Т.к. вопросы 2,4 схожи, 7,8 и 18 схожи, то разбираются вместе. Поэтому второй вопрос там же, где четвёртый, а 7-й и 8-й – где 18-й.

Вопрос 1

1. Понятие случайного события. Алгебры и сигма-алгебры событий. Примеры и свойства алгебр и сигма-алгебр.

Пусть есть множество Ω всех возможных исходов.

Например, отрезок: 

На нём мы можем выделять различные подмножества:

например, такое 

Естественно будет дать определение вероятности как некой функции, сопоставляющей каждому подмножеству Ω число от 0 до 1, т.е. некую вероятность:

р есть: подмножество Р -> число от [0..1]

Однако математики в теории множеств открыли всякие говномножества типа такого (можете мне поверить, у меня был курсач по этой теме):

Не будем обсуждать, каким образом они строятся – ОНИ ЕСТЬ. И вот если мы определим вероятность для них, у нас будут проблемы (не будут выполняться аксиомы вероятности, которые мы позже сформулируем).

Какой выход? Сделать «элитный клуб» подмножеств Ω (куда не пускать «плохие» множества), и определить вероятность только для членов «элиты».

Такой «элитный клуб» называется σ-алгеброй.

Просто алгебра событий - это

Алгеброй событий называется класс $\mathcal{F}$ событий (система множеств), замкнутый относительно операций $A \cup B$ и $\bar{A}$, то есть

1. Из $A, B \in \mathcal{F}$ следует $A \cup B \in \mathcal{F}$,
2. Из $A \in \mathcal{F}$ следует $\bar{A} \in \mathcal{F}$.

А если σ-алгебра, то вдобавок нужно

3. $A_n \in \mathcal{F}, n = 1, 2, \ldots, \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Примеры? Например, если Ω – конечно (например, если мы кидаем кубик, то у нас всего 6 возможных исходов), то «плохих» множеств не бывает, и множество всех подмножеств Ω образует сигма-алгебру.

Вопрос 3

3. Аксиомы вероятности и простейшие следствия из них.

Если у нас есть пространство исходов Ω, а на нём введена алгебра F, то мы можем ввести вероятность как функцию, сопоставляющей каждому событию число от 0 до 1:

1) $\mathrm{P}(\cdot)\colon \mathrm{F} \to [0,1]$, где F σ-алгебра,
2) $0 \le P(A) \le 1, \quad \forall A \in F$.

при этом мы также должны потребовать, чтобы

3) $P(\Omega) = 1$ (нормировка).
4) Если $AB = \emptyset$, то P(A+B) = P(A) + P(B) (аддитивность).
5) σ-аддитивность. Для несовместных событий $A_1, A_2, \ldots, A_n, \ldots$ выполняется:

$$P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j).$$

Четвёртое и пятое требования очень похожи друг на друга, только в 4) объединяются два несовместных события, а в 5) бесконечного много.

Это всё были требования к вероятности в её определении. А вот свойства вероятности:

1) Монотонность: если $A \subset B$, то $P(A) \le P(B)$.
2) Если $A \subset B$, то $P(B \backslash A) = P(B) - P(A)$. Док-во:
3) $P(A) = 1 - P(\bar{A})$ и наоборот. Док-во:
4) $P(A \cup B) = P(A) + P(B) - P(AB)$. Док-во:

Вопросы 2 и 4

2. Понятие предела последовательности событий. Сходимость монотонных последовательностей событий.

Вопрос 2:

Пусть у нас есть бесконечная последовательность вложенных событий B1⊂B2⊂B3

И бесконечная последовательность антивложенных: A1⊂A2⊂A3…

Такие последовательность называются монотонной.

Понятие предела:

Нижний предел: $\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k$

Верхний предел: $\bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k$

Если они совпадают, то просто предел.

Теорема о сходимости монотонных последовательностей:

Последовательность вложенных событий B и антивложенных A сходится, т.е. для любой из них существуют верхний и нижний пределы, и они равны.

Вопрос 4:

4. Теорема о непрерывности вероятности для монотонных и произвольных последовательностей событий.

Вероятность предела есть предел вероятнотей:

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$$

У этой теоремы есть доказательство, и оно, как положено доказательствам из теорвера, страшно занудное. Обсудим его.

Сначала мы доказываем, что $P(\lim_{n\to\infty} A_n)$ равно нижнему пределу:

$$P(A_{\sim}) = \lim_{n\to\infty} P\left(\bigcap_{k\geq n} A_k\right) \leq \varliminf_{n\to\infty} P(A_n)$$

А затем, что нижнему пределу. это будет дольше:

$$\varliminf_{n\to\infty} P(A_n) \le \varlimsup_{n\to\infty} P(A_n) \le \lim_{n\to\infty} P\left(\bigcup_{k\ge n} A_k\right)$$

При замене A_n на объединение уже можем записать предел, потому что объединение – это бесконечно убывающая и ограниченная последовательность:

$$\bigcup_{k\ge n} A_k = D_n, \qquad \lim_{n\to\infty} D_n = \bigcap_{n=1}^{\infty} D_n.$$

В силу непрерывности вероятности относительно монотонной сходимости:

$$\lim_{n\to\infty} P\left(\bigcup_{k\ge n} A_k\right) = P\left(\lim_{n\to\infty} \bigcup_{k\ge n} A_k\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\ge n} A_k\right) = P(\tilde{A}).$$

Обратите внимание, что для нижнего предела мы представляли А в виде ***пересечения*** последовательности Ашек, а для верхнего – в виде ***объединения***

$$\bigcup_{k\ge n} A_k$$

Наконец,

Поскольку верхний и нижний пределы совпадают, все неравенства превращаются в равенства:

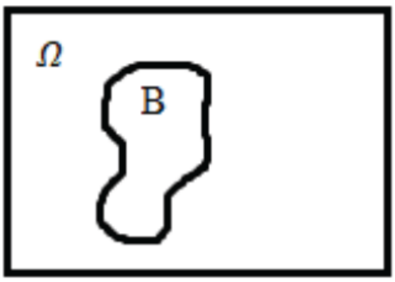
$$\Rightarrow \varliminf_{k\to\infty} P(A_k) = \varlimsup_{k\to\infty} P(A_k) = P(\lim_{n\to\infty} A_k).$$

Вопрос 5

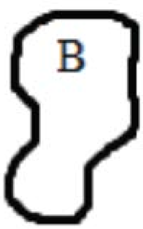
5. Пространство условной вероятности. Формула полной вероятности и формула Байеса.

Пространство условной вероятности.

Было у нас пространство исходов Ω, и вдруг случилось В:



Очевидно, что точки вне В уже тоже недействительно, и нам нужно оставить

лишь само В: 

Но вероятностное пространство состоит из 3 частей:

Ω,
алгебры F
и вероятностей P для каждого события из F.

Коль мы укорачиваем Ω до B, то должны скорректировать и алгебру F с вероятностями P.
В алгебре F каждому событию A ставится в соответствие событие AB, а вместо вероятности P(A) приходится вводить понятие условной вероятности P(A|B).

Полный набор (группа) несовместных событий. Формула полной вероятности
Пожалуй, поясню это на примере.

Пример. Экзамен по ОММ в группе студента принимают Токмачёв, Боголюбов Александр Николаевич и Мухартова.
Вероятность попасть к Токмачу – 45%
Вероятность попасть к Боголюбову – 40%
Вероятность попасть к Мухартовой – 15%.
Они образуют полную группу несовместных событий – это все возможные исходы (четвёртого преподавателя нет!) и при этом несовместны

Вероятность получить «4» или «5» у Токмача 20%, у Боголюбова 70%, у Мухартовой 80%.

Тогда вероятность получить хорошую или отличную оценку считается как 0,45*0,2+0,4*0,7+0,15*0,8=0,09+0,28+0,12=0,49.

Теперь запишем это «буквами». Вшки образуют полную группу несовместных событий, если:

- $B_i \cap B_j = \emptyset,$
- $\sum B_j = \Omega.$

Тогда вероятность события A с их помощью можно рассчить как

$$P(A) = \sum_{j=1}^{\infty} P(AB_j) = \sum_{j=1}^{\infty} P\left(A\middle|B_j\right)P\left(B_j\right)$$

Формула Байеса

$$P(AB) = P(A|B)P(B),$$

$$P(AB) = P(B|A)P(A),$$

$$P(A|B)P(B) = P(B|A)P(A).$$

позволяет переходить от P(A|B) к P(B|A).

Вопрос 6

6. Независимость событий: попарная и в совокупности. Свойства независимых событий.

Определение ***попарной*** независимости, которое все знают: P(AB)=P(A)P(B)

Свойства:

1) $\text{A, B – независимы} \Longrightarrow P(A|B) = \dfrac{P(AB)}{P(B)} = \dfrac{\mathrm{P(A)P(B)}}{\mathrm{P(B)}} = \mathrm{P(A)}$

2) Пустое множество и 3) пространство всех исходов Ω независимы от любого события A:

$$P(A)P(\underset{=0}{\varnothing}) = 0 = P(A\varnothing)$$

$$P(A\Omega) = P(A) = P(A)P(\Omega)$$

Как говорится, независимы от всех и могут плевать на всех бомж (нулевое множество) и арабский шейх (пространство всех исходов Ω).

4) $\text{A, B – независимые} \Longrightarrow \text{A}, \bar{B};\ \bar{A}, \bar{B};\ \text{B}, \bar{A}$ – независимы.

Это что касается попарная независимость, а есть ещё независимость в совокупности.

События $A_1, A_2, \dots, A_n$ независимы в совокупности, если:

для любых пар

$P(A_iA_j)=P(A_i)(A_j)$

для любых троек

$P(A_iA_jA_k)=P(A_i)(A_j)(A_k)$

…

для любых наборов из n-1 событий

$P(A_iA_j \dots A_k)=P(A_i)*(A_j)*\dots*(A_k)$

и, наконец, для набора из n событий

$P(A_iA_j \dots A_k)=P(A_i)*(A_j)*\dots*(A_k)$

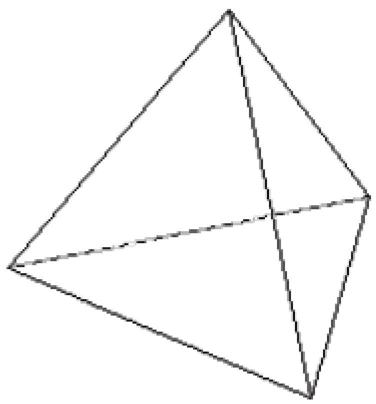
У Чуличкова сказано то же самое, но в гораздо более запутанной формулировке:

$$\forall k = 2,3,4,\dots,n \qquad \forall i_1, i_2, \dots, i_k \in \{1,\dots,n\} \qquad P(A_{i_1}A_{i_2}\dots A_{i_k}) = \prod_{k=1}^{n} P(A_{i_k}).$$

Такое огромное количество равенств требуется, во-первых, чтобы исключить вероятность попадания пустого множества, во-вторых, величины могут быть попарно независимы, но зависимы в совокупности:

Из независимости в совокупности $\Rightarrow$ попарная независимость. Обратное неверно.

Пример: тетраэдр Бернштейна, что-то около игральной кости:



Одна из граней красная, вторая синяя, третья зелёная, а четвёртая покрашена во все 3 цвета.

Бросают кубик и смотрят на наличие цвета на нижней грани.
Вероятность найти там красный: 2/4=½ (если выпадет красная грань или разноцветная)
Вероятность найти там зелёный: 2/4=½ (если выпадет зелёная грань или разноцветная)
Вероятность найти там красный и зелёный одновременно: ¼ (только разноцветная грань)
Проверяем P(зел)*P(красн)=P(зел И красн)
½*1/2=1/4 – сходится.

Аналогично для двух других пар цветов.

Но вот P(зел)*P(красн)*P(син)=1/2*1/2*1/2=1/8, а P(зел И красн И син)=1/4. Есть попарная независимость, но зависимость в совокупности!

Вопрос 9

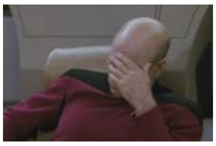
9. Определение случайной величины. Свойства функции распределения. Плотность вероятности и ее свойства.

Вопрос обманчиво кажется простым – он очень коварный.

История от Газарян:

Экзаменатор: Что такое случайная величина?

Студент: Ну это хрень такая, ξ. Она может быть то одной, то другой. Случайная. random.org знаете?

Экзаменатор: 

Дадим же нормальное определение случайной величине!

Откроем Чуличкова:

$$(\Omega, F, P) - \text{вероятностное пространство.}$$

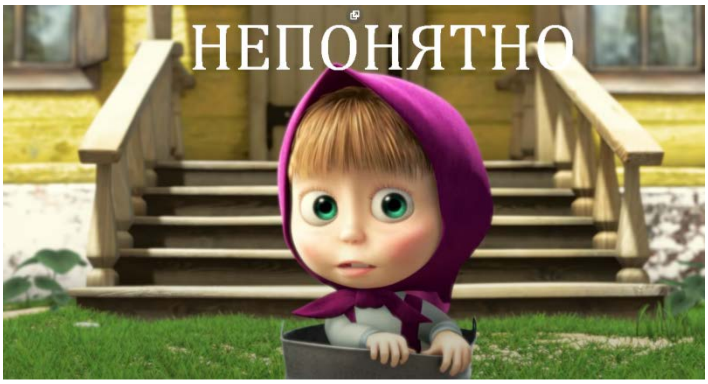
Опр.: Случайная величина $\xi(\cdot)$- это функция, задана $\forall\omega \in \Omega$ и принимает числовые значения $\xi(\omega) \in R_1$, причем $\forall x \in R_1$ $\{\omega: \xi(\omega) < x\} \in F$.

Все ω, $\xi(\omega)$ которых попадают в эту полупрямую, должны попадать в алгебру F. Если вместо полупрямой взять ее дополнение от x до ∞, то оно тоже должно принадлежать алгебре F, т.е. $\{\omega: \xi(\omega) > x\} \in F$. Любые пересечения этих полупрямых тоже будут таковы, что ω принадлежат алгебре F. Поскольку для любых элементов из F задана вероятность, то она переносится на все борелевские множества.

$$P(\xi < x) - \text{определена.}$$

Опр.: Функцией распределения случайной величины ξ называется функция

$$\forall x \in R_1: \quad F_\xi(x) = P(\xi < x).$$



Обратите внимание, что Чуличков вводит случайную величину не просто на множестве исходов Ω, а на вероятностном пространстве (Ω, F, P), на котором ***уже*** (!!!) введена вероятность как функция от события:

Опр.: Вероятностью называется функция $P(\cdot)$:

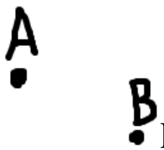
1) $P(\cdot)\colon F \to [0,1]$, где F σ-алгебра,
2) $0 \le P(A) \le 1, \ \forall A \in F$.
3) $P(\Omega) = 1$ (нормировка).
4) Если $AB = \emptyset$, то $P(A+B) = P(A) + P(B)$ (аддитивность).
5) σ-аддитивность. Для несовместных событий $A_1, A_2, \ldots, A_n, ..$ выполняется:

Кстати, тоже нюанс: от события, а не от исхода. Событие – это подмножество Ω (естественно, «хорошее»), а исход – лишь точка. Событие может включать в себя множество исходов, конечное или бесконечное.

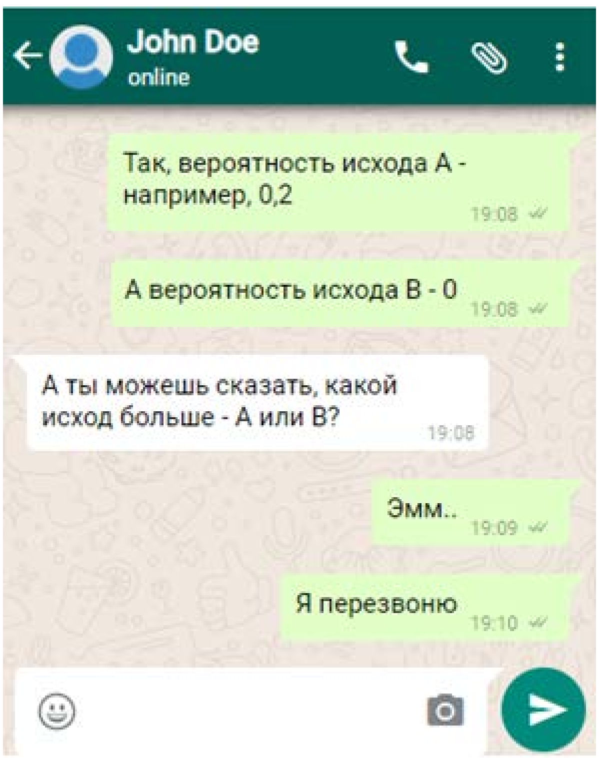
Ну ладно, мы отвлеклись. Главное у нас что?
Случайной величины ещё нет, а вероятность уже есть!
Тогда что же такое случайная величина и нахрен она нужна, когда вероятность у нас уже так и есть и без неё?

Случайная величина не имеет никакого отношения к вероятности. У неё совершенно другая задача – упорядочить множество, приведя его к отрезке [0..1].

До случайной величины мы могли сказать вероятность каждого исхода!



Вероятность каждого события! Но вопрос «а какая точка из двух точек множества Ω больше?» вводил нас в ступор:



Ну это как-то странно. А в идеале – вообще отобразить наше пространство исходов Ω на прямую R_1. Если нам удастся это сделать, то у нас и возможность сравнить точки появится, и возможность ввести плотность вероятности с интегральной функцией распределения.

Именно это и делает случайная величина! Она ставит в соответствие каждому исходу число: ξ(исход)->число. Тем самым она взаимно однозначно отображает Ω на вещественную прямую R_1.
Никакого отношения случайная величина к вероятности не имеет, у неё другая задача: пронумеровывать все исходы, ставя каждому в соответствие его номер.

И теперь, когда у каждого исхода из Ω есть номер, мы можем перейти от опостылевшего пространства исходов Ω, с которым неудобно работать, к R_1, с которой работать очень удобно. Можем также ввести понятия интегральной функции распределения и плотности вероятности.

Пример. Представим, что мы не имеем никакого представления о единицах измерениях длины – метрах, астрономической единице, световом годе и прочем парсеке.

Мы берём кирпич и меряем её длину линейкой в разных местах. Мы знаем, что будем получать всё время разные результаты, а погрешность будет подчиняться нормальному распределению.
Стоп! Нормальное распределение – это вообще-то функция от числа:

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

вот, видите, в формуле число x!

Меж тем погрешность – это не число, а размерная физическая величина (длина). Из-за этого мы никак не можем установить связь между нормальным распределением и распределением погрешностей в эксперименте.

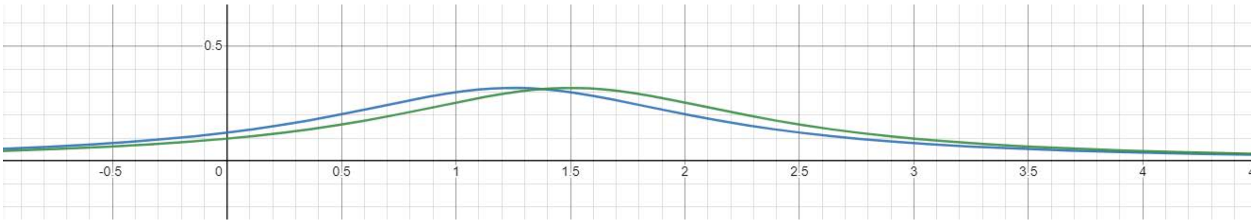
И тут появляется Парижская палата мер и весов с метром, который нам позволяет перейти от размерной длины к безразмерному числу! Всё, теперь у нас есть этот «мостик»! Именно он в данном случае и есть случайная величина.

Не правда, звучит странно – случайная величина есть упорядочивающее отображение, а к вероятности не имеющее никакого отношения? Ну что ж, такова терминология…

А что такое тогда две случайные величины, ξ и η? Два способа нумерации одно и того же множества Ω, что ли? Именно. Кстати, теперь по-новому начинает играть понятие ковариации – чем больше общего имеют нумерации ξ и η (т.е. чем большему количеству точек они дадут общий номер), тем больше будет $\mathrm{cov}(\xi , \eta)$, и наоборот.

Более корректно: если у точек, которым ξ и η дали одинаковую или почти одинаковую вероятность, будет одна или почти одна и и та же вероятность, $\mathrm{cov}(\xi , \eta)$ окажется большой.

Например, пусть ξ дала наиболее вероятной точке номер 1,25, а η номер 1,5. Вообще η даёт номера всем точкам на 0,25 больше, чем ξ.
Нарисуем графики плотностей вероятностей, составленные по двум случайным величинам:

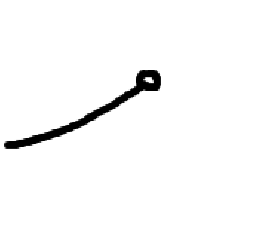


Чёт похоже… У них наверняка ковариация окажется нехилой ☺

Я сказал «плотности вероятностей, составленные по двум случайным величинам». Это более корректно: вероятность была до этого, случайная величина лишь пронумеровала исходы, отобразив Ω на R_1. Но чаще говорят «плотность вероятности случайной величины». Это короче… но теперь вы знаете, что на самом деле это не совсем корректно ☺

Когда мы пронумеровали пространство событий, у нас автоматически появляется интегральная функция распределения. От неё требуется МЗФ от 0 до 1:

$0 \le F_\xi(x) \le 1$ и неубывание. Также непрерывность слева:

т.е. так можно  , а так нельзя: 

Плотность вероятности
Производная от интегральной функции распределения.

Какие у неё свойства? Наверное, самое главное – это

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

В 2D аналогично

$$\int_{-\infty}^{+\infty} p(x,y)dxdy = 1$$

Вопрос 10

10. Независимость случайных величин: попарная и в совокупности. Свойства моментов суммы и произведения независимых случайных величин.

Случайные величины ξ_1, ξ_2, … , ξ_N называются независимыми в совокупности, если:

$$F_{\xi_1,\xi_2,\dots,\xi_n}(x_1, \dots, x_n) = \prod_{j=1}^{n} F_{\xi_j}(x_j).$$

Напомним, что:

$$F_{\xi_1,\xi_2,\dots,\xi_n}(x_1, \dots, x_n) = P(\xi_1 < x_1, \dots, \xi_n < x_n),$$

$$F_{\xi_j}(x_j) = P(\xi_j < x_j).$$

Или можно вместо интегральных функций распределения использовать плотности вероятности:

$$P_{\xi_1,\xi_2,\dots,\xi_n}(x_1, \dots, x_n) = \prod_{j=1}^{n} P_{\xi_j}(x_j), \qquad P_j(x_j) - \text{плотность вероятности } \xi_j$$

Попарная же независимость – это когда мы то же самое записываем для двух переменных:

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

или

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

Если у нас есть N переменных, то если независимы в совокупности, то любая пара независима, но из независимости всех пар не следует независимость в совокупности.

Пример: пусть у нас 3 случайных величины: ξ_1, ξ_2, $\xi_1+\xi_2$; причём ξ_1, ξ_2 независимы. Тогда видно, что любая из 3 пар независима, но в целом совокупность зависима.

Моменты.

Опр.: Моментом порядка k случайной величины ξ называется

$$M_\xi^k = \begin{cases} \sum_{j=1}^{\infty} x_j^k p_j\,, & \text{если } \xi - \text{дискретная случайная величина} \\ \int_{-\infty}^{\infty} x^k p_\xi(x)dx\,, & \text{если } \xi - \text{абсолютно непрерывная сл. величина} \end{cases}$$

То есть:

первый момент - $\int_{-\infty}^{+\infty} xp(x)dx$ (кстати, это матожидание)

второй момент - $\int_{-\infty}^{+\infty} x^2p(x)dx$

третий момент - $\int_{-\infty}^{+\infty} x^3p(x)dx$

и так далее…

Как правило, моменты второго и более высокого порядка центрируют:

второй центрированный момент - $\int_{-\infty}^{+\infty}(x-Mx)^2p(x)dx$ - кстати, это дисперсия

третий центрированный момент - $\int_{-\infty}^{+\infty}(x-Mx)^3p(x)dx$

и так далее.

В общем случае

Опр.: k – м центральным моментом случайной величины называется число:

$$M(\xi - M_\xi)^k, \qquad \text{если } M\left|\xi - M_\xi\right|^k < \infty.$$

Центральный момент отличается от простого момента тем, что из значения ξ вычитается математическое ожидание, и дальше усредняется отклонение ξ от своего математического ожидания (т.е. центра).

k-тый момент ($\xi + \eta$) = k-тый момент ξ + k-тый момент η - это равенство выполняется для любых ξ и η. Иная ситуация, если моменты не обычные, а центральные.

k-тый центральный момент ($\xi + \eta$) = k-тый центральный момент ξ + k-тый центральный момент η ***гарантированно*** выполняется лишь, если ,

ξ и η независимы. Если зависимы, то k-тый центральный момент ($\xi + \eta$) может как =, так и != k-тый центральный момент ξ + k-тый центральный момент η.

Вопрос 11

11. Математическое ожидание случайной величины. Свойства математического ожидания. Неравенство Чебышёва.

Матожидание – это момент первого порядка

$\int_{-\infty}^{+\infty} xp(x)dx$ для непрерывной величины,

$\sum_i p_i \omega_i$ для дискретной

Свойства:

1) Линейность:

а) Mξ+Mη=M(ξ+η)

б) aMξ=M(aξ)

2) Если ξ и η независимы, то M(ξη)=M(ξ)M(η). Если ξ и η зависимы, то так делать нельзя, и в этом случае разность M(ξη)-M(ξ)M(η) будет называться ковариацией ξ и η.

3) Неравенство Чебышёва. Вообще у Чебышёва много неравенств в теорвере, но коли 11-й вопрос посвящён матожиданию, то требуется именно то, которое про матожидание:

$$\boxed{P(\,|\xi| > \varepsilon) \le \frac{M_{|\xi\,|}}{\varepsilon}}$$

С этой формулы можно делать оценки, с какой вероятностью ξ улетит от 0 дальше, чем на ε. Универсальной, её конечно, считать нельзя, хотя бы потому, что M|ξ| может не всегда существовать (при существовании при этом Mξ).

Вопрос 12

12. Дисперсия случайной величины и матрица ковариаций случайного вектора, их свойства. Неравенство Коши–Буняковского.

Дисперсия

Конечно, вы знаете, что такое дисперсия:

$\int_{-\infty}^{+\infty} (x - Mx)p(x)dx$ для непрерывной величины,

$\sum_i p_i(\omega_i - M\,\omega_i)$для дискретной

Свойства:

Аддитивность: Dξ+Dη=D(ξ+η), если ξ и η независимы!

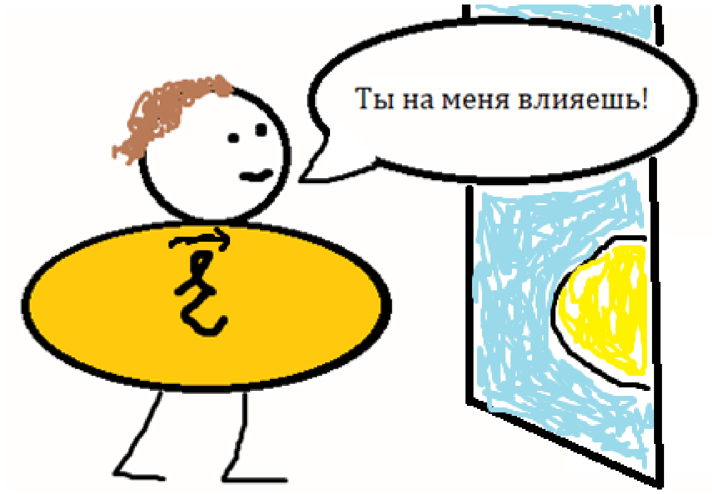
Вынос константы: D(aξ)=a^2Dξ

Что такое ***ковариация?***

Пусть у нас есть случайные величины ξ и β, и мы хотим понять, зависят ли они друг от друга.

Тогда и вычисляется ковариация cov(ξ,β). Её можно вычислять как
M[(ξ-Mξ)(β-Mβ)]
А можно по альтернативной формуле
M(ξβ)-MξMβ
Если она 0, то они не коррелируют (это не значит 100%, что они независимы, величины могут не коррелировать, но быть зависимыми).
Если же ковариация не 0, то 100% какая-то связь есть – величины зависимы.

Можно подсчитать ковариацию и от самого себя. Представим себе героя, смотрящего на себя в зеркале и считающего ковариацию cov(ξ,ξ).



cov(ξ,ξ)= M[(ξ-Mξ)(ξ -Mξ)]=Dξ – обычной дисперсии.

Матрица ковариаций
Сложная тема. Положение осложняется тем, что есть матрица ковариаций одного случайного вектора, а есть матрица ковариаций двух случайных векторов, и во множестве пособий и роликов на Ютубчике эти понятия успешно путают.

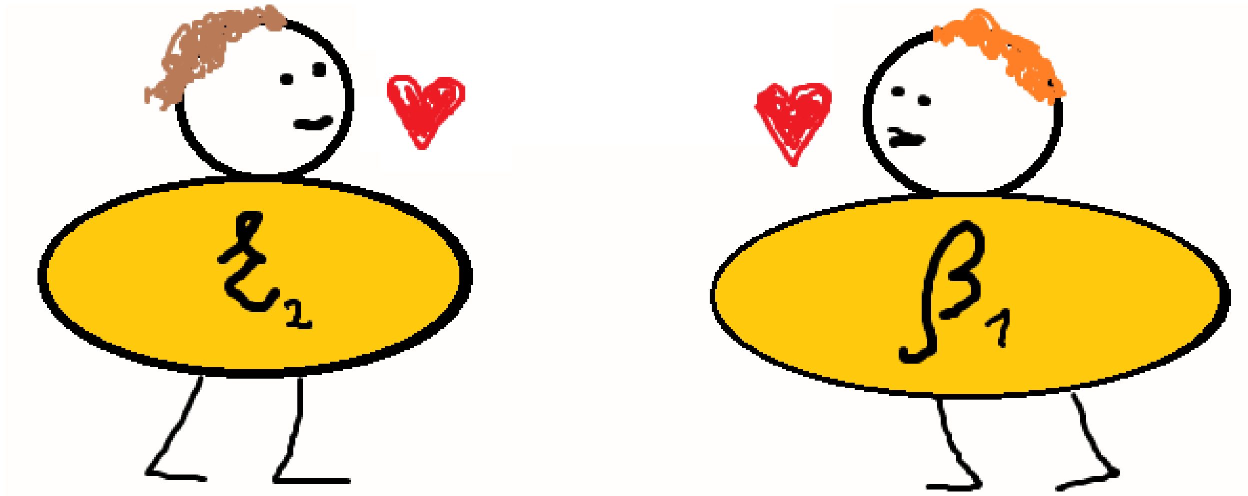
Пусть у нас есть два случайных вектора – например, трёхмерный и двумерный:
$\xi = \{\xi_1, \xi_2, \xi_3\}$

$\boldsymbol{\beta} = \{\beta_1, \beta_2\}$

И мы хотим понять, есть ли между ними какой-то «коннект», какая-то связь.

Причём связь может быть от любой проекции одного вектора к любой проекции другого вектора.

Например, если связь обнаружилась хоть между одной парой



то вектора будут всё равно зависимыми и в итоге будет

Так что придётся считать все попарные ковариации:
$cov(\xi_1,\beta_1\}$
$cov(\xi_1,\beta_2\}$
$cov(\xi_2,\beta_1\}$
$cov(\xi_2,\beta_2\}$
$cov(\xi_3,\beta_1\}$
$cov(\xi_3,\beta_2\}$
Ну а их естественно засунуть в матрицу:

$$\begin{matrix} cov(\xi_1,\beta_1) & cov(\xi_1,\beta_2) \\ cov(\xi_2,\beta_1) & cov(\xi_2,\beta_2) \\ cov(\xi_3,\beta_1) & cov(\xi_3,\beta_2) \end{matrix}$$

Которая и будет матрицей ковариации двух векторов.

Теперь давайте обсудим матрицу ковариации одного вектора. Это частный случай, когда $\boldsymbol{\xi} = \boldsymbol{\beta}$, то есть мы считаем ковариацию случайного вектора от себя же.
Она уже будет квадратной:

$$\begin{matrix} cov(\xi_1,\xi_1) & cov(\xi_1,\xi_2) & cov(\xi_1,\xi_3) \\ cov(\xi_2,\xi_1) & cov(\xi_2,\xi_2) & cov(\xi_2,\xi_3) \\ cov(\xi_3,\xi_1) & cov(\xi_3,\xi_2) & cov(\xi_3,\xi_3) \end{matrix}$$

И вдобавок ещё и симметричной, т.к. $cov(\xi_i,\beta_k) = cov(\xi_k,\beta_i)$

По диагонали, кстати, у нас будут дисперсии:

$$\begin{matrix} D(\xi_1) & cov(\xi_1,\xi_2) & cov(\xi_1,\xi_3) \\ cov(\xi_2,\xi_1) & D(\xi_2) & cov(\xi_2,\xi_3) \\ cov(\xi_3,\xi_1) & cov(\xi_3,\xi_2) & D(\xi_3) \end{matrix}$$

В таком виде – с дисперсиями - матрицу ковариаций часто и приводят. Но это частный случай! В общем случае матрица ковариации считается от двух разных векторных случайных величин и дисперсий там нет:

$$\begin{matrix} cov(\xi_1,\beta_1) & cov(\xi_1,\beta_2) \\ cov(\xi_2,\beta_1) & cov(\xi_2,\beta_2) \\ cov(\xi_3,\beta_1) & cov(\xi_3,\beta_2) \end{matrix}$$

Неравенство Коши-Буняковского-Шварца (КБШ)
Неравенство, всплываемое во многих областях математики. Иностранцы-русофобы выкидывают Буняковского, некоторые наши – Шварца. Я предлагаю быть терпимыми и уважать всех трёх математиков ☺
Пусть есть случайные величины ξ и η. Тогда

$$\left(M_{\xi\eta}\right)^2 \leq M_{\xi^2} M_{\eta^2}$$

Следствием из него является

$$D_\xi D_\eta \geq cov^2(\xi, \eta)$$

Вопрос 13

13. Условные распределения и условные плотности вероятности. Формулы полной вероятности и формулы Байеса для условных плотностей вероятности.

Мы уже разбирали всё это в вопросе 5, здесь всё то же самое, но со случайными величинами, и здесь будут иные формулы.

Если у нас одна случайная величина ξ, то условие нормировки запишется как

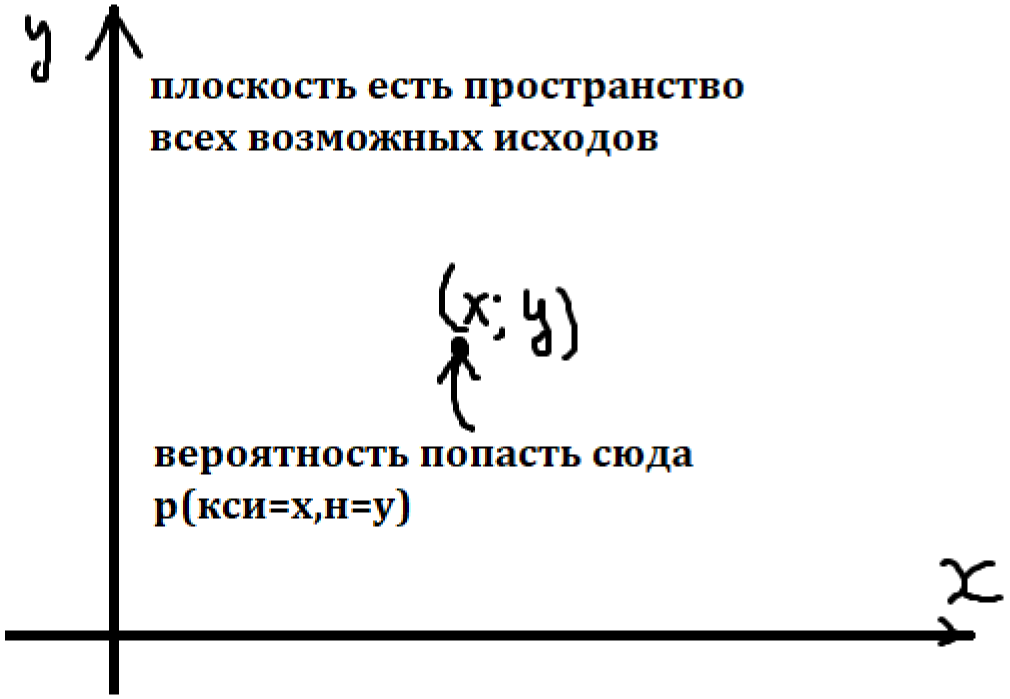
$$\int_{-\infty}^{+\infty} p(\xi = x)dx = 1$$

Где $p(\xi = x)$ - вероятность того, что ξ примет значение x. У Чуличкова чуть иные обозначения: вместо $p(\xi = x)$ у него $p_\xi(x)$.

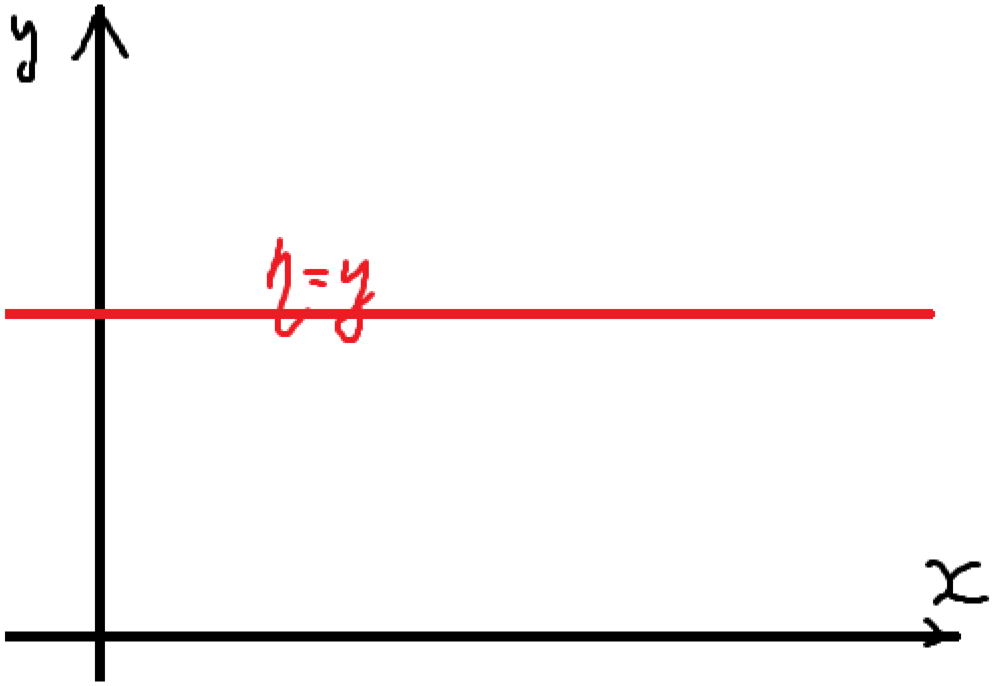
Теперь пусть у нас есть две случайные величины ξ и η. Тогда условие нормировки запишется как

$$\iint_{\text{по плоскости}}^{\text{по плоскости}} p(\xi = x, \eta = y)dxdy = 1$$

Где интегрирование будет вестись по всем возможным исходам (x,y), а $p(\xi = x, \eta = y)$ - вероятность того, что одновременно ξ равно x, а η равно y.



Где же здесь условная вероятность? А вот где. Представим себе, что мы сначала измерили η – она оказалась равна y (вероятность этого $p(\eta = y)$).



А теперь давай измерим ξ ***при условии, что η мы измерили до и она оказалась =y***. Вероятность получить x теперь будет условной, т.к. мы уже знаем. что η=y, и она запишется как $p(\xi = x|\eta = y)$.

В итоге получили η=y, ξ=x, т.е. $p(\xi = x, \eta = y)$. Можем записать, что

$$p(\eta = y) * p(\xi = x|\eta = y) = p(\xi = x, \eta = y)$$

Это аналог формулы

p(B)*p(A|B)=p(AB)

Теперь давайте получим формулу Байеса.

Коль

$$p(\eta = y) * p(\xi = x|\eta = y) = p(\xi = x, \eta = y)$$

то и

$$p(\xi = x) * p(\eta = y|\xi = x) = p(\xi = x, \eta = y)$$

Приравниваем:

$$p(\eta = y) * p(\xi = x|\eta = y) = p(\xi = x) * p(\eta = y|\xi = x)$$

Это формула Байеса, аналог

p(B)*p(A|B)=p(A)*p(B|A)

От нас спрашивают ещё формулу полной вероятности. Давайте вспомним, как она записывалась в вопросе 5:

$$P(A) = \sum_{j=1}^{\infty} P(AB_j) = \sum_{j=1}^{\infty} P(A|B_j)P(B_j)$$

Мы искали вероятность А с помощью условных вероятностей А при наборе несовместных событий B_j. В данном случае в роли несовместных событий B_j будет $\eta = y$.

$$p(\xi = x) = \int_{-\infty}^{+\infty} p(\eta = y) p(\xi = x | \eta = y) dy$$

Вопрос 14

14. Приближение случайной величины функциями от другой случайной величины.

1-й сюжет от Чуличкова: Пусть у нас есть две случайные величины: наблюдаемая ξ и ненаблюдаемая η. Запретный плот сладок: нас ОЧЕНЬ интересует вторая.
Мы один раз измерили ξ, пусть оно оказалось =x. Чему равно η? Точно сказать нельзя, т.к. нельзя померить. А вот оценку Mη мы сказать можем.
Теорема. Лучшей оценкой для матожидания η в ситуации, когда ξ оказалось x, будет M(η|ξ=x).

Не будем останавливаться на страшно занудном доказательстве этой бесполезной теоремы (бесполезной, потому что на практике мы зачастую и M(η|ξ=x) не знаем).
Хотелось бы прояснить фразу «лучшая оценка». Это такая оценка f(ξ), что матожидание квадрата невязки $M(f(\xi)\text{-}n)^2$ было минимальным. Вот эта величина - $M(f(\xi)\text{-}n)^2$ – называется хи-квадрат. Возможно, с ней вы уже знакомы.

2-й сюжет от Чуличкова, гораздо более осмысленный.
Жена Иванова заметила, что муж Иванова приносит каждый месяц разную зарплату. То 100 тыс, то 120, то 110… Спрашивала - муж всё отнекивался. Ну, им надо как-то планировать семейный бюджет (хватит на поездку в Крым летом?), и оценивали случайную величину «зарплата Иванова» η матожиданием, т.е. средним арифметическим (а чем же ещё?).

Но однажды жена пронюхала, что на работе есть понятие «выполненное число задач за месяц» (случайная величина ξ). Ага, подумала она, зарплата η наверняка зависит от ξ! Но как зависит? Предположим, что линейно: η=aξ+b. Ну а далее начинается МНК: Чуличков находит a и b так, чтобы хи-квадрат было как можно меньше.
В итоге для η получаем такую зависимость:

$$M_\eta + \frac{cov(\xi,\eta)}{D_\xi}(\xi - M_\xi)$$

Mη – «базовая» зарплата
ξ-Mξ – успешность работы Иванова в этом месяце. Если ξ-Mξ>0, то Иванов работал хорошо, если ξ-Mξ<0 – плохо.

$\frac{cov(\xi,\eta)}{D_\xi}$ - коэф пропорциональности, показывающий, насколько сильно будет скакать зарплата от месяца к месяцу.

Заметьте, кстати, что если коварииция ξ,η будет равна 0, это будет означать, что начальнику вообще пофиг, что там делает Иванов (это не значит, что он каждый месяц должен платить одну и ту же сумму. Это значит лишь то, что η может зависеть от чего-то другого (например, случайной величины «температура левой пятки начальника»), но уж точно нет от успехов Иванова ξ).
Ещё обратите внимание на Dξ в знаменателе. Помните, как Митин, объясняя МНК, писал в знаменателе что-то в духе $\sum_i w_i{}^2 - (\sum_i w_i)^2$? Это и есть дисперсия, особенно если её записать в явном виде $\sum_i (w_i - Mw)^2$.

Можно привести и более «физичный» пример.

Влияет ли атмосферное давление на высоту рессорных пружин? (Понятно, что масса вагона влияет, а вот про атмосферное давление уже не так очевидно).

$$M_\eta + \frac{cov(\xi,\eta)}{D_\xi}(\xi - M_\xi)$$

Mη – длина пружины (в нормальных условий, при 760 мм.рт.ст)
ξ – атмосферное давление в этот день, ξ-Mξ – отклонение от среднего давления

$\frac{cov(\xi,\eta)}{D_\xi}$ - коэффициент влияния атмосферного давления на высоту пружины.

Если что, примеры про Иванова и рессоры мои, Чуличков их не знает ☺

Ещё Чуличков рассматривает векторные случайные величины $\vec{\xi}$, $\vec{\eta}$ и обобщает полученные результаты для них. В частности, в сюжете 1 лучшей оценкой по наблюдению $\vec{\xi}$ ожидаемо будет $M(\vec{\eta}|\vec{\xi})$, а наилучшая линейная оценка -

$$\hat{\vec{\eta}} = M_{\hat{\vec{\eta}}} + \Sigma_{\eta,\xi}{\Sigma_\xi}^{-1}(\vec{\xi} - M_{\vec{\xi}}).$$

Некоторые пояснения: $(\vec{\xi} - M_{\vec{\xi}})$ есть вектор, а значком $\Sigma_{\eta,\xi}$ Чуличков пометил матрицу ковариаций:

$$(\Sigma_{\xi,\eta})_{ij} = cov(\xi_i, \eta_j)$$

Так что в записи $\hat{\vec{\eta}} = M_{\hat{\vec{\eta}}} + \Sigma_{\eta,\xi}{\Sigma_\xi}^{-1}(\vec{\xi} - M_{\vec{\xi}}).$ скрывается домножение матрицы на вектор.

Вопрос 15

15. Неравенство Чебышёва. Сходимость по вероятности последовательности случайных величин. Закон больших чисел в форме Чебышева. Теорема Бернулли.

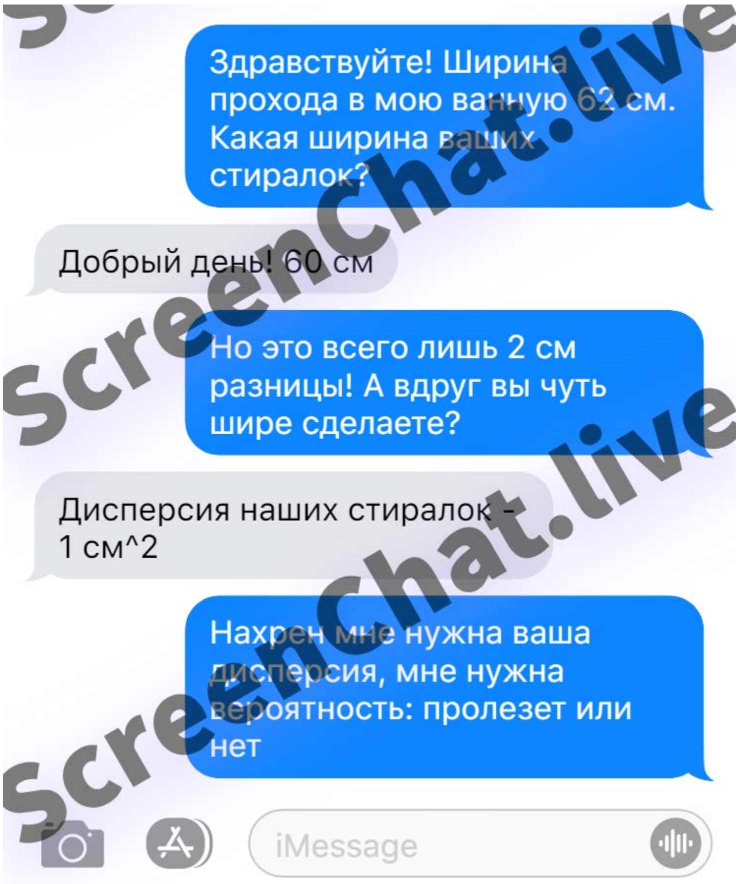
Неравенство Чёбышева

Пусть есть случайная величина ξ, и нам известно её дисперсия Dξ.

Тогда вероятность того, что ξ будет отличаться от маотжидания более, чем на ξ, равна

$$\boxed{P\left(\left|\xi - M_\xi\right| \geq \varepsilon\right) \leq \frac{D_\xi}{\varepsilon^2}}$$ /

Пример - наша старая-добрая сказка про стиральную машину:



Сделаем расчёты. $D\xi=1$ см^2, $\varepsilon=2$ см. Откуда вероятность НЕпролезания можно оценить как $\frac{1\text{ см}^2}{(2\text{ см})^2} = 0{,}25$. Это оценка сверху, в реальности вероятность того, что стиралка не пролезет, ещё меньше.

Сходимость по вероятности

Опр.: последовательность случайных величин $\{\xi_n\}$ называется сходящейся по вероятности к величине ξ_0, если:

$$P(|\xi_n - \xi_0| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0, \qquad \forall\varepsilon > 0.$$

$$P(|\xi_n - \xi_0| < \varepsilon) \xrightarrow[n\to\infty]{} 1 - \text{эквивалентное утверждение.}$$

Сходимость по вероятности обозначается следующим образом: $\xi_n \xrightarrow{P} \xi_0$.

Подробнее мы обсудим это определение в вопросе 16, где будет ещё одна сходимость, и мы как раз сравним их на контрасте.

Закон больших чисел утверждает:

Пусть есть N ***попарно независимых*** случайных величин ξ. Тогда закон больших чисел утверждает, что

$\lim_{N\to\infty}(\frac{1}{n}\sum_{j=1}^{N}\xi_j) = M\xi$

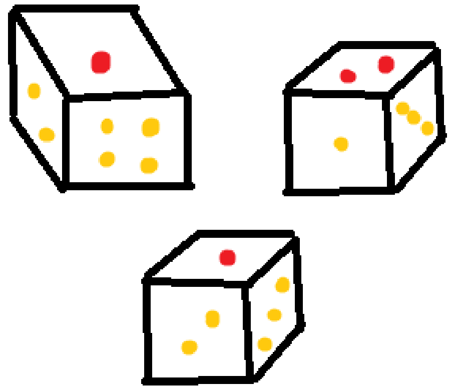
Т.е. среднее арифметическое величин будет стремиться к их матожидание.

Покажем, как работает закон больших чисел на примере:

- Блин, Аня, походу я сломала теорвер!

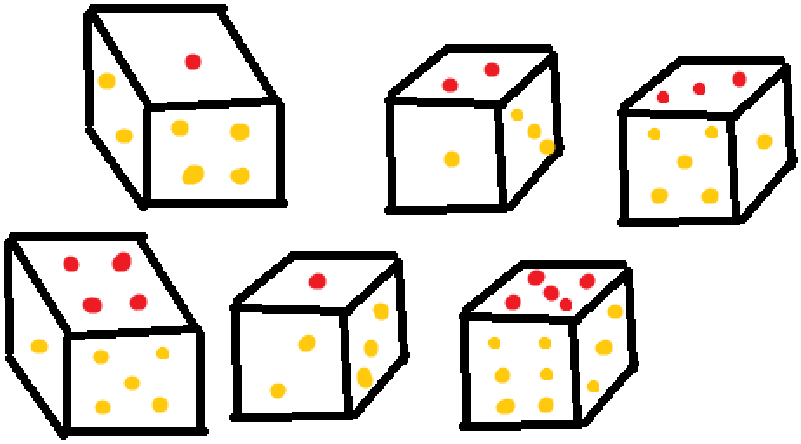
- Почему ты так думаешь, Катя?

- А вот посмотри:



Посмотри на верхние грани трёх кубиков! Среднее арифметическое $\frac{1+1+2}{3} = \frac{4}{3}$. Хотя теоретически (матожидание) должно быть 3,5!

- Ну у тебя небольшая выборка. Давай я тебе ещё кубики принесу.



И теперь среднее уже $\frac{1+1+2+3+4+5}{6} = 3$, что уже гораздо ближе к 3,5.

- Слушай, Ань, а если у меня будет 100 кубиков, среднее, наверное, будет ещё ближе к матожиданию 3,5?
- Да, Кать. Именно так и работает закон больших чисел!

ЗБЧ $\lim_{N\to\infty}(\frac{1}{n}\sum_{j=1}^{N}\xi_j) = M\xi$ можно записать иначе, перекинув матожидание налево и внеся под знак суммы: $\frac{1}{n}\sum_{j=1}^{n}(\xi_j - \mu_j)\xrightarrow{P}0$. Именно так и записывает ЗБЧ Чуличков. Обратите внимание на стрелочку: $\xrightarrow{P}$. Это и есть сходимость по вероятности.

Теорема Бернулли

является частным случае ЗБЧ, на случай, когда у ξ есть только два исхода: 1 с вероятностью p и 0 с вероятьностью 1-p.

Тогда среднее число успехов в серии из N испытаний при стремлении N->∞ стремится к p.

Пример:
На китайской фабрике выпускают поддельные Айфоны. Вероятность брака 20%. Если мы возьмём партию из 5 Айфонов, то, скорее всего, средняя доля бракованных будет 20% (т.е. 1 Айфон). А может быть, и 0, а может быть, и 40% (если 2 Айфона) и т.д.
А вот если мы возьмём партию из 50000 Айфонов, то тут мы уже можем сказать гораздо увереннее, что в среднем будет 20% бракованных айфонов (т.е. 10 тысяч). Вероятности того, доля бракованных айфонов будет сильно отличаться от 20% (скажем, 0% или 20%) будут крайне малы.

Ещё один способ понять ЗБЧ в том виде, в котором его записывает Чуличков:

$$\frac{1}{n}\sum_{j=1}^{n}(\xi_j - \mu_j) \xrightarrow{P} 0$$

Среднее арифметическое отклонений стремится к 0.
Представьте себе много-много случайных величин. Т.к. они случайные, то они могут немного отклоняться от своего матожидания вправо-влево. Так вот, мы эти отклонения суммируем, и после усреднения получим 0:

Какие-то отклонения будут >0, какие-то <0 (как в басне про лебедя, рака и щуку), но в среднем будет 0.

Вопрос 16

16. Лемма Бореля–Кантелли, усиленный закон больших чисел.
Усиленный ЗБЧ очень похож на простой ЗБЧ. Вот простой ЗБЧ:

$$\frac{1}{n}\sum_{j=1}^{n}(\xi_j - \mu_j) \xrightarrow{P} 0$$

А вот усиленный ЗБЧ:

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k - M_k)\xrightarrow{\text{п.н.}} = 0).$$

Разница лишь в сходимости!
В простом ЗБЧ используется сходимость по вероятности, в усиленном – сходимость «почти наверное» (отсюда сокращение «п.н.»).

Что же такое сходимость «почти наверное»? Её можно сравнить с поточечной сходимостью (вспоминаем матан-3):
Последовательность функций $f_n(x)$ поточечно сходится к f(x) на отрезке [a..b], если для любой точки x на отрезке $\lim_{n\to\infty} f_n(x) \to f(x)$.
Переформулируем язык теорвера:
Последовательность интегральных функций распределения $f_n(\omega)$ поточечно (почти наверное) сходится к интегральной функции распределения f(ω) на множестве исходов Ω, если для любого исхода ω на отрезке $\lim_{n\to\infty} f_n(\omega) \to f(\omega)$.
Т.к. случайная величина ⇔ интегральная функция распределения (вторая случайную величину и задаёт), то
Последовательность случайных величин $\xi_n(\omega)$ поточечно (почти наверное) сходится к случайной величине ξ(ω) на множестве исходов Ω, если для любого исхода ω на отрезке $\lim_{n\to\infty} \xi_n(\omega) \to \xi(\omega)$.
Последний штрих: условие «для всех ω на Ω» заменим более слабым: «ω, для которых это верно, образуют множество, вероятность которого 1»:
Последовательность случайных величин $\xi_n(\omega)$ почти наверное сходится к случайной величине ξ(ω) на множестве исходов Ω, если мн-во ω таких, что $\lim_{n\to\infty} \xi_n(\omega) \to \xi(\omega)$, образует множество вероятностью 1.

Например, если из отрезка выколоть несколько точек

то вероятность оставшихся точек всё равно будет 1. Поэтому и «ПОЧТИ наверное». Другое название этой сходимости – «сходимость с вероятностью 1».

Если сходимость «почти наверное» - аналог поточечной сходимости (и то ослабленный – в некоторых точках можно «не сходиться»), то сходимость по вероятности – та, что была раньше, в вопросе 15 – может показаться аналогом равномерной сходимости. Сравним:
Последовательность функций $f_n(x)$ равномерно сходится к f(x) на отрезке [a..b], если $\lim_{n\to\infty} sup|f_n(x) - f(x)| = 0$

Последовательность случайных величин ξ_n сходится к ξ по вероятности, если $P(|\xi_n - \xi_0| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$.

Однако обращаю внимание на отсутствие супремума в сходимости по вероятности. На самом деле сходимость по вероятности слабее. Именно поэтому усиленный ЗБЧ и называется усиленным – там используется сходимость «почти наверное», которая помощнее.

Лемма Бореля-Кантелли особого значения не имеет, лишь помогает в доказательстве усиленного ЗБЧ. Сформулируем её:

Пусть

$A_1, A_2, \dots, A_n, \dots$ - последовательность событий, такая что ряд $\sum\limits_{k=1}^{\infty} P(A_k)$ сходится.

Тогда с вероятностью 1 происходит конечное число событий.

Лемма достаточно очевидная: чтобы ряд сходился, вероятность $P(A_k)$ должна с ростом индекса k достаточно быстро падать, а тогда вероятность бесконечного числа событий… крайне мала. Доказательство в лучших традициях теорвера крайнее занудное, надо возюкаться с верхними пределами.

Вопрос 17

17. Сходимость по распределению последовательности случайных величин. Характеристические функции и их свойства.

Ещё одна сходимость – по распределению. Она основана на принципе «хм, у каждой случайной величины есть интегральная функция распределения, для которой есть определения из матана-3»!

Опр.: $\xi_n \xrightarrow[n\to\infty]{} \xi_0$ по распределению, если

$F_{\xi_n}(x)$ – функция распределения ξ_n,

$F_{\xi_0}(x)$ – функция распределения ξ_0,

$F_{\xi_n}(x) \xrightarrow[n\to\infty]{} F_{\xi_0}(x)$ в каждой точке непрерывности $F_{\xi_0}(x)$.

Сходимость по распределению обозначается следующим образом: $\xi_n \xrightarrow{d} \xi_0$

А есть ещё одна сходимость, четвёртая:

Сходимость в среднем квадратичном:

$$\xi_n \xrightarrow{\text{с.к.}} \xi_0, \qquad \text{при } n \to \infty$$

$$\text{если } M(\xi_n - \xi_0)^2 \xrightarrow[n\to\infty]{} 0.$$

Связь между 4 типами сходимости:

$$\begin{matrix} \xi_n \xrightarrow{\text{с.к.}} \xi_0 \Rightarrow \\ \xi_n \xrightarrow{\text{п.н.}} \xi_0 \Rightarrow \end{matrix} \quad \xi_n \xrightarrow{P} \xi_0 \Rightarrow \xi_n \xrightarrow{d} \xi_0$$

Самой сильной оказалась сходимость по распределению, т.е. та, где нет вероятностей, а есть матан-3. (Всё логично – Чуличков и Сердобольская добрые, а Щепетилов сильный ☺)

Теперь про характеристические функции.

Характеристические функции случайных величин

Опр.: Характеристической функцией случайной величины ξ называется функция, заданная на всей числовой прямой:

$$f_\xi(t) = Me^{it\xi} = M(\cos t\xi + i\sin t\xi), \qquad t \in (-\infty, \infty).$$

$$\xi = Re\xi + i\, Im\xi, \qquad \text{тогда } M_\xi = M\, Re\xi + iM\, Im\xi.$$

$$f_\xi(t) = \begin{cases} \sum\limits_{k=1}^{\infty} e^{itx_k} p_k, & \text{дискретная} \\ \int\limits_{-\infty}^{\infty} e^{itx} p(x)dx, & \text{абсолютно непрерывная} \end{cases}$$

Свойства характеристических функций:

1) Характеристическая функция $\exists\ \forall\xi$.

Чтобы это показать, надо доказать что существует мат ожидание от комплексной экспоненты:

$$M\left|e^{it\xi}\right| = M1 = 1 - \text{всегда существует.}$$

2) $f_\xi(0) = 1, \quad \left|f_\xi(t)\right| \le 1$

Действительно:

$$Me^{it\xi}\Big|_{t=0} = Me^0 = 1.$$
$$\left|Me^{it\xi}\right| \le M\left|e^{it\xi}\right| = 1.$$

3) $\xi_1, \xi_2, \dots, \xi_n$ – независимые в совокупности случайные величины.

$$\alpha = \sum_{j=1}^{n} \xi_j$$

Тогда

$$f_{\xi_1+\xi_2+\dots+\xi_n}(t) = Me^{it(\xi_1+\xi_2+\dots+\xi_n)} = Me^{it\xi_1}e^{it\xi_2}\dots e^{it\xi_n} = \prod_{k=1}^{n} Me^{it\xi_k} = \prod_{k=1}^{n} f_{\xi_k}(t).$$

4)

$$\xi \to f_\xi(t)$$
$$\sigma\xi + \mu \to f_{\sigma\xi+\mu}(t) = e^{it\mu} f_\xi(\sigma t)$$

Зачем нужны характеристические функции? Для доказательства теорем о сходимости случайных величин (например, ЦПТ, которая будет в 18-м вопросе). Там в доказательства используется теорема о непрерывности, которую мы сейчас сформулируем:

Теорема о непрерывности характеристических функций

Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ - последовательность случайных величин, ξ_0- случайная величина. Каждой случайной величине соответствуют функция распределения и характеристическая функция:

$$F_1(x), F_2(x), \dots, F_n(x);\ \ F_0(x),$$
$$f_1(x), f_2(x), \dots, f_n(x);\ \ f_0(x).$$

Пусть $f_n(t) \xrightarrow[n\to\infty]{} f_0(t) \quad \forall t \in (-\infty, \infty), \quad f_0(t)$ – непрерывна в точке $t = 0$. Это условие необходимо и достаточно для того, чтобы $F_n(t) \xrightarrow[n\to\infty]{} F_0(t)$ в каждой точке непрерывности $F_0(x)$.

Если $F_0(x)$- непрерывна на $(-\infty, \infty)$, то сходимость $F_n(t) \xrightarrow[n\to\infty]{} F_0(t)$ равномерна.

Т.е. если предел последовательности интегральных функций распределения непрерывен, то предел последовательности характеристических функций распределения непрерывен.

Вопросы 7,8 и 18

7. Вероятностное пространство биномиальной схемы независимых испытаний. Биномиальное распределение. Отрицательное биномиальное распределение.

8. Биномиальное распределение. Теорема Пуассона. Интегральная теорема Муавра–Лапласа.

18. Центральная предельная теорема. Интегральная теорема Муавра–Лапласа.

Вероятностное пространство бином. Схемы независимых испытаний

Вероятностное пространство исходов одного испытания – это набор из пространства исходов Ω, алгебры F и функции вероятности P: (Ω, F, P). Что будет, если испытаний несколько?

Пространство исходов n испытаний – это Ω^n.
Например, если Ω – отрезок, то для двух испытаний будет квадрат, а для трёх – куб.
Первая координата (абсцисса) будет означать исход первого испытания, вторая (ордината) – исход второго и т.д.

Также модифицируется и алгебра F – она теперь называется F^n и включает «хорошие» множества не в Ω, а в Ω^n.
А вот с вероятностью ничего не происходит, так что конечная вероятностная тройка (Ω^n, F^n, P).

Биномиальное распределение

Конечно, это вы уже знаете. В каждом из n испытаний может быть или «быть» с вероятностью p и «не быть» с вероятностью 1-p.
Тогда в серии из n испытаний общее число «быть» может изменяться от 0 до n, причём

$$p(k \text{ успешных исходов в серии из } n \text{ испытаний}) = C_n^k p^k (1-p)^{n-k}$$

$$p(k \text{ успешных исходов в серии из } n \text{ испытаний}) = C_n^k p^k (1-p)^{n-k}$$

Формула, которую вы многократно использовали.

Теорема Пуассона

При стремлении n к бесконечности, а p к нулю (так, чтобы матожидание np=λ оставалось конечным), то биномиальное распределение стремится к пуассоновскому:

$$C_n^k p^n q^{n-k} \to \frac{\lambda^k}{k!} exp(-\lambda)$$

Доказательство:

$$C_n^k p^n q^{n-k} = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} = \frac{1}{k!} n(n-1) \cdot\ldots\cdot (n-k+1) p^k (1-p)^{n-k}$$
$$= \frac{1}{k!} 1 \left(1 - \frac{1}{n}\right) \cdot\ldots\cdot \left(1 - \frac{k+1}{n}\right) (np)^k (1-p)^{n-k}.$$
$$1 \left(1 - \frac{1}{n}\right) \cdot\ldots\cdot \left(1 - \frac{k+1}{n}\right) \to 1, n \to \infty, k = fix.$$
$$(np)^k \to \lambda^k, np \to \lambda, k = fix.$$
$$(1-p)^{-k} \to 1, p \to 0, k = fix.$$
$$ln(1-p)^n = n \cdot ln(1-p) = n \cdot \big(-p + o(p)\big) \to -\lambda \Rightarrow (1-p)^n \to e^{-\lambda}.$$

Интегральная теорема Муавра-Лапласа

Если точно так же устремить n к бесконечности, но при этом

$$n \to \infty, \left|\frac{k-np}{\sqrt{npq}}\right| \leqslant C, p = fix$$

то дискретное распределение станет напоминать непрерывное – а именно, нормальное.

Вероятность того, что будет не более k успехов, будет стремиться к

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{k-np}{\sqrt{npq}}} e^{z^2/2}\, dz$$

.

Интегральная теорема Муавра-Лапласа является следствием более общей – центральной предельной теоремы.

Пусть $\xi_1, \xi_2, \dots, \xi_n, \ \dots$ - последовательность независимых в совокупности и одинаково распределенных случайных величин. $M_{\xi_i} = \mu, D_{\xi_i} = \sigma^2$. Тогда

$$\frac{1}{\sqrt{n}\,\sigma}\sum_{j=1}^{n}(\xi_j - \mu) \xrightarrow{d} \eta \sim N(0,1),$$

ЦПТ утверждает, что случайная величина - сумма любых центрированных (т.е. с матожиданием 0) случайных величин - есть нормальное распределение.

Вопросы 19 и 20

19. Цепи Маркова. Свойства матриц перехода.

20. Финальные распределения в цепи Маркова. Теорема Маркова. Эргодичность цепи Маркова.

Что такое цепи Маркова?

Пусть у нас есть набор несовместных событий $\omega_1, \omega_2, \dots, \omega_N$. И у нас есть некий «игровой автомат», способный нам выдавать каждый раз одно из этих событий. Один раз дёрнули за ручку – нам выдали одно из событий $\omega_1, \omega_2, \dots, \omega_N$, другой раз – снова выдали какое-то из событий $\omega_1, \omega_2, \dots, \omega_N$, и так k раз. Каждый раз мы записали результат, и в итоге у нас получилась строчка:

$\{\omega^1, \omega^2, \dots, \ \omega^k\}$ (верхними индексами я буду обозначать хронологический номер события, а нижними – номер события в исходном списке $\omega_1, \omega_2, \dots, \omega_N$).

Влияют ли предыдущие события за вероятность последующего? Если да, то это псевдомаркова цепь.

Нужен пример, да? Вспомним переписку с мобильного телефона. Вот мы печатаем, а мобила предлагает нам несколько следующих слов на выбор. Как она делает? А это тоже псевдомарковы цепи. В русском языке конечное число словоформ (словоформа – слово с учётом окончаний), пусть это число N. Каждое слово – это событие из списка-словаря $\omega_1, \omega_2, \ldots, \omega_N$. Наше сообщение – это строчка, набор слов. Влияют ли предыдущие слова на то, какое будет следующим? Ещё как! После слов «Больше всего на ФФ мне не нравятся» может последовать «праки», «практикумы», «экзамены», «первые пары», но никак не «бульвар», «небо», «ещё» и «причём». Вот вам и пример псевдомарковой цепи. (Кстати, это и обуславливает интерес IT-шников к марковским цепям).

А маркова цепь – это псевдомаркова, обладающая двумя свойствами:
1) Вероятность последующего события зависит только от последнего предыдущего.
2) При известном настоящем прошлое и будущее независимы (но это обычно неважно, важно именно первое свойство).

Мобильная переписка уже не будет марковой цепью, т.к. там вероятность последующего события зависит от всех предыдущих, а не только от последнего. Так что пример уже марковой цепи будет такой – **случайное блуждание на отрезке:**
У нас есть отрезок. Левая точка 1, правая N. Мы рассматриваем целочисленные точки на нём.
Мы на каждом шаге с вероятностью p идём на 1 влево и с вероятностью q на 1 вправо.
Если мы попадаем в любую из крайних точек, мы там залипаем и более из неё не выбираемся.
После каждого шага записываем координату. В итоге имеем нечто вроде
5, 4, 5, 4, 3, 2, 3, 2, 1, 1, 1, 1, 1…
Очевидно, что тут вероятность оказаться в какой-либо точке будет зависеть уже только от последнего места, и не зависеть от того, что было до этого.

Матрица перехода
Для марковских цепей вводят так называемую матрицу перехода
$\hat{\pi} = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$ (галочку обычно не пишет, это я пишу, т.к. считаю, что всё некоммутирующее (в т.ч. матрицы) должно писаться с галочкой).
 Где p_{ij} – вероятность перехода за 1 шаг из i-того состояния в j-тое (т.е. $p(\omega_j|\omega_i)$, условная вероятность перейти в j-тое состояние при условии, что мы сейчас в i-том).

Например, для случайного блуждания, если мы сейчас в i-той точке и она промежуточная, вероятность попасть в i-1-ю p, в i+1-ю q, в остальные 0. Для крайних точек вероятность остаться в них 1, вероятность попасть в остальные 0. Поэтому матрица перехода имеет вид (показан случай N=6):

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 & 0 \\ 0 & p & 0 & q & 0 & 0 \\ 0 & 0 & p & 0 & q & 0 \\ 0 & 0 & 0 & p & 0 & q \\ 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$

Данная матрица обладает двумя свойствами:

1) Все элементы >0 (т.к. это вероятности)

2) Сумма элементов в каждой строчке =1 (т.к. сумма вероятностей=1).

Подобные матрицы называются стохастическими.

Ещё пример: случай, когда у системы «отсутствует память», каждый раз она как новенькая: вероятности новой попытки совсем не зависят от того, что было последний раз.

Тогда матрица будет иметь вид:

$$\begin{pmatrix} p_1 & p_2 & \cdots & p_N \\ p_1 & p_2 & \cdots & p_N \\ \vdots & \vdots & \ddots & \vdots \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}$$

А что, если нам надо найти вероятность перехода из i-того в j-тое состояние не за 1 шаг, а за ν шагов? Оказывается, надо просто возвести матрицу перехода в ν-тую степень: $\hat{\pi}^{\nu}$, и ij-тый элемент уже полученной матрицы будет означать вероятность перехода из i-того состояния в j-тое за ν шагов.

Доказательство этой теоремы проще произвести по мат.индукции. База при ν=1 есть, делаем переход.

Вероятность перехода из i-того состояния в j-тое за ν шагов равна сумме по k вероятности перехода из i-того состояния в k-тое за ν-1 шагов * вероятность перехода из k-того в j-тое за 1 шаг.

Но вероятность перехода из i-того состояния в k-тое за ν-1 шагов – это $\hat{\pi}^{\nu-1}{}_{ik}$

Вероятность перехода из k-того в j-тое за 1 шаг – это $\hat{\pi}_{kj}$

В итоге имеем $\sum_k \hat{\pi}^{\nu-1}{}_{ik}\,\hat{\pi}_{kj}$, но это же $\hat{\pi}^{\nu}{}_{ij}$, ч.т.д.

Эргодичность.
Скорее всего, вы сталкивались с этим понятием в курсе молекулы. Там оно формулировалось как «среднее по ансамблю равно среднему по времени». Математики же придумали своё определение, более строгое и понятное:

Цепь Маркова называется эргодичной, если

$$\lim_{\nu \to \infty} \hat{\pi}^{\nu} = \begin{pmatrix} p_1 & p_2 & \dots & p_n \\ p_1 & p_2 & \dots & p_n \\ & & \dots & \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Т.е. через ν шагов вероятность придти в j-тое состояние одна и та же для всех i. Вот эти вот вероятности p_1, p_2, … , p_N – вероятности в j-тое состояние ***в конце концов*** – называются финальными вероятностями.

Представим себе замкнутую комнату, внутри которой молекулы болтаются. Возьмём промежуток 1 микросекунду и кубик стороной 1 мм внутри этой комнаты. Про каждую молекулу, находящуюся в этом кубике, спрашиваем, где она была 1мкс назад. Скорее всего, в окрестностях этого кубика, потому что из более дальних мест она долететь за эту 1 мкс она не успела.
А теперь возьмём промежуток 1 год. Где ты, молекула, была год назад? Да где угодно, за этот год всё там так перемешалось, что вероятности, что она была год назад, одинаковы для любого места комнаты.

Об этом и эргодичность.

Но это я физику приплёл, а нам могут дать просто матрицу и сказать: «Слышь, чел, во тебе матрица перехода, соответствующая некоторой цепи Маркова. Можешь проверить, эргодична ли цепь или нет?»
И вы такие… ну эээ… а как? Определение неудобно, потому что в нём предел.

Есть, к счастью, признак эргодичности - теорема Маркова:
Если существует ν такое, что все N^2 элементов матрицы $\hat{\pi}^{\nu}$ >0, то цепь Маркова эргодична.

Как вы понимаете, это очень легко запрогать: for (ν=1; ν<дофига; ++ν) вы поочерёдно проверяете N^2 элементов, если все >0 – ура, эргодичненько.
Но это именно признак эргодичности, а не свойство – в обратную сторону не работает.