одного наблюдения $p(x_i, \theta)$. Тогда плотность совокупности независимых наблюдений есть произведение.

$$\prod_{i=1}^{n} p(x_i, \theta) = L(\theta).$$

Это называется функцией правдоподобия. Пусть θ_0 — истинное значение оцениваемого параметра. Возьмём логарифм от функции правдоподобия

$$\log L(\theta) = \sum_{i=1}^{n} \log p(x_i, \theta).$$

И возьмём производную по θ

$$\frac{1}{n} \frac{\partial}{\partial \theta} \log L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log p(x_i, \theta) = 0.$$

Распишем по формуле Тейлора то, что здесь у нас написано.

$$\frac{1}{n} \frac{\partial}{\partial \theta} \log L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log p(x_i, \theta) + (\theta - \theta_0) \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \log p(x_i, \theta) + \frac{1}{2} (\theta - \theta_0)^2 \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^3}{\partial \theta^3} \log p(x_i, \theta_i) = 0.$$

Мы допустим, что $\left| \frac{\partial^3}{\partial \theta^3} \log p(x, \theta) \right| \leqslant M(x)$ так, что если мы подставим вместо x случайную величину и посчитали бы мат ожидание, то получилась бы ограниченная величина. Тогда получится, что остаточный член мал

Дело в том, что $\mathbb{E} \frac{\partial}{\partial \theta} \log p(x_i, \theta) = 0$, а дисперсия такой суммы равна суммы дисперсий. Берём мат ожидание. Получаем по втором слагаемом информацию Фишера. ПОлучается, что если только сумму первых двух членов приравнять к нулю, то получается, что

$$\widetilde{\theta} - \theta_0 \sim \frac{1}{\sqrt{n}}$$