# TV_lesson-04. Конструирование законов распределения вероятностей составных ДСВ

> Построение распределения вероятности для суммы, разности и произведения СВ:
>
> - сумма очков при бросании двух, трех, ... кубиков;
> - произведение очков при бросании двух кубиков.

> Построение вероятностных распределений ДСВ для составных испытаний:
>
> - биномиальное распределение,
> - отрицательное биномиальное распределение,
> - распределение Пуассона.

> Смеси распределений. Пример испытания, приводящего к смеси вероятностных распределений.

## Построение распределения вероятности для результатов операций со случайными величинами

- Со СВ, как и с обычными величинами можно оперировать - складывать/вычитать, уножать/делить, ...
- отличие только в том, что в результате операции получается также СВ со своим распределением вероятностей ее значений.

### Сумма, разность и произведение двух ДСВ

Рассмотрим две СВ X и Y, для которых известны их законы распределения, т.е. их функции вероятностей:

$$p_X() : x_i \to p_i,$$
$$p_i \ge 0,\ i = 1..n;\ \sum_{i=1}^{n} p_i = 1.$$
$$p_Y() : y_j \to q_j,$$
$$q_j \ge 0,\ j = 1..m;\ \sum_{j=1}^{m} q_j = 1$$

Какой закон распределения будет у СВ $Z = X + Y$, $Z = X - Y$, $Z = X \cdot Y$ ?

Рассмотрим случай независимых ДСВ.

## Примеры постановок задач

### Пример **“Продолжение знакомства с девушкой”**.

> Ура, мы познакомились с симпатичной девушкой и договорились с нею встретиться в субботу в парке на том же самом месте! Мы уже опытные ловеласы и понимаем, что успех - т.е. возможность дальнейшего общения и продолжение серьезных отношений зависит от

- в каком настроении придет девушка - в хорошем (X = 1), в нейтральном (X = 0.5) или не в настроении (X = 0);

- в каком состоянии будет парень (в том числе и в финансовом состоянии): в хорошем и финансово и физически (Y = 2), в хорошем физически, но не очень финансово (Y = 1), в плачевном состоянии (Y = 0).

Практика показывает, что шансы на продолжение отношений хорошие:

- при знакомстве с состоятельными девушками, если $X + Y \geq 2.5$;
- с просто милыми девушками, если $X \cdot Y \geq 1$.

> Определите шансы парня на продолжение отношений
> - а) с состоятельной красавицей;
> - б) с просто милашкой;

> если известны законы распределения для (будем считать) независимых СВ X и Y:

| $x_i$ | 0 | 0.5 | 1 |
|---|---|---|---|
| $p_i$ | 0.2 | 0.5 | 0.3 |

| $y_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $p_i$ | 0.1 | 0.5 | 0.4 |


> ВОПРОС. Как можно описать содержательно событие $X + Y \geq 2.5$?

## Пример. **Анализ резюме претендентов на вакантную должность**

Рассмотрим две независимые СВ X и Y. Например, кол-во проанализированных за день резюме на вакантную должность двумя разными работниками. Так как каждый из них, будем считать, работает независимо от другого (возможно, что они работают удаленно и не знают друг друга), то и результаты их работы не связаны никак.

Также их результат может зависеть от очереди резюме претендентов. Они просто берут из очереди первую необслуженную заявку и ее рассматривают. Если в очереди заявки завершаются, то тогда и результат их работы будет зависеть от количества заявок в очереди на этот день. Как вырожденный случай такой зависимости - это когда нет вакансий, а значит и рассматривать нечего. Учет возможной ограниченности очереди - это мостик к рассмотрению смеси распределений. А пока будем считать очередь неограниченной. Т.е. кол-во рассмотренных резюме ограничивается только производительностью работников.

Ясно, что заранее спрогнозировать какое кол-во резюме i-й работник сможет сегодня проанализировать сложно, так как резюме могут сильно отличаться друг от друга объемом и составом ключевых слов, с помощью которых описываются скиллы. Поэтому X = "кол-во проанализированных резюме за день работником №1" и Y = "кол-во проанализированных резюме за день работником №2" можно считать случайными величинами.

Пусть мы набрали достаточно данных, чтобы утверждать, что полученные эмпирические распределения вероятностей для СВ X и Y достаточно устойчивы.

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $p_i$ | 0 | 0 | 0.2 | 0.3 | 0.3 | 0.2 |

| $y_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $p_i$ | 0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 |

> Нас будут интересовать ответы на следующие вопросы:

- что можно сказать о суммарном кол-ве проанализированных резюме за день (распределение суммы)?
- можно ли сказать, что какой-то из работников лучше работает или нет (распределение разности)?


> ВОПРОС. Как можно сформулировать словесно событие $A = (X = 4) \cdot (Y = 5)$?

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

> Рассмотрим две независимые СВ:
>
> - X = "кол-во дней, затраченных на проверку заявки (MRM), которая завершилась успешно";
> - Y = "кол-во дней на подключение клиента после успешной проверки заявки";

Нас будет интересовать общее кол-во дней на обработку заявки клиента, которое завершилось успешным его подключением.


> ВОПРОС. Как можно сформулировать словесно событие $A = (X + Y \leq 5)$?

## Построение функции вероятностей для $Z = X\ op\ Y$, где $X, Y$ - НЕзависимые СВ.

Рассмотрим процесс построения функции вероятностей для совместного распределения двух незавсиьымх СВ $X, Y$ на примере “Продолжение знакомства с девушкой”.

В это примере нам известны законы распределения для (будем считать) независимых СВ X (настроение девушки) и Y (состояние парня):

Функция вероятности для СВ X

| $x_i$ | **0** | **0.5** | **1** | $\sum$ |
|---|---|---|---|---|
| $p_i$ | 0.2 | 0.5 | 0.3 | 1.0 |

Функция вероятности для СВ Y

| $y_j$ | **0** | **1** | **2** | $\sum$ |
|---|---|---|---|---|
| $q_j$ | 0.1 | 0.5 | 0.4 | 1.0 |

Построим закон распределения вероятности для пары (X, Y). Для этого нам надо рассмотреть все возможные пары значений этих величин $(x_i, y_j)$ и найти соответствующие вероятности их появления одновременно в одном испытании.

Для этого подготовим такую таблицу:

| $x_i \setminus y_j$ | **0** | **1** | **2** | $p_i$ |
|---|---|---|---|---|
| 0 | . | . | . | 0.2 |
| 0.5 | . | . | . | 0.5 |
| 1 | . | . | . | 0.3 |
| ------------- | --- | --- | --- | ---- |
| $q_j$ | 0.1 | 0.5 | 0.4 | 1.0 |

Чтобы ее заполнить, нам надо найти для каждой пары значений $(x_i, y_j)$ вероятность ее появления. Это все равно, что найти вероятность произведения событий:

$$p_{i,j} = p(x_i, y_j) = p((X = x_i) \cdot (Y = y_j))$$

Так как мы считаем данные СВ независимыми, то независимы и события $(X = x_i), (Y = y_j)$, а это означает, что

$$p_{i,j} = p((X = x_i) \cdot (Y = y_j)) = p(X = x_i) \cdot p(Y = y_j) = p_i \cdot q_j$$

Используя последнее равенство, мы легко заполним таблицу, задающую закон распределения вероятности для пары X,Y:

| $x_i \backslash y_j$ | **0** | **1** | **2** | \| $p_i$ |
|---|---|---|---|---|
| 0 | 0.02 | 0.1 | 0.08 | \| 0.2 |
| 0.5 | 0.05 | 0.25 | 0.2 | \| 0.5 |
| 1 | 0.03 | 0.15 | 0.12 | \| 0.3 |
| ------------ | ---- | ---- | ---- | ---- |
| $q_j$ | 0.1 | 0.5 | 0.4 | \| 1.0 |

Зная закон совместного распределения СВ X, Y, мы можем теперь легко найти закон распределения для результата любой бинарной операции с этими СВ.


---

ВОПРОС. Чему равна вероятность того, что $X + Y = 3$ ?

---

## Построение функции вероятностей для $Z = X + Y$, где $X, Y$ - независимые СВ

> Воспользуемся построенным законом распределения для векторной СВ (X,Y) и найдем закон распределения для СВ
> - $Z = X + Y$

Для этого воспользуемся той же таблицей для наглядности и в каждую ячейку таблицы впишем получающееся значение $z_{i,j} = x_i + y_j$:

| $x_i \backslash y_j$ | **0** | **1** | **2** |
|---|---|---|---|
| 0 | 0 (0.02) | 1 (0.1) | 2 (0.08) |
| 0.5 | 0.5 (0.05) | 1.5 (0.25) | 2.5 (0.2) |
| 1 | 1 (0.03) | 2 (0.15) | 3 (0.12) |

В скобках мы оставили вероятности появления данных пар значений X,Y.

Теперь мы можем выписать функцию вероятности для $Z = X + Y$. Для этого надо просто просуммировать вероятности всех случаев, когда $Z = z_k$. Например,

$$p(Z = 1) = p((X = 0) \cdot (Y = 1) + (X = 1) \cdot (Y = 0)) = p((X = 0) \cdot (Y = 1)) + p((X = 1) \cdot (Y = 0)) = 0.1 + 0.03 = 0.13$$

Получаем такое распределение:

| $z_k$ | **0** | **0.5** | **1** | **1.5** | **2** | **2.5** | **3** | \| $\sum$ |
|---|---|---|---|---|---|---|---|---|
| $p_k$ | 0.02 | 0.05 | 0.13 | 0.25 | 0.23 | 0.2 | 0.12 | \| 1.0 |


ВОПРОС. Чему равна вероятность успешного продолжения знакомства с состоятельными девушками, если

$$p(\text{" УСПЕХ "}) = p(X + Y \geq 2.5)$$

## Построение функции вероятностей для $Z = X \cdot Y$, где $X, Y$ - независимые СВ

> Воспользуемся построенным законом распределения для векторной СВ (X,Y) и найдем закон распределения для СВ
> - $Z = X \cdot Y$

Аналогично примеру со сложением впишем в таблицу совместного распределения значения СВ $Z = X \cdot Y$:

| $x_i \backslash y_j$ | **0** | **1** | **2** |
|---|---|---|---|
| 0 | 0 (0.02) | 0 (0.1) | 0 (0.08) |
| 0.5 | 0 (0.05) | 0.5 (0.25) | 1 (0.2) |
| 1 | 0 (0.03) | 1.0 (0.15) | 2 (0.12) |

$$p(Z = 1) = p((X = 0.5) \cdot (Y = 2) + (X = 1) \cdot (Y = 1)) = p((X = 0.5) \cdot (Y = 2)) + p((X = 1) \cdot (Y = 1)) = 0.2 + 0.15 = 0.35$$

Получаем такое распределение:

| $z_k$ | **0** | **0.5** | **1** | **2** | $\mid\sum$ |
|---|---|---|---|---|---|
| $p_k$ | 0.28 | 0.25 | 0.35 | 0.12 | \| 1.0 |


ВОПРОС. Чему равна вероятность успешного продолжения знакомства с просто милыми девушками, если

$$p(\text{" УСПЕХ "}) = p(X \cdot Y \geq 1)$$

## Построение функции вероятностей для $Z = XopY$, где $X, Y$ - зависимые СВ

здесь алгоритм тот же, за исключением того, что теперь мы не можем посчитать вероятность $p((X = x_i) \cdot (Y = y_j))$ как произведение вероятностей. Теперь эту вероятность надо оценивать с помощью наблюдений и вычисления относительной частоты появления данной пары значений вместе.

Также можно использовать формулу с условной вероятностью $p(Y = y_j / X = x_i)$, если она известна или может быть легко найдена:

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

для нашего случая она будет выглядеть так:

$$p((X = x_i) \cdot (Y = y_j)) = p(X = x_i) \cdot p(Y = y_j / X = x_i))$$

### ПРИМЕР. В списке резюме на вакантную должность имеется 3 подробно расписанных резюме и 5 кратких.

Какова вероятность того, что второй рецензент получит краткое резюме, если 1-й рецензент получил подробное резюме на анализ.

Здесь

- событие A = "1-й рецензент получил подробное резюме на анализ"
- событие B = "2-й рецензент получил краткое резюме на анализ".

$$p(B/A) = \frac{5}{5+2} = \frac{5}{7}$$
$$p(A \cdot B) = p(A) \cdot p(B/A) = \frac{3}{8} \cdot \frac{5}{7} = \frac{15}{56}$$


ВОПРОС. Какова вероятность того, что второй рецензент получит краткое резюме, если неизвестно, какое резюме получил на анализ 1-й рецензент?

## Построение вероятностных распределений ДСВ для составных испытаний

### **Биномиальное распределение**

> Данное распределение получается, если мы проводим серию из N испытаний Бернулли. Обозначим через $X_j, j = 1..N$ - С.В., каждая из которых подчиняется закону Бернулли с параметром $p$ и принимающие значение 1 или 0 в зависимости от того наступил Успех или НеУспех в одном испытании.

> Вероятность успеха в каждом испытании, $p$ одна и та же и и не зависит от результатов предыдущих испытаний.

> Рассмотрим случайную величину Y, которая показывает количество успехов в серии из N испытаний. Мы можем считать, что С.В. $Y$ получается при сложении С.В. $X_j, j = 1..N$, каждая из которых подчиняется закону Бернулли с параметром $p$:

$$Y = X_1 + X_2 + \ldots + X_N.$$

Закон распределения для $X_j$ результата j-го испытания Бернулли:

| $x_i$ | 0 | 1 |
|---|---|---|
| $p_i$ | 1-p | p |

Тогда мы можем утверждать, что:

- вероятность того, что во всех испытаниях был успех равна $$p(Y=N)=p((X_1=1)\cdot(X_2=1)\cdot..\cdot(X_N=1))=p^N;$$
- вероятность того, что во всех испытаниях был Неуспех равна $$p(Y=0)=(1-p)^N;$$
- вероятность того, что в серии из N испытаний успех наступил только один раз равна $$p(Y=1)=N\cdot p^1\cdot(1-p)^{N-1}.$$

---

Но чтобы построить биномиальный закон распределения, нам надо определить значение функции вероятности для всех возможных исходов, а их ровно N+1:

| $x_i$ | 0 | 1 | 2 | ... | N |
|---|---|---|---|---|---|
| $p_i$ | (1-p)**N | Np(1-p)**(N-1) | .. | .. | $p^N$ |

> Чтобы найти вероятность события, что "успех наступил ровно $k$ раз в серии из $N$ испытаний", надо:
>
> - найти вероятность того, что успех произошел в каких-то (пусть первых) k испытаниях: $$p(X_1=1,..,X_k=1,X_{k+1}=0,..,X_N=0)=p^k\cdot(1-p)^{N-k}$$
> - найти количество комбинаций $k$ испытаний, в которых мог произойти успех в общем числе испытаний N: $$C_N^k=\frac{N!}{(N-k)!\cdot k!}$$
>   Фактически, мы решаем задачу - сколькими способами мы можем $k$ неразличимых шаров (успехов) разместить в N перенумерованных лунках (испытаниях).

Просуммировав все эти несовместные исходы в серии испытаний, мы получим формулу $k$ успехов в серии из $n$ испытаний Бернулли с вероятностью успеха $p$:

$$p_n(k)=C_n^k\cdot p^k\cdot(1-p)^{n-k}.$$

Это и есть **функция вероятности для биномиального распределения** для серии из N испытаний с вероятностью успеха $p$.

Про такую С.В. говорят, что она подчиняется биномиальному закону распределения с параметрами $n$ и $p$:

$$X\in Bin(n,p)$$

---

## Почему биномиальный?

---

> Биномиальным данное распределение называется потому, что здесь фактически зайдествован бином Ньютона:
>
> $$1=((1-p)+p)^n=C_n^0\cdot p^0\cdot(1-p)^n+C_n^1\cdot p^1\cdot(1-p)^{n-1}+..+C_n^k\cdot p^k\cdot(1-p)^{n-k}+..+C_n^n\cdot p^n\cdot(1-p)^0$$

> например, для серии из двух испытаний имеем: $$1=p_2(0)+p_2(1)+p_2(2)=((1-p)+p)^2=(1-p)^2+2p(1-p)+p^2$$

---

Иллюстрация Биномиального распределения при N=10. [Доска Гальтона]


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 А = "монета упала орлом вверх", В = "монета упала решкой вверх" полную группу событий (из примера о бросании монеты)?

---

## Нормальный закон распределения - как предел биномиального распределения

Рассмотрим иллюстрацию биномиального закона распределения:

https://www.edumedia-sciences.com/ru/media/905

А теперь рассмотрим СВ "Сумма очков в серии бросания кубика". Давайте рассмотрим иллюстрацию того, как меняется закон распределения суммы очков в серии из 2-х, 3-х и более бросаний кости:

Иллюстрация распределения вероятности для суммы очков в серии бросания кости

## Закон Пуассона - как аппроксимация биномиального распределения

Рассмотрим СВ Y = "Кол-во заявок на обслуживание за час". Мы уже отмечали ранее, что данную величину можно рассматривать как серию испытаний Бернулли с вероятностью успеха p (УСПЕХ = "вызов в эту минуту").

Значение данной СВ мы можем найти просто просуммировав значения СВ Xi, i=1..60 полученные при ежеминутном испытании Бернулли:

$$Y = \sum_{i=1}^{60} X_i$$

Давайте рассмотрим иллюстрацию того, как меняется закон распределения кол-ва вызовов в серии из 10, 20 и более минут, с использованием биномиального закона и сраним его с законом Пуассона:

$$p_\lambda(k) = \frac{(\lambda)^k \cdot e^{-\lambda}}{k!}$$

где $\lambda = n \cdot p = 0.04 \cdot n$.

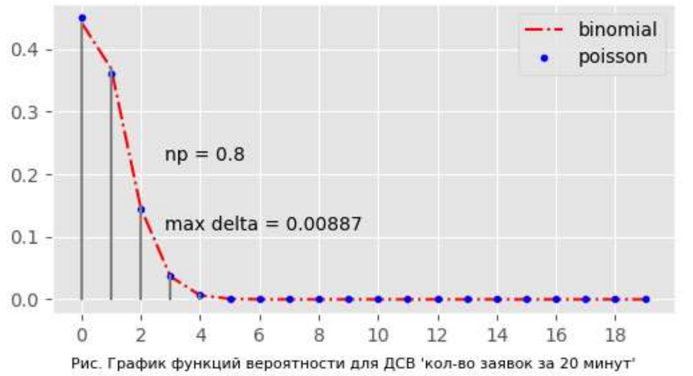
```
# импорт библиотеки для работы с массивами, последовательностями и для построения графиков
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

plt.style.use('ggplot')
```

используем готовые функции для законов распределения дискретных СВ из библиотеки scipy
https://docs.scipy.org/doc/scipy/reference/stats.html#discrete-distributions

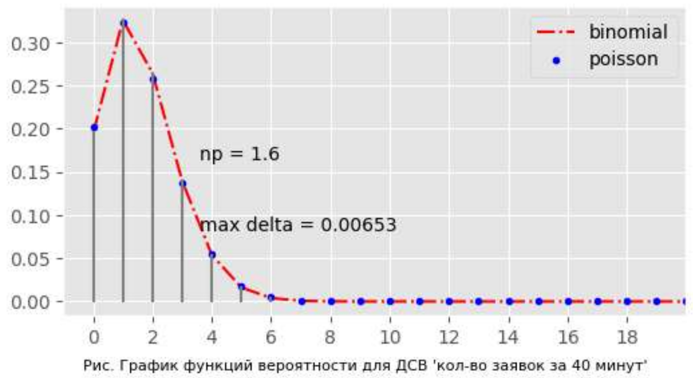
```
# Загрузим библиотеку с вероятностными законами распределения
from scipy.stats import binom, poisson

def create_bin_prob(n, p):
    def bin_prob(k):
        return binom.pmf(k, n, p)
    return bin_prob

def create_poisson_prob(n, p):
    lamda = n * p
    def poisson_prob(k):
        return poisson.pmf(k, lamda)
    return poisson_prob

def drawing(n, p=0.04, xmax=20):
    plt.figure(figsize=(6, 3))
    # совместим график функции вероятности биномиального закона с законом Пуассона
    bin_prob = create_bin_prob(n, p)
    poisson_prob = create_poisson_prob(n, p)
    ki = np.array(range(n))
    bin_pi = bin_prob(ki)
    poisson_pi = poisson_prob(ki)
    plt.plot(ki, bin_pi, '-.', c='red', label="binomial")
    plt.plot(ki, poisson_pi, '.', c='blue', label="poisson")
    for i in range(len(ki)):
        plt.plot([ki[i], ki[i]], [0, bin_pi[i]], c='grey')
    plt.text(n*p + 2, max(bin_pi) / 2, s=f"np = {n * p}")
    max_delta = np.round(np.max(np.abs(poisson_pi - bin_pi)), 5)
    plt.text(n*p + 2, max(bin_pi) / 4, s=f"max delta = {max_delta}")
    plt.suptitle(f"Рис. График функций вероятности для ДСВ 'кол-во заявок за {n} минут'", y=0, fontsize=8)
    plt.xlim(-1, 20)
    plt.xticks(np.arange(0, xmax, 2))
    plt.legend()
```

```
    plt.show()
```

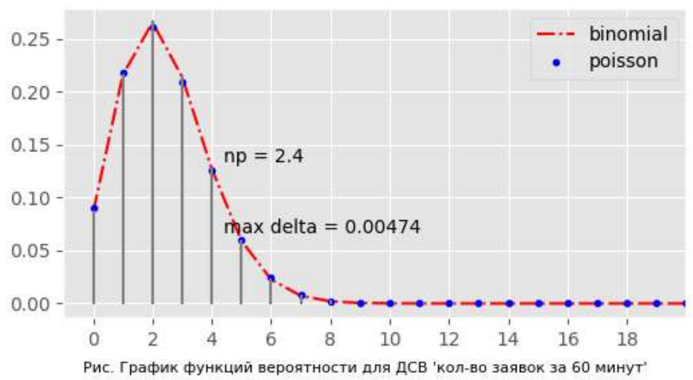
```
drawing(n=20)
```



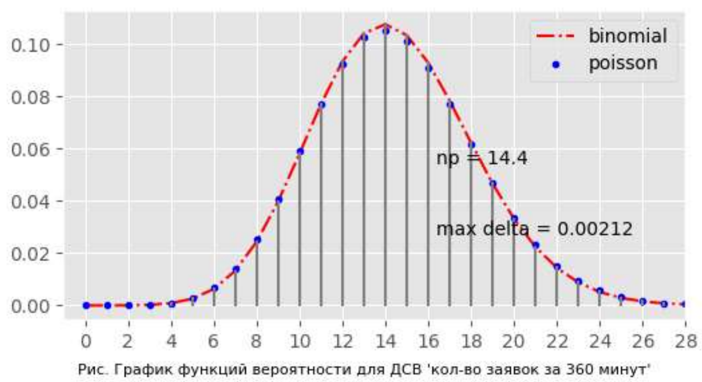

```
drawing(n=40)
```




```
drawing(n=60)
```




```
drawing(n=60*6, xmax=30)
```




## Отрицательный биномиальный закон

Это закон распределения кол-ва испытаний в серии испытаний Бернулли до появления "Успеха". Если нам известна вероятность успеха в одном испытании, то можно легко построить данный закон:

| $x_i$ | 1 | 2 | 3 | 4 | ... |
|---|---|---|---|---|---|
| p($x_i$) | p | qp | $q^2p$ | $q^3p$ | ... |

### Пример "**Бросание монеты до появления решки**"

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

| $x_i$ | 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:

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

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 С.В. $X$ подчиняется отрицательному биномиальному распределению с вероятностью успеха $p=0.3$.

Найдите вероятность события, что с.в. $X$ примет значение больше 2:

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

Что более вероятней, что СВ окажется больше 2-х или противоположное событие ($X\leq 2$)?

---

## Смеси распределений. Пример испытания, приводящего к смеси вероятностных распределений.

### Пример формирования смеси распределений

Рассмотрим формирование смеси распределений на примере задачи "**Анализ резюме претендентов на вакантную должность**".

Пусть у нас есть один рецензент и два сорта резюме "легкие" и "тяжелые" (см. пример выше). Рассмотрим функции вероятности для СВ X = "Кол-во рассмотренных резюме за день" в двух случаях:

- случай A = "на рецензию поступают только подробные (тяжелые) резюме"
- случай B = "на рецензию поступают только краткие (легкие) резюме".

Пусть мы накопили достаточно кол-во наблюдений и выяснили, что "тяжелые" резюме рецензент анализирует и пишет рецензию в среднем за два часа (4 за день), а легкие за 1 час (8 - за день). Воспользуемся законом Пуассона, чтобы описать распределение СВ X

Функция вероятности для $p_A(x)$

$$p_A(k)=\frac{4^k\cdot e^{-4}}{k!}$$

Функция вероятности $p_B(x)$

$$p_B(k)=\frac{8^k\cdot e^{-8}}{k!}$$

```
def poisson_drawing(lamda, xmax=20):
    plt.figure(figsize=(6, 3))
    ki = np.arange(xmax)
    poisson_pi = poisson.pmf(ki, lamda)
    plt.plot(ki, poisson_pi, '.', c='blue', label="poisson")
    for i in range(len(ki)):
        plt.plot([ki[i], ki[i]], [0, poisson_pi[i]], c='grey')
    plt.text(lamda + 3, max(poisson_pi) / 2, s=f"$\lambda$ = {lamda}")
    plt.suptitle(f"Рис. График функций вероятности для ДСВ 'кол-во рецензий за день'", y=0, fontsize=8)
```

```
        plt.xlim(-1, xmax)
        plt.xticks(np.arange(0, xmax, 2))
        plt.legend()
        plt.show()
```

```
poisson_drawing(lamda=4, xmax=20)
```

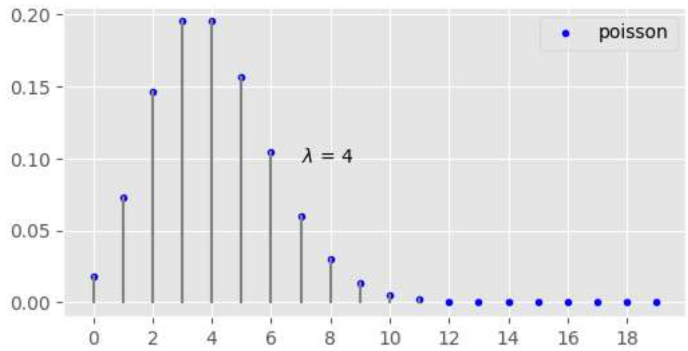


Рис. График функций вероятности для ДСВ 'кол-во рецензий за день'

```
poisson_drawing(lamda=8, xmax=20)
```

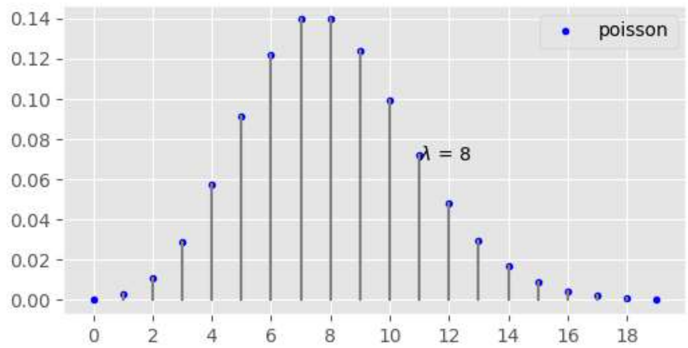


Рис. График функций вероятности для ДСВ 'кол-во рецензий за день'

> Нас будут интересовать ответы на следующие вопросы:
>
> - А как описать распределение СВ X в более общем случае, когда будут попадаться и легкие и тяжелые, если нам известны доля тяжелых и доля легких резюме?
> - можно ли использовать имеющиеся распределения для описания более общего случая?

Давайте смоделируем ситуацию случайного выбора заявки из очереди. Пусть у нас имеется СВ Y, которая принимает два значения:

- 0 - если заявка оказалась тяжелой;
- 1 - если заявка оказалась легкой

Пусть мы знаем, что доля легких заявок составляет 0.7. Тогда имеем следующее распределение Бернулли:

| $y_i$ | **0** | **1** |
|---|---|---|
| $p_i$ | 0.3 | 0.7 |

Наша гипотеза, что в итоге мы получим смесь этих двух распределений в пропорции, задаваемой этим распределением Бернулли

```
def mix_drawing(lamda1, lamda2, p1, xmax=20):
    plt.figure(figsize=(6, 3))
    ki = np.arange(xmax)
    p1_i = poisson.pmf(ki, lamda1)
    p2_i = poisson.pmf(ki, lamda2)
    pmix_i = p1 * p1_i + (1 - p1) * p2_i
    plt.plot(ki, p1_i, '-.', c='red', label=f"$\lambda$ = {lamda1}")
    plt.plot(ki, p2_i, '-.', c='blue', label=f"$\lambda$ = {lamda2}")
    plt.plot(ki, pmix_i, '-', c='green', label=f"mix p1={p1}")
    for i in range(len(ki)):
        plt.plot([ki[i], ki[i]], [0, pmix_i[i]], c='grey')
    plt.suptitle(f"Рис. Графики функций вероятности для ДСВ 'кол-во рецензий за день'", y=0, fontsize=8)
    plt.xlim(-1, xmax)
    plt.xticks(np.arange(0, xmax, 2))
    plt.legend()
    plt.show()
```

```
mix_drawing(lamda1=4, lamda2=8, p1=0.3, xmax=20)
```

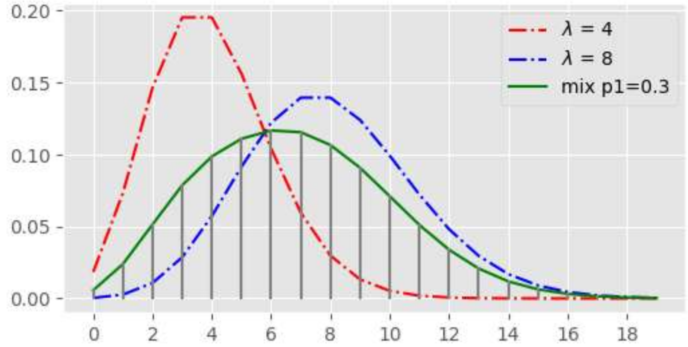


Рис. Графики функций вероятности для ДСВ 'кол-во рецензий за день'

В следующих уроках мы смоделируем эту ситуацию с использованием распределения времени рассмотрения рецензии и посмотрим насколько полученное в эксперименте распределение совпадает с нашей теоретической смесью.

## Резюме

- в результате арфиметических операций со СВ получается СВ;
- чтобы найти закон распределения для результата оперирования со СВ X и Y нужно построить закон их совместного распределения или распределения векторной СВ (X, Y):
$$p(x_i, y_j)$$
- если СВ X и Y независимы, то построить закон их совместного распределения легко, просто перемножая соответствующщие вероятности:
$$p(x_i, y_j) = p(X = x_i) \cdot p(Y = y_j)$$
- биномиальный закон описывает количество успехов в серии из n испытаний и определяется двумя параметрами n и p;
- СВ, распределенная по биномиальному законц получается в результате суммирования СВ - результатов испытаний Бернулли;
- биномиальный закон в пределе (при $n \to \infty$) дает закон Пуассона (для ДСВ) и нормальный закон распределения (для НСВ);
- на практике мы имеем чаще всего смесь известных законов, описывающих поведение СВ.