В прошлый раз мы говорили, что дисперсия некоторым образом характеризует разброс значений случайной величины. Покажем, какой точный смысл можно придать этому утверждению.

Теорема 1. (Неравенство Чебышёва) Пусть ξ — случайная величина. Тогда для любого $k > 1$ выполнено $Pr(|\xi - M\xi| \geqslant k\sqrt{D\xi}) \leqslant \frac{1}{k^2}$.

Доказательство 1.

Лемма 1. Пусть случайная величина η принимает только неотрицательные значения. Тогда для любого $k > 1$ выполнено $Pr(\eta \geqslant kM\eta) \leqslant \frac{1}{k}$.

Доказательство 2. Рассмотрим сумму значений, умноженных на вероятность принятия этих значений, равную ожиданию η. Рассмотрим только те слагаемые, которые соответствуют значениям не меньше $kM\eta$. Если суммарная вероятность, соответствующая этим значениям, превышает $\frac{1}{k}$, то уже эта часть суммы превысит $M\eta$, то есть значение всей суммы. Но остальные слагаемые неотрицательны. Противоречие доказывает, что суммарная вероятность, соответствующая значениям η не меньше $kM\eta$, не больше $\frac{1}{k}$, что и требовалось.

Теперь заметим, что $|\xi - M\xi| \geqslant k\sqrt{D\xi} \Leftrightarrow (\xi - M\xi)^2 \geqslant k^2 D\xi$ и $D\xi = M(\xi - M\xi)^2$. Таким образом, положив $\eta = (\xi - M\xi)^2$ в лемме, получим требуемое.

Доказательство 3. Другое доказательство.

Лемма 2. Пусть для всех исходов $\xi \geqslant \eta$. Тогда $M\xi \geqslant M\eta$.

Доказательство 4. Напишем ожидания как суммы по элементарным событиям произведений вероятности и значения. Тогда в каждом слагаемом первой суммы первый сомножитель неотрицательный и такой же, как во второй, а второй не меньше.

Рассмотрим для заданных ξ и k из условия теоремы величину η, равную 0 при $|\xi - M\xi| < k\sqrt{D\xi}$ и 1 иначе. Очевидно, что $M\eta = Pr(|\xi - M\xi| \geqslant k\sqrt{D\xi})$. Рассмотрим также величину $\rho = \frac{1}{k^2 D\xi}(\xi - M\xi)^2$. Заметим, что $\rho \geqslant \eta$. Тогда $Pr(|\xi - M\xi| \geqslant k\sqrt{D\xi})$ равно $M\eta \leqslant M\rho = \frac{1}{k^2 D\xi} M(\xi - M\xi)^2 = \frac{1}{k^2}$, что и требовалось.

Теорема 2. (Закон больших чисел) Пусть мы повторяем эксперимент (отдельные повторы независимы) и считаем долю испытаний, в которой произошло данное событие A, имеющее вероятность p. Тогда вероятность того, что при n повторениях отклонение доли от вероятности превысит заданное $\varepsilon > 0$ стремится к нулю с ростом n.

Замечание 1. Мы рассматриваем вероятности, связанные с конечным количеством повторений эксперимента, а потом переходим к пределу в ответе. Поэтому мы по-прежнему можем ограничиваться рассмотрением конечных наборов элементарных событий.

Доказательство 5. Рассмотрим случайные величины ν_k, каждая из которых равна 1, если в k-м испытании произошло событие A и ν, равную сумме все ν_k с 1-го по n-е, делённой на n. Каждая из ν_k имеет математическое ожидание p и дисперсию $p(1 - p) \leqslant$. Ожидание ν равно p. Дисперсия суммы ν_k в силу независимости величин равна $np(1 - p)$, а дисперсия ν — по формуле выноса константного множителя из дисперсии — равна $\frac{p(1-p)}{n}$. Теперь можно применить к ν неравенство Чебышёва. Так как дисперсия стремится к нулю, то вероятность фиксированного отклонения тоже стремится к нулю.

Замечание 2. Здесь могла бы возникнуть разница между попарной независимостью событий из набора и независимостью в совокупности. Независимость в совокупности означает, что вероятность пересечения равна произведению вероятностей для произвольного подмножества рассматриваемого набора событий. Здесь это неважно, потому что дисперсия выражается как нечто второй степени и при раскрытии скобок в вычислении дисперсии суммы не возникнут более чем попарные произведения.

Примером независимых попарно, но не в совокупности, величин могут служить результаты двух бросаний монетки и их сумма по модулю два: любая пара независима, но по любым двум однозначно восстанавливается третья величина.

Определение 1. Марковская цепь задаётся количеством состояний n, выбором начального состояния и набором переходных вероятностей P_{ij}, $\sum\limits_{k=1}^{n} P_{ik} = 1$. Эксперимент происходит следующим образом: сначала частица ставится в начальное состояние. Потом на каждом ходу выбирается следующее состояние, переход из состояния i в состояние j происходит с вероятностью P_{ij}. Случайный выбор при переходах выполняется независимо.

Одним из вариантов описания независимости выбора в соответствии с вероятностями, зависящими от результата предыдущих испытаний, следующий: на каждом шаге производится n случайных испытаний с вероятностями, заданными n строками таблицы, и все такие испытания на всех шагах независимы. После этого выбирается результат, полученный при выборе по нужной строке.

Посмотрим каким может быть поведение марковской цепи. Для простоты будем считать, что события с нулевой вероятностью в конечном по времени эксперименте невозможны.

Во-первых, возможно, что некоторые состояния недостижимы из начального. Такие можно просто игнорировать.

Во-вторых, возможно, что попав в некоторые состояния мы уже никогда не сможем попасть в некоторые другие. Рассмотрим для каждого состояния множество тех состояний, в которые из него можно попасть. Со временем (при блуждании частицы) это множество может только уменьшаться, но никогда не станет пустым. Рассмотрим все состояния, после которых это множество не может уменьшиться. Они разбиваются на компоненты связности — множества попарно достижимых вершин. Нетрудно видеть, что вероятность попасть внутрь одной из таких компонент с ростом числа шагов стремится к 1, так как из каждого состояния достижима хотя бы одна компонента; поэтому за n шагов вероятность не попасть ни в одну финальную компоненту связности умножается на число меньше 1. Будем в дальнейшем считать, что у нас исходно из любой вершины можно попасть в любую.

В-третьих, возможно, что в некоторые состояния можно попасть только за число шагов, делящееся за некоторое k. Например, на шахматной доске, переходя через стороны клеток, мы каждый ход обязаны менять цвет клетки на которой стоим. При этом за сколь угодно большое фиксированное число ходов мы можем попасть не во все клетки (если требовать попасть в них последним ходом).

Теорема 3. Пусть имеется марковская цепь, в которой можно из любого состояния перейти в любое ровно за k шагов. Тогда вероятности перехода из состояния i в состояние j ровно за N шагов имеют пределы при $N \to \infty$ и эти пределы не зависят от i.

Доказательство 6. Обозначим $P_{ij}^{(N)}$ вероятность перейти в состояние j ровно за N шагов, если мы начинаем из состояния i.

Пусть за k шагов минимальная вероятность перехода равна ε (а максимальная, соответственно, не больше $1-\varepsilon$). Пусть за некоторое N шагов минимальная (по исходным вершинам) вероятность перехода в состояние j равна $P_{min,j}^{(N)}$, а максимальная — $P_{max,j}^{(N)}$. Рассмотрим $k+N$ шагов.

Эти шаги мы разобьём на первые k и последние N. По формуле полной вероятности $P_{ij}^{(k+N)} = \sum\limits_l P_{il}^{(k)} P_{lj}^{(N)}$, так как мы должны за k шагов куда-то прийти. Оценим $P_{ij}^{(k+N)}$ сверху. $P_{ij}^{(k+N)} \leqslant \varepsilon P_{min,j}^{(N)} + (1-\varepsilon) P_{max,j}^{(N)}$, так как с вероятностью не менее ε (минимальная вероятность перехода за k шагов) мы придём за k шагов в вершину, из которой минимальна вероятность попасть в j за следующие N шагов, а в любом другом случае вероятность попасть в j за N шагов всё равно не превысит $P_{max,j}^{(N)}$. Если вероятность попасть в самую неудобную вершину больше ε, то результирующая вероятность ещё меньше. Аналогично, $P_{ij}^{(k+N)} \geqslant (1-\varepsilon) P_{min,j}^{(N)} + \varepsilon P_{max,j}^{(N)}$. Эти оценки, в частности, справедливы для $P_{min,j}^{(k+N)}$ и $P_{max,j}^{(k+N)}$. Поэтому если мы рассмотрим минимальную и максимальную вероятности перехода в состояние j, то с ростом N шагами по k они будут стремиться к общему пределу.

С другой стороны, $P_{ij}^{(1+N)} = \sum\limits_l P_{il} P_{lj}^{(N)}$ лежит в пределах $[P_{min,j}^{(N)}; P_{max,j}^{(N)}]$. Поэтому между членами нашей подпоследовательности отношение вероятностей перехода не портится, что и требовалось.

Введём теперь понятие условного математического ожидания.

Рассмотрим произвольное множество исходов, Ω, вероятностное пространство на нём и событие в нём B, имеющее положительную вероятность. Тогда можно считать, что B само является вероятностным пространством, так как условные вероятности при условии B обладают всеми свойствами вероятностей. Если у нас есть случайная величина ξ, то есть функция на исходах, то можно рассмотреть её ограничение на множество исходов B как случайную величину. Её математическое ожидание назовём математическим ожиданием ξ при условии B.

Далее, пусть у нас есть множество событий, имеющих положительную вероятность, $\{B_1, \ldots, B_n, \ldots\}$. Тогда можно для каждого события из множества найти математическое ожидание ξ при условии этого события. Назовём *условным математическим ожиданием* величины ξ при условии множества событий $\{B_1, \ldots, B_n, \ldots\}$ отображение, переводящее B_k в математическое ожидание ξ при условии B_k.

Наибольший интерес для нас сейчас представляет случай, когда $B_1, \ldots, B_n, \ldots$ попарно не пересекаются. В этом случае условное математическое ожидание является функцией не только на событиях, но и на самих исходах (никакому исходу не поставлено в соответствие два разных значения), определённой на каком-то событии, то есть случайной величиной.

Определим также для двух случайных величин ξ и η условное математическое ожидание ξ при условии η. Рассмотрим B — множество событий вида "η принимает значение x". По определению, $M(\xi|\eta) = M(\xi|B)$.

Как пример можно привести условное математическое ожидание среднего трёх оценок при условии первой для школьника, который равновероятно и независимо каждый раз получает 3, 4, или 5. Элементарных событий здесь будет 27. Так как независимо от первой оценки математическое ожидание суммы двух оставшихся равно 8, то это случайная величина равная $3\frac{2}{3}$ на девяти элементарных событиях, где первая оценка 3, 4 на девяти элементарных событиях, где первая оценка 4, и $4\frac{1}{3}$ на девяти событиях, где первая оценка 5.

Если у нас есть разбиение множества исходов на события положительной вероятности $B_1, \ldots, B_n$, а событие B является объединением каких-то из этих событий, то $M(\xi|B) = M(M(\xi|\{B_1, \ldots, B_n\})|B)$. Чтобы доказать это, надо рассмотреть математическое ожидание как сумму по элементарным событиям в левой части и как сумму по $B_1, \ldots, B_n$ в правой части. После раскрытия скобок в правой части заметим, что $P(B_j)P(\sigma|B_j) = P(\sigma)$.

Приведём пример задачи, в которой ответ удобно определять через УМО. Пусть есть случайная величина ξ, которую мы хотим оценить по результатам эксперимента, и случайная величина η, которую нам сообщают. При этом будем считать, что мы сообщаем как оценку ξ величину $f(\eta)$ и выбираем функцию f для минимизации $M(f(\eta)-\xi)^2$.

Найдём, какое y нам выгодно сообщить при условии $\eta = x$. Вычислим

$$M((y-\xi)^2|\eta = x) = M(y^2 - 2y\xi + \xi^2|\eta = x) = y^2 - 2yM(\xi|\eta = x) + M(\xi^2|\eta = x) =$$
$$= (y - M(\xi|\eta = x))^2 + M(\xi^2|\eta = x) - M^2(\xi|\eta = x) = (y - M(\xi|\eta = x))^2 + D(\xi|\eta = x)$$

Ясно, что минимум достигается при $y = M(\xi|\eta = x)$. Выбирая таким образом значения f, получаем, что при наилучшей f будет выполнено $f(\eta) = M(\xi|\eta)$.