Высшая школа экономики, 2010-11 учебный год
Факультет прикладной политологии
Высшая математика (вводный курс)
И.А.Хованская, И.В.Щуров, К.И.Сонин (РЭШ)

Лекция 6. Дискретные случайные величины

Распределение случайной величины

Какая температура завтра будет на улице в 10 утра? Какую оценку данный студент получит на экзамене по курсу математики? Сколько будут стоить акции некоторой компании в следующем месяце? Сколько очков выпадет на игральном кубике?

Всё перечисленное — примеры *случайных величин*. Неформально говоря, случайная величина — это такая величина, значение которой выбирается в результате проведения случайного испытание. Поскольку исход случайного испытания заранее неизвестен, вообще говоря, неизвестно, чему будет равна случайная величина. Тем не менее, что-то о случайных величинах мы обычно всё-таки знаем. Например, если речь идет о завтрашней температуре, можно с уверенностью сказать, что она будет не меньше -300 градусов по Цельсию (из физических соображений, температура в принципе не может быть меньше абсолютного нуля, то есть -273.15 градусов Цельсия), а оценка на экзамене — будет не меньше нуля или не больше 10 баллов (при оценивании по десятибальной шкале). Можно получить еще больше информации о случайной величине: например, если известно, что студент успешно делал все домашние задания, разумно считать, что вероятность получения оценки больше 5 будет больше, чем вероятность получения оценки ниже 4. И так далее. Рассмотрим несколько примеров.

Пример 1. *Мы подкидываем монетку три раза. Случайная величина (обозначим её за x) — число выпавших орлов.*

Нетрудно видеть, что такая величина может принимать одно из конечного множества значений — $\{0,\ 1,\ 2,\ 3\}$. (Этим она отличается, например, от температуры за окном.) Мы будем рассматривать только такие случайные величины.

Напомним, как устроены элементарные исходы в ситуации с подбрасыванием монетки: мы можем их кодировать последовательностями символов «О» и «Р» — например, «ООО» — все три раза выпали орлы, или «ОРО» — в первый раз выпал орёл, во второй — решка, в третий — снова орёл. (Всего элементарных исходов в этом случае $2^3 = 8$.) Теперь каждому элементарному исходу можно сопоставить значение случайной величины x, которое она принимает, если реализовался этот исход. Например, если реализовался исход «ООО», то выпало 3 орла, и значит $x = 3$. А если реализовался исход «ООР», орёл выпал всего один раз, то есть $x = 1$. (См. табл. 1.)

По таблице нетрудно видеть, что различные значения могут приниматься случайной величиной с разными вероятностями. Например, если $x = 0$, то орёл не выпадал ни разу, а значит выпали все решки. Этому событию соответствует единственный исход «РРР», и его вероятность таким образом равна 1/8. С такой же вероятностью x может принимать и значение 3. А вот для $x = 1$ возможностей гораздо больше: такому значению соответсвует три исхода, в которых выпадает ровно один орёл: «ОРР», «РРО», «РОР». Значит, вероятность события $x = 1$ равна 3/8. (См. таблицу 2.)

исход	значение x
ООО	3
ООР	2
ОРО	2
ОРР	1
РОО	2
РОР	1
РРО	1
РРР	0

Таблица 1: Значения случайной величины x (число орлов) при реализации различных исходов

значение x	исходы	вероятность
0	РРР	1/8
1	РРО, РОР, ОРР	3/8
2	ООР, ОРО, РОО	3/8
3	ООО	1/8

Таблица 2: Распределение значений случайной величины x

Таблица, в которой указано, с какой вероятностью случайная величина принимает то или иное значение, называется *распределением (дискретной) случайной величины.* Фактически, случайная величина задана, если задано её распределение.

Перейдем к следующему примеру.

Пример 2. *Игральный кубик кидают один раз. Случайная величина y — выпавшее число очков.*

Каково распределение такой случайной величины? Она y принимает одно из 6 значений $\{1, 2, 3, 4, 5, 6\}$. Каждое значение принимается с вероятностью $1/6$. Таблица распределения в этом случае очень простая (см. табл. 3).

значение y	1	2	3	4	5	6
вероятность	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Таблица 3: Распределение случайной величины y

Характеристики случайных величин

Допустим, случайная величина принимает два значения — 0 и 5000. Какое у неё будет *среднее значение*? На первый взгляд кажется разумным взять среднее арифметическое:

$$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 5000 = 2500. \tag{1}$$

Однако, рассмотрим такой пример:

Пример 3. *Вероятность выиграть в мгновенную лотерею равна* $1/100$*. Сумма выигрыша — 5000 руб. Если рассмотреть выигрыш как случайную величину, каково её среднее?*

Распределение такой случайной величины (мы будем обозначать её через z) приведено в табл. 4. Сейчас кажется очевидным, что «средний выигрыш» в такой игре не бу-

значение z	0	5000
вероятность	$1/100$	$99/100$

Таблица 4: Распределение выигрыша в лотерею y

дет равен 2500 руб, а будет существенно меньше. Чему же он тогда равен? Допустим, мы играем в лотерею $100'000$ раз. Из них мы примерно $100'000 \cdot \frac{99}{100} = 99'000$ раз проиграем, и примерно $100'000 \cdot \frac{1}{100} = 1000$ раз выиграем. За каждый выигрыш мы получим 5000 руб., а за проигрыш не получим ничего. То есть общий выигрыш составит примерно $99'000 \cdot 0 + 1000 \cdot 5000 = 5'000'000$ руб. В среднем на одну игру выигрыш составит

$$\frac{99'000 \cdot 0 + 1000 \cdot 5000}{100'000} = \frac{99}{100} \cdot 0 + \frac{1}{100} \cdot 5000 = 50 \text{ руб.} \tag{2}$$

Это и есть «настоящее» среднее описанной случайной величины. (Или, что то же самое — «честная» стоимость лотерейного билетика. Если билетик стоит дороже — излишек идет в карман организатору лотереи, если дешевле — организатор быстро разорится.) Нетрудно видеть, что в отличие от среднего арфиметического, каждое слагаемое здесь берется с множителем, равным вероятности его реализации. (Ср. формулы (1) и (2).) Аналогичным образом — просуммировав значения, которые принимает случайная величина, умноженные на соответствующие вероятности — можно найти среднее (оно называется *математическим ожиданием*) любой дискретной случайной величины.

Вернемся к примеру 1. Каково «среднее» число орлов, выпадающих в трёх бросаниях? Поскольку орлы и решки равновероятны, представляется логичным, что среднее число орлов равно среднему числу решек. С другой стороны, в каждом испытании сумма числа выпавших орлов и числа выпаших решек равна 3 (количеству подкидываний монетки). Значит, в «среднем» должно выпадать 1,5 орла и 1,5 решки. Давайте проверим, так ли это. Используя таблицу 2, вычислим математическое ожидание случайной величины x (оно обозначается Ex, от *Expected value*):

$$Ex = \frac{1}{8} \cdot 0 + \frac{3}{8} \cdot 1 + \frac{3}{8} \cdot 2 + \frac{1}{8} \cdot 3 = \frac{12}{8} = \frac{3}{2} = 1{,}5 \tag{3}$$

Интуиция нас не обманула: среднее действительно равно 1,5.

Математическое ожидание — важная характеристика случайной величины. Но зачастую знания только мат.ожидания недостаточно, чтобы делать какие-либо выводы. Давайте представим себе «честное казино». Вы можете в нём поставить произвольную сумму, например, на «красное» (или на «нечет»), и с вероятностью $1/2$ выиграть и удвоить свою ставку (то есть выиграть столько же, сколько поставили), а с вероятностью $1/2$ — проиграть и потерять её. Допустим, один человек ставит 10 рублей, а другой — миллиард рублей. Можно рассмотреть две случайные величины — выигрыш первого игрока (x_1) и выигрыш второго игрока (x_2). (Если игрок проигрывает свою ставку, то его выигрыш отрицателен.) Распределения этих случайных величин приведены в таблице 5. Нетрудно видеть, что в

вероятность	1/2	1/2
значение x_1	-10	10
значение x_2	-1000'000'000	1000'000'000

Таблица 5: Распределение выигрыша в «честном казино»

обоих случаях математическое ожидание равно нулю (поэтому мы и назвали такое казино «честным» — на практике, конечно, так не бывает, и математическое ожидание выигрыша в казино всегда отрицательно). Тем не менее, очевидно, что второй игрок рискует гораздо большими суммами (хотя может и выиграть гораздо больше), чем первый. На практике, это явно различные ситуации — вы, вероятно, были бы готовы рискнуть десятью рублями (имея некоторую толику азартности), но вряд ли согласитесь поставить на кон большую сумму, даже если знаете, что «в среднем» в обоих случаях ничего не потеряете и не приобретете.

Случайные величины x_1 и x_2 отличаются тем, насколько далеко они отклоняются от своего среднего значения (в данном случае, нуля). Если мы попытаемся просто посчитать «среднее отклонение от среднего значения», то есть $E(x - Ex)$, то получим всегда нуль (потому что отклонение в положительную сторону и в отрицательную берутся с разными знаками и компенсируют друг друга при сложении). Чтобы избежать этой проблемы рассматривают *средний квадрат отклонения от среднего значения*, то есть $E((x - Ex)^2)$. Эта величина и называется *дисперсией* случайной величины.

Пример 4. *Вычислим дисперсию средней величины y — «число очков, выпавших на кубике (см. пример 2).*

Математическое ожидание Ey равно 3,5 (проверьте!). Чтобы вычислить дисперсию, вместе с величиной y нужно рассмотреть еще две случайные величины: $y - Ey = y - 3.5$ (например, если y принимает значение 1, случайная величина $y - 3.5$ принимает значение $1 - 3.5 = -0.5$) и $(y - Ey)^2$ (если $y - Ey$ принимает значение 1.5, то $(y - Ey)^2$ принимает значение $1.5^2 = 2.25$). Математическое ожидание последней и будет дисперсией y. Рассмотрим соответствующие таблицы распределений (см. табл. 6).

Таким образом,

$$Dy = E((y - Ey)^2) = \frac{1}{6} \cdot 6.25 + \frac{1}{6} \cdot 2.25 + \frac{1}{6} \cdot 0.5 + \frac{1}{6} \cdot + \frac{1}{6} \cdot 2.25 + \frac{1}{6} \cdot 6.25 = 15/6 \qquad (4)$$

Это и есть дисперсия величины y.

вероятность	1/6	1/6	1/6	1/6	1/6	1/6
значение y	1	2	3	4	5	6
значение $y - Ey$	-2.5	-1.5	-0.5	0.5	1.5	2.5
значение $(y - Ey)^2$	6.25	2.25	0.25	0.25	2.25	6.25

Таблица 6: Вычисление дисперсии y