

▼ TV_lesson-02. Дискретная случайная величина

монета |

МАТБАЗА

Дискретная Случайная Величина (ДСВ).
Вероятностное распределение ДСВ.
Функция распределения. Примеры вероятностных распределений ДСВ для простых испытаний: распределение Бернулли.

13 сентября
занятие 1.2.



Игорь Нехаев
Специалист по анализу данных
и машинному обучению

▼ Вероятностная модель: эксперименты, элементарные исходы и события

▼ Вероятностная модель

Даже если мы не можем заранее предсказать какой будет исход, мы можем описать частоту появления того или иного исхода - задать его вероятность. Построить вероятностную модель.

Вероятностная модель нам может понадобиться для того, чтобы:

- уметь по выборке экспериментальных данных строить модели, описывающие случайную составляющую наблюдения.
- уметь пользоваться свойствами этих моделей для оценки неопределенности результата прогноза, анализа аномалий и выбросов;
- пользоваться моделями для генерации данных и моделировании и исследовании свойств процесса.



Эксперимент — это повторяемый опыт, который может закончиться разными *исходами* (элементарными исходами): исход либо случился, либо не случился и не могут два разных исхода случиться одновременно.

Множество всех элементарных исходов эксперимента, для каждого из которых определена вероятность его появления, принято называть вероятностным пространством:

$$\Omega = \{\omega_i, i = 1..N\} (\Omega = \{U, F, \Sigma\})$$

Мы пока рассматриваем дискретные вероятностные пространства, в котором счетное количество элементарных исходов.

Из него можно выделить подмножества элементарных исходов, которые принято называть событиями:

$$A = \{\omega_j, j = 1..m\}; A \subset \Omega$$

считается, что событие A произошло, если наступил один из исходов, благоприятствующих этому событию;

ВОПРОС. Как называется событие $A = \Omega$?

ВОПРОС. Как называется событие $A = \emptyset$?

Пример “Процесс знакомства с девушкой в парке”.

Элементарных исходов здесь может быть очень много. По-разному девушки могут дать ответ на предложение познакомиться. Но мы можем объединить эти ответы и разбить на три подмножества, три события:

- A = "отказалась знакомиться";
- B = "познакомились, но без продолжения";
- C = "познакомились, дала телефончик".

Пример “Процесс подключения клиента к системе интернет-платежей”.

Здесь можно выделить также несколько элементарных исходов в зависимости от особенностей подключения каждого клиента. Но можно ограничиться только значимыми для нас исходами-событиями:

- A = "не прошел проверку MRM" (сайт не готов, несуществующий клиент, уже есть такой, ...);
- B = "не прошел процедуру CRM" (не заполнен ЛК, не подключил счет);
- C = "подключен";
- D = "транзачит".

Пример "Бросание монеты"

Исходы:

- A = "монета упала орлом вверх";
- B = "монета упала решкой вверх";
- C = "монета упала на ребро";
- D = "монета упала";
- E = "монета зависла в воздухе".

Определение вероятности

Определение. Пусть в результате какого-то испытания могут происходить (или не происходить) некоторые события. Если испытания повторять много (бесконечное) число раз (и при этом сохранять условия его проведения неизменными), то каждому событию можно поставить в соответствие его вероятность - долю испытаний, завершившихся наступлением этого события:

$$p(A) = \lim_{N \rightarrow \infty} \frac{n(A)}{N}$$

Обоснование того, что это возможно - перейти от эксперимента и частоты наступления событий к фиксированной вероятности помогает закон больших чисел. Он доказывает, что при увеличении числа испытаний относительная частота события стремится к постоянной величине, к его вероятности.

Невозможное событие - событие, которое не произойдет никогда, вероятность его появления равна 0. Достоверное событие - событие, которое точно произойдет, вероятность его появления равна 1. Вероятность появления других событий находится в промежутке от 0 до 1:

$$0 \leq p(A) \leq 1$$

Для любого события A можно определить противоположное событие "не A " (A не произошло), вероятность такого события равно:

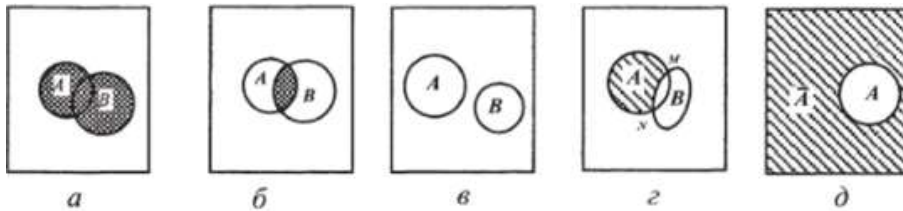
$$p(\bar{A}) = 1 - p(A).$$

Это следует из логики, что одно из двух событий обязательно наступит - или A или \bar{A} . Поэтому сумма относительных частот этих событий в сумме всегда равна 1.

▼ Алгебра событий

Алгебра событий = Множество всевозможных событий и операции над ними (сумма, произведение, разность). Свойства этих операций.

Так как алгебре событий соответствует алгебра множеств, то свойства достаточно понятны (см. диаграммы Эйлера-Венна).



Сумма событий

Опр. Под суммой (объединением) двух событий A и B будем подразумевать событие $C = A + B$, которое считается наступившим, если произошло одно из двух событий или A или B (рис.а).

Пример. A = 'число выпавших очков на кубике четно', B = 'число выпавших очков на кубике больше трех'. Тогда событие $A + B$ - это когда выпало либо четное количество либо больше трех, т.е. когда выпала одна из граней: 2, 4, 5, 6.

Произведение событий

Опр. Произведением (пересечением) двух событий A и B будем подразумевать событие $C = A \cdot B$, которое считается наступившим, если произошли оба события и A и B (рис.б).

Пример. A = 'число выпавших очков на кубике четно', B = 'число выпавших очков на кубике больше трех'. Тогда событие $A \cdot B$ - это когда выпало и четное количество и больше трех, т.е. когда выпала одна из граней: 4, 6.

▼ Разность событий

Опр. Разностью двух событий A и B будем называть событие $C = A - B$, которое считается наступившим, если произошло событие A , но не произошло событие B (рис.г).

Пример. A = 'число выпавших очков на кубике четно', B = 'число выпавших очков на кубике больше трех'. Тогда событие $A - B$ - это когда выпало четное количество, но не больше трех, т.е. когда выпала 2.

Кстати, событие, противоположное событию A , можно определить как разность:

$$\bar{A} = \Omega - A$$

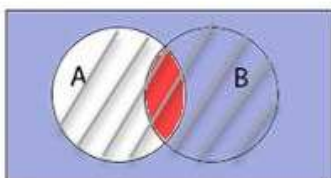


Рис. Диаграмма Венна: прямоугольник включает все элементарные исходы, точки в круге A символизируют исходы, благоприятствующие событию A , область, закрашенная красным цветом - исходы, соответствующие событию $A \cdot B$, заштрихованная область - событию $A + B$.

Нас будет интересовать как подсчитать вероятность суммы или произведения событий, если мы знаем вероятности этих событий:

$$p(A + B) = ?$$

$$p(A \cdot B) = ?$$

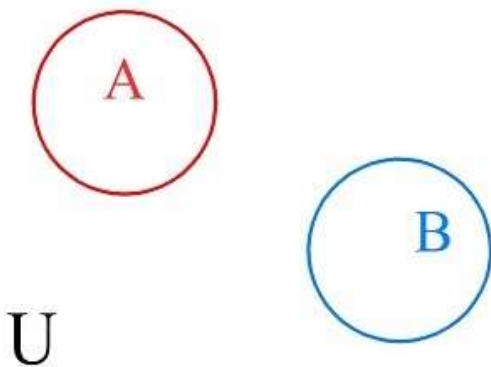
ВОПРОС. "Как связаны между собой $p(A)$, $p(B)$, $p(A + B)$ и $p(A \cdot B)$?"

▼ Полная группа событий. Вероятность суммы событий.

Нас вряд ли будут интересовать все возможные события (все возможные подмножества элементарных исходов). Нам важно выделить определенное множество событий - несовместные (взаимоисключающие) и образующие полную группу.

Опр. События A и B называются несовместными, если в результате эксперимента они не могут произойти одновременно (нет общих благоприятных исходов):

$$A \cap B = \emptyset$$



Для таких событий можно утверждать, что вероятность их суммы равна сумме их вероятностей:

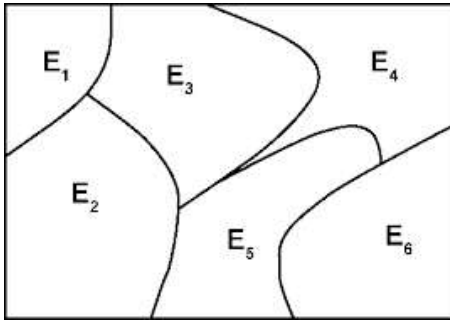
$$p(A + B) = p(A) + p(B), \quad A, B - \text{несовместные.}$$

Опр. Конечный или счетный набор попарно несовместных событий E_1, E_2, \dots, E_m , таких что:

- никакое из них не является невозможным: $p(E_i) > 0, \forall i$;
- в результате эксперимента одно из этих событий обязательно произойдет:

$$E_1 \cup E_2 \cup \dots \cup E_m = \Omega$$

называется Полной группой событий или разбиением пространства Ω .



Вот такое разбиение нам и нужно, чтобы перейти к понятию случайной величины.



ЗАДАНИЕ

ВОПРОС: Является ли пара противоположных событий (A, \bar{A}) полной группой?

ВОПРОС: Какое событие является противоположным для события D = "транзачит" (из примера о подключении клиентов)?

ВОПРОС: Образуют ли события A = "монета упала орлом вверх", B = "монета упала решкой вверх" полную группу событий (из примера о бросании монеты)?

▼ Независимые события. Вероятность произведения событий

Опр. События A и B называются **независимыми**, если вероятность наступления одного из них не зависит от того, наступило другое событие или нет.

Например, события A = "она придет в черном платье" и B = "она придет в хорошем настроении" могут быть как зависимыми, так и не зависимыми (зависит от характера девушки).

Если мы бросаем два кубика, то события A = 'число выпавших очков на первом кубике четно', B = 'число выпавших очков на втором кубике больше трех' явно независимые.

Для таких событий можно утверждать, что вероятность их произведения равна произведению их вероятностей:

$$p(A \cdot B) = p(A) \cdot p(B), \quad A, B — \text{независимые.}$$



ВОПРОС. Являются ли несовместные события независимыми?

ПРИМЕР. Пусть мы бросаем два кубика.

Событие A = 'число выпавших очков на первом кубике четно', B = 'число выпавших очков на втором кубике больше трех'.

Имеем:

$$p(A) = p(\text{'число выпавших очков на первом кубике четно'}) = \frac{3}{6}$$

$$p(B) = p(\text{'число выпавших очков на втором кубике больше трех'}) = \frac{3}{6}$$

и если мы подсчитаем вероятность произведения этих двух событий по формуле, учитывающая кол-во благоприятствующих событий, то получим:

$$p(A \cdot B) = \frac{n(AB)}{n} = \frac{3 \cdot 3}{36} = \frac{1}{4} = p(A) \cdot p(B)$$

▼ Случайная величина. Дискретная случайная величина

Для того, чтобы можно было проводить исследования и использовать вероятностные распределения для построения более сложных вероятностных пространств, удобно закодировать исходы эксперимента (или события из полной группы интересующих нас событий) числами и считать, что в результате эксперимента какая-то величина (параметр) принял какое-то значение.

Опр. Случайная Величина (СВ) - это параметр, значения которого представляют собой закодированные числами исходы некоторого случайного явления или эксперимента. Так как результат эксперимента заранее неизвестен, то и неизвестно заранее какое значение примет данная величина. Поэтому она называется Случайной Величиной.

Опр. Дискретной случайной величиной (ДСВ) будем называть случайную величину, которая принимает конечное или счетное количество возможных числовых значений.

Это нас не ограничивает при работе с категориальными параметрами. Мы можем каждый исход для категориального параметра закодировать числом. Однако с такими данными как с числами работать бессмысленно, если не определен хотя бы порядок на множестве значений СВ. И в рамках урока мы предложим путь как можно кодировать такие значения.

Обозначения

Случайные величины обозначают заглавными латинскими буквами (X, Y, Z), а их значения - строчными буквами (x_i, y_i, z_i).

▼ Вероятностное распределение ДСВ. Функция вероятности

Теперь, чтобы задать закон распределения вероятностей, надо сопоставить каждому возможному значению С.В. его вероятность появления. Например, если С.В. X может принимать три значения x_1, x_2, x_3 , то надо поставить им в соответствие вероятности:

$$(p_1, p_2, p_3).$$

В общем случае **дискретная случайная величина X** может принимать счетное число значений, которые можно перенумеровать x_1, x_2, x_3, \dots с вероятностями соответственно p_1, p_2, p_3, \dots , для которых выполняется условие неотрицательности и нормировки:

$$\sum_{i=1}^{\infty} p_i = 1; p_i \geq 0, \forall i.$$

Таким образом, задать соответствие вида:

$$p(X = x_i) = p_i, \forall x_i,$$

удовлетворяющее условию неотрицательности и нормировки, это значит определить **закон распределения случайной величины** в виде **функции вероятности** на множестве исходов.

Условие нормировки логично вытекает из того, что мы учли все возможные исходы и эти исходы несовместны.

Так как эти события несовместны, то:

$$p((X = 1) + (X = 2) + (X = 3) + \dots) = p(X = 1) + p(X = 2) + p(X = 3) + \dots = \sum_{i=1}^m p_i = 1$$

Как и обычную функцию ее можно задать в виде таблицы, если множество элементарных исходов конечно, или в виде формулы или алгоритмически в общем случае.

▼ Пример "Процесс знакомства с девушкой в парке".

Ранее мы решили описать этот процесс с помощью полной группы из 3-х исходов/событий:

- A = "отказалась знакомиться";
- B = "познакомились, но без продолжения";
- C = "познакомились, дала телефончик".

Введем С.В. X и закодируем ее значения с учетом их полезности для нас:

исход	x_i	$p(x_i)$
"отказалась знакомиться"	0	0.25
"познакомились, но без продолжения"	0.5	0.5
"познакомились, дала телефончик"	1	0.25

Такое кодирование может быть информативным при суммировании значений данных величин в серии из N испытаний. Оно будет отображать полезность наших попыток знакомства.



ЗАДАЧА.

Пусть мы имеем вероятностное распределение С.В. X из примера выше

Найдите вероятность события, что с.в. X примет значение больше 0:

$$p(X > 0) = ?$$

▼ Пример "Бросание монеты до появления решки"

Пусть наш эксперимент состоит в подбрасывании монеты до тех пор, пока не выпадет решка. С.В. X - количество таких бросаний. Таблично мы не сможем задать распределение вероятности из-за бесконечного числа возможных значений С.В.:

x_i	1	2	3	4	...
$p(x_i)$	0.5	0.25	0.125	0.0625	...

(считаем монету идеальной). Но можем задать закон распределения вероятности теперь с помощью отображения:

$$p(X = k) = 2^{-k}, \quad k = 1, 2, 3, \dots$$

Можно убедиться, что данный закон отвечает условию нормировки:

$$\sum_{k=1}^{\infty} p(X = k) = 0.5 * (1 + 0.5 + 0.25 + 0.125 + \dots) = 0.5 * \frac{1}{1 - 0.5} = 1$$



ЗАДАЧА

Пусть мы имеем вероятностное распределение С.В. X из примера выше

Найдите вероятность события, что с.в. X примет значение больше 2 (придется бросать монету более 2-х раз):

$$p(X > 2) = ?$$

▼ Распределение Бернулли

Это распределение является элементарным кирпичиком при построении большого числа других более сложных распределений.

Мы рассматриваем полную группу событий, состоящую всего из двух исходов "УСПЕХ" и ему противоположное событие "НеУСПЕХ".

Вводим ДСВ, которая кодирует эти исходы как 1 ("УСПЕХ") и 0 ("НеУСПЕХ").

Достаточно задать вероятность успеха, т.е. $p(X=1) = p$, чтобы задать распределение вероятностей:

x_i	0	1
$p(x_i)$	$1-p$	p

Например, если в примере со знакомством мы определяем исход УСПЕХ, если мы познакомились с девушкой, то распределение Бернулли будет выглядеть так:

x_i	0	1
$p(x_i)$	0.25	0.75

для примера с подключением клиентов к системе интернет-платежей, распределение Бернулли будет выглядеть так (успех = "подключение"):

x_i	0	1
$p(x_i)$	0.77	0.23

Само распределение Бернулли не впечатляет, но нас будут интересовать серии из испытаний Бернулли. И вот здесь мы найдем много интересного в следующих занятиях.

▼ Графическое представление Закона распределения ДСВ. Полигон относительных частот.

Графически Закон распределения часто задают (или иллюстрируют) в виде графика функции вероятности или полигона относительных частот (если имеются данные наблюдений С.В.).

Рассмотрим график функции вероятности для примера с бросанием монеты (испытание Бернулли!) до выпадения решки.

```
import numpy as np
import matplotlib.pyplot as plt
plt.figure(figsize=(4, 3))
plt.style.use('ggplot')
```

```
# описываем функцию вероятности
def prob_moneta(n):
    return 1 / 2 ** n
```

```
# для графика возьмем 10 первых значений
n = 10
```

```
nn = np.arange(1, n+1, dtype=int)
p_n = prob_moneta(nn)
```

```
# построим график функции вероятности
```

```
plt.scatter(nn, p_n, s=30)
```

```
for i in range(len(nn)):
```

```
    plt.plot([nn[i], nn[i]], [0, p_n[i]], c='grey')
```

```
plt.suptitle("Рис. График функции вероятности для ДСВ 'кол-во бросаний монеты'", y=0, fontsize=8);
```

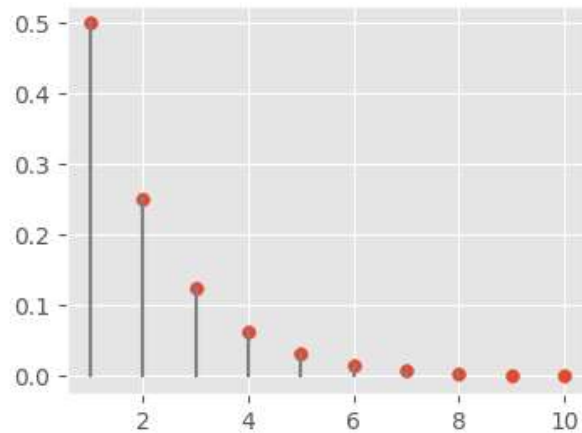


Рис. График функции вероятности для ДСВ 'кол-во бросаний монеты'

Смоделируем бросание монеты, сделаем 100 наблюдений - бросаний до появления решки (1) и построим полигон относительных частот. Совместим его с графиком функции вероятности.

```
import numpy as np
import pandas as pd
```

```
def experiment_fun():
```

```
    n = 1
```

```
    x = np.random.random()
```

```
    while x < 0.5:
```

```
        n += 1
```

```
        x = np.random.random()
```

```
    return n
```

```
n_x = 100
```

```
n_exp = pd.Series([experiment_fun() for _ in range(n_x)])
```

```
n_exp = n_exp.value_counts()
```

```
p_n_exp = n_exp.values / n_x
```

```
n_exp, p_n_exp
```

```
(1    54
 2    22
 3     9
 4     8
 5     4
 6     2
 9     1
 dtype: int64,
 array([0.54, 0.22, 0.09, 0.08, 0.04, 0.02, 0.01]))
```

```
# построим полигон относительных частот
```

```
plt.figure(figsize=(4, 3))
```

```
n_exp = len(p_n_exp)
```

```
plt.scatter(nn[:n_exp], p_n[:n_exp], s=30, label='модель')
```

```
plt.scatter(nn[:n_exp], p_n_exp, s=30, c="blue", label='эксперимент')
```

```
for i in range(n_exp):
```

```
    plt.plot([nn[i], nn[i]], [0, p_n_exp[i]], c='grey')
```

```
plt.suptitle("Рис. Полигон относительных частот для ДСВ 'кол-во бросаний монеты'", y=0, fontsize=8)
```

```
plt.legend()  
plt.show()
```

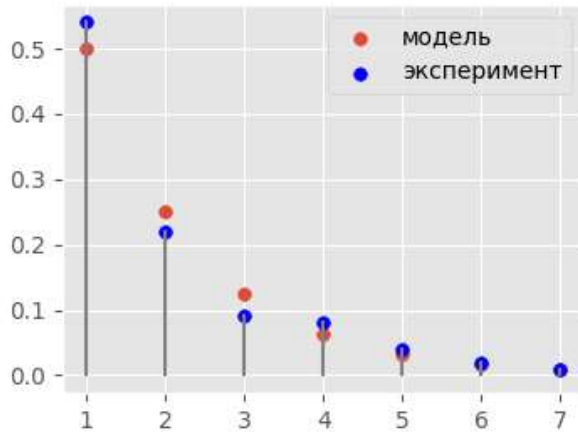


Рис. Полигон относительных частот для ДСВ 'кол-во бросаний монеты'

Резюме

- любой результат, получаемые данные, можно рассматривать как результат некоторого эксперимента;
- каждому результату эксперимента, событию, можно сопоставить его относительную частоту появления (вероятность) и множество благоприятных ему элементарных исходов;
- на множестве всех исходов/событий можно определить алгебру, ввести операции сложения (объединения), умножения (пересечения), отрицания (дополнения).
- мы можем выделить интересующие нас исходы эксперимента (важно чтобы они составляли полную группу событий!);
- задать С.В. - означает закодировать интересующие нас исходы эксперимента числами (значения С.В.) и задать функцию вероятности, определенную на множестве возможных значений С.В. (удовлетворяющая условию неотрицательности и нормировки);
- распределение Бернулли - это самое простое распределение задает С.В. с двумя значениями 0 и 1 и

✓ 0 сек. выполнено в 11:49

● ✕