Тема 5. Дискретные случайные величины

Дискретные случайные величины и их распределения

Случайными величинами называются функции от элементарных событий ω. Обычно их обозначают большими латинскими буквами, чтобы не путать с переменными и параметрами: $X(\omega), Y(\omega), Z(\omega), \ldots$. В классической вероятностной модели с равновероятными исходами любая случайная величина принимает лишь конечное число значений ввиду конечности пространства Ω. В общей вероятностной модели случайные величины могут иметь бесконечно много значений. Если множество значений является конечным или счётным, то случайная величина называется *дискретной* (случайные величины с несчётным множеством значений будут рассмотрены в теме 9.)

Простейшей случайной величиной является индикатор I_{A} некоторого события А:

$$I_{\mathrm{A}}(\omega) = \begin{cases} 1, & \text{если } \omega \in \mathrm{A}, \\ 0, & \text{если } \omega \notin \mathrm{A}. \end{cases}$$

Пример 1. Напомним условия задачи 2.1: среди m экзаменационных билетов l — лёгкие, n студентов по очереди наудачу берут билеты. Пусть билеты выбираются с возвращением. Рассмотрим события

$$\mathrm{A}_k = \{k\text{-й студент выбрал лёгкий билет}\}, \quad k = 1, 2, \ldots, n.$$

Обозначим через I_k индикатор события A_k. Согласно решению задачи 2.1 имеем

$$\mathbf{P}(I_k = 1) = \mathbf{P}(\mathrm{A}_k) = l \cdot m^{n-1} / m^n = l / m = p,$$

где $p = l/m$ — доля лёгких билетов. Положим $S_n = I_1 + \ldots + I_n$. Тогда S_n — общее число лёгких билетов, вынутых всеми n студентами. Найдем для $0 \le i \le n$ вероятность события $\{S_n = i\}$. Выбрать подмножество размера i из множества, состоящего из n студентов, можно C_n^i способами. Для каждого из i студентов этого подмножества есть l вариантов выбора номера билета. Для каждого из $(n-i)$ студентов, не вошедших в это подмножество, есть $(m-l)$ вариантов выбора номера билета. Перемножая все возможности, находим

$$\mathbf{P}(S_n = i) = \frac{C_n^i l^i (m-l)^{n-i}}{m^n} = C_n^i p^i (1-p)^{n-i}. \quad (1)$$

Определение. *Распределением* случайной величины $X(\omega)$ называется набор пар (x_i, p_i), где x_i — всевозможные значения случайной величины, $p_i = \mathbf{P}(X = x_i)$ — соответствующие вероятности.

Например, индикаторная случайная величина I_k из примера 1 имеет распределение

x_i	0	1
p_i	$1-p$	p

Оно называется ***распределением Бернулли***[1] с параметром p, где $0 < p < 1$.

---

[1] В честь Якоба Бернулли (1654-1705), который сформулировал и доказал в книге «Искусство предположений» частный случай важнейшей теоремы теории вероятностей — закона больших чисел.

В свою очередь, случайная величина S_n принимает значения $i = 0, 1, \ldots, n$ с соответствующими вероятностями из правой части формулы (1). Распределение этой случайной величины называется ***биномиальным***:

x_i	0	1	$\ldots$	i	$\ldots$	n
p_i	$C_n^0 p^0 (1-p)^{n-0}$	$C_n^1 p^1 (1-p)^{n-1}$	$\ldots$	$C_n^i p^i (1-p)^{n-i}$	$\ldots$	$C_n^n p^n (1-p)^{n-n}$

Его изучение сыграло важную роль в формировании теории вероятностей как отдельного направления математики.

Пример 2. В урне лежат шары, занумерованные числами от 1 до m. Наудачу с возвращением извлекают n шаров с номерами $i_1, i_2, \ldots, i_n$ соответственно. Положим $M(\omega) = \max\{ i_1, i_2, \ldots, i_n\}$. Согласно ответу на вопрос 2 из темы 2

$$p_i = \mathbf{P}(M = i) = \left(\frac{i}{m}\right)^n - \left(\frac{i-1}{m}\right)^n, \quad \text{где } i = 1, 2, \ldots, m. \tag{2}$$

Отметим, что для любого распределения вероятностей выполняется равенство $\sum p_i = 1$. Оно следует из аксиомы счётной аддитивности, поскольку события $\mathrm{A}_i = \{\omega\colon X(\omega) = x_i\}$ не пересекаются и в объединении составляют всё Ω.

Формула включений-исключений

Формула включений-исключений позволяет находить вероятность объединения нескольких событий. Её частными случаями служат следующие два равенства (в них и ниже знак пересечения множеств $\bigcap$ будем опускать ради краткости, т. е. вместо $\mathrm{A} \bigcap \mathrm{B}$ будем записывать AB):

$$\mathbf{P}(\mathrm{A}_1 \bigcup \mathrm{A}_2) = \mathbf{P}(\mathrm{A}_1) + \mathbf{P}(\mathrm{A}_2) - \mathbf{P}(\mathrm{A}_1\mathrm{A}_2),$$

$$\mathbf{P}(\mathrm{A}_1 \bigcup \mathrm{A}_2 \bigcup \mathrm{A}_3) = \mathbf{P}(\mathrm{A}_1) + \mathbf{P}(\mathrm{A}_2) + \mathbf{P}(\mathrm{A}_3) - \mathbf{P}(\mathrm{A}_1\mathrm{A}_2) - \mathbf{P}(\mathrm{A}_1\mathrm{A}_3) - \mathbf{P}(\mathrm{A}_2\mathrm{A}_3) + \mathbf{P}(\mathrm{A}_1\mathrm{A}_2\mathrm{A}_3).$$

Доказательство этих формул очевидно из рис. 1.

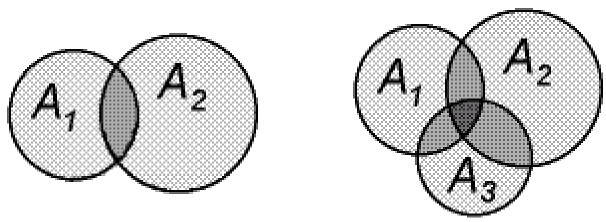

Рис. 1

В общем случае вероятность, что произойдёт <u>хотя бы одно</u> из событий $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$, вычисляется по ***формуле включений-исключений***

$$\mathrm{P}\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) = \sum_{i=1}^{n} \mathrm{P}(\mathrm{A}_i) - \sum_{i<j} \mathrm{P}(\mathrm{A}_i\mathrm{A}_j) + \sum_{i<j<k} \mathrm{P}(\mathrm{A}_i\mathrm{A}_j\mathrm{A}_k) - \ldots + (-1)^{n-1}\mathrm{P}(\mathrm{A}_1\mathrm{A}_2\ldots\mathrm{A}_n). \tag{3}$$

Использование индикаторов существенно упрощает вывод данной формулы. Применим их.

Доказательство. Для произвольных событий A и B справедливы равенства $I_{\bar{A}} = 1 - I_A$ и $I_{AB} = I_A I_B$. Воспользовавшись ими и *формулой Буля* $\overline{\bigcup A_i} = \bigcap \bar{A}_i$, запишем: $I_{\cup A_i} = 1 - (1 - I_{A_1})\ldots(1 - I_{A_n})$. Раскрывая скобки, получим соотношение

$$I_{\cup A_i} = 1 - \left(1 - \sum_i I_{A_i} + \sum_{i<j} I_{A_i A_j} - \ldots + (-1)^n I_{A_1 A_2 \ldots A_n}\right).$$

Для завершения вывода формулы (3) осталось воспользоваться равенством $\mathbf{P}(A) = \dfrac{1}{|\Omega|}\sum_{\omega\in\Omega} I_A(\omega)$.

Случайные величины со счётным множеством значений

Рассмотрим два примера дискретных случайных величин со счётным множеством значений.

Пример 3. *Геометрическое распределение*. Немного обобщим задачу о ключах из темы 4. В урне находятся m занумерованных шаров. Шары с номерами $1, 2, \ldots, l$ белого цвета, а остальные шары — чёрного. Наудачу с возвращением извлекают шары. Появление белого шара назовём «успехом», чёрного — «неудачей». Определим случайную величину L как число извлечений до первого «успеха» (включительно). Она может принимать счётное множество значений: 1, 2, … . Найдём распределение случайной величины L, т. е. подсчитаем вероятности $p_i = \mathbf{P}(L = i)$, $i = 1, 2, \ldots$.

Выбор шаров описывается бесконечной последовательностью $\omega = (i_1, i_2, \ldots)$, где i_n — это номер шара (от 1 до m) при n-м извлечении. Аналогично решению задачи о ключах для $n \geq i$ получаем:

$$p_i = (m-l)^{i-1} \cdot l \cdot m^{n-i} / m^n = (1 - l/m)^{i-1}(l/m) = (1-p)^{i-1} p, \tag{4}$$

где $i = 1, 2, \ldots,$ $p = l/m$ — доля белых шаров в урне. Вероятности p_i образуют геометрическую прогрессию со знаменателем $q = 1 - p$. Поэтому это распределение называется ***геометрическим***.

Пример 4. *Пуассоновское распределение*. Определим вероятности p_i формулой

$$p_i = \lambda^i e^{-\lambda} / i!, \tag{5}$$

где $i = 0, 1, 2, \ldots$, параметр $\lambda > 0$. Распределение из пар (i, p_i) называется ***законом Пуассона***.[2]

Зададим явно случайную величину $N(\omega)$, имеющую пуассоновское распределение, как функцию на пространстве элементарных событий $\Omega = [0,1]$. Положим $r_{-1} = 0$, $r_i = p_0 + \ldots + p_i$. Пусть на промежутке $A_i = [r_{i-1}, r_i)$ функция $N(\omega)$ равна i. Доопределим её в 1 нулём: $N(1) = 0$. График функции $y = N(\omega)$, состоящий из счётного числа ступенек, изображён на рис. 2. Случайная величина N принимает значения $i = 0, 1, 2, \ldots$ с соответствующими вероятностями $r_i - r_{i-1} = p_i$, что и требуется.

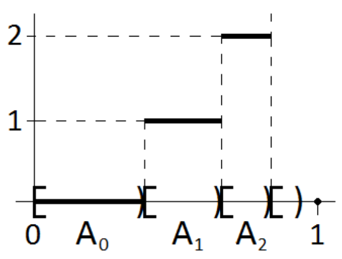

Рис. 2

[2] Симеон Дени Пуассон (1781-1840) — французский математик, механик и физик. Опубликовал более 300 научных работ.

Домашнее задание

Если **второй** буквой вашего **имени** служит:

А, Б, В, Г, Д, Е, Ё, то «своими» являются задачи 5.1 и 5.5;
Ж, З, И, Й, К, Л, М, то «своими» являются задачи 5.2 и 5.6;
Н, О, П, Р, то «своими» являются задачи 5.3 и 5.7;
С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я, то «своими» являются задачи 5.4 и 5.8.

5.1) Какое распределение в модели выбора наудачу с возвращением 2-х шаров из урны с 3-я шарами имеет случайная величина: а) $X = \min\{i_1, i_2\}$; б) $Y = \max\{i_1, i_2\}$? Составить таблицу.

5.2) Какое значение случайной величины S_n с биномиальным распределением $(\mathbf{P}(S_n = i) = C_n^i p^i (1-p)^{n-i},\ i = 0, 1, \ldots, n;\ 0 < p < 1)$ является наиболее вероятным? (*Указание*. Изучите поведение отношения $\mathbf{P}(S_n = i)/\mathbf{P}(S_n = i-1)$ при увеличении i.)

5.3) По n конвертам в случайном порядке раскладываются n писем. Найти вероятность, что в свои конверты попадут письма с заданными номерами $i_1, i_2, \ldots, i_k$, где $k \le n$.

5.4) Какое значение случайной величины N с распределением Пуассона $(\mathbf{P}(N = i) = \lambda^i e^{-\lambda}/i!,\ i = 0, 1, 2, \ldots;\ \lambda > 0)$ является наиболее вероятным? (*Указание*. Изучите поведение отношения $\mathbf{P}(N = i)/\mathbf{P}(N = i-1)$ при увеличении i.)

5.5) Из урны с l белыми и $(m-l)$ чёрными шарами n раз наудачу с возвращением извлекаются шары. Какова вероятность, что среди извлечённых шаров присутствует хотя бы один белый шар и хотя бы один чёрный шар? (*Указание*. Используйте формулу $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.)

5.6) Каждый из шести участников вечеринки пришёл в шляпе. Отправляясь после вечеринки домой, шляпы надевали наугад. Какова вероятность, что ровно половина участников вечеринки надела чужие шляпы? Записать ответ в виде несократимой дроби.

5.7) В условиях задачи 5.3 обозначим через p_n вероятность, что хотя бы одно письмо попадёт в свой конверт. В следующей таблице приведены несколько значений p_n:

n	2	3	4	5	6
p_n	0,500	0,667	0,625	0,633	0,632

Вывести формулу для вероятности p_n и найти предел этой вероятности при $n \to \infty$.

5.8) Среди m экзаменационных билетов l — лёгкие, n студентов по очереди берут билеты без возвращения ($n < m$). Найти распределение случайной величины X — количества студентов, получивших лёгкие билеты, и предел вероятности $\mathbf{P}(X = i)$ при $m \to \infty$, $l = pm$, $0 < p < 1$.

5.9)* Вывести формулу для вычисления вероятности, что при случайном размещении n занумерованных шариков по m ящикам ($n \ge m$) все ящики окажутся занятыми. (*Указание*. Рассмотрите события A_j = {j-й ящик пуст}, $j = 1, \ldots, m$, и примените формулу включений-исключений.)