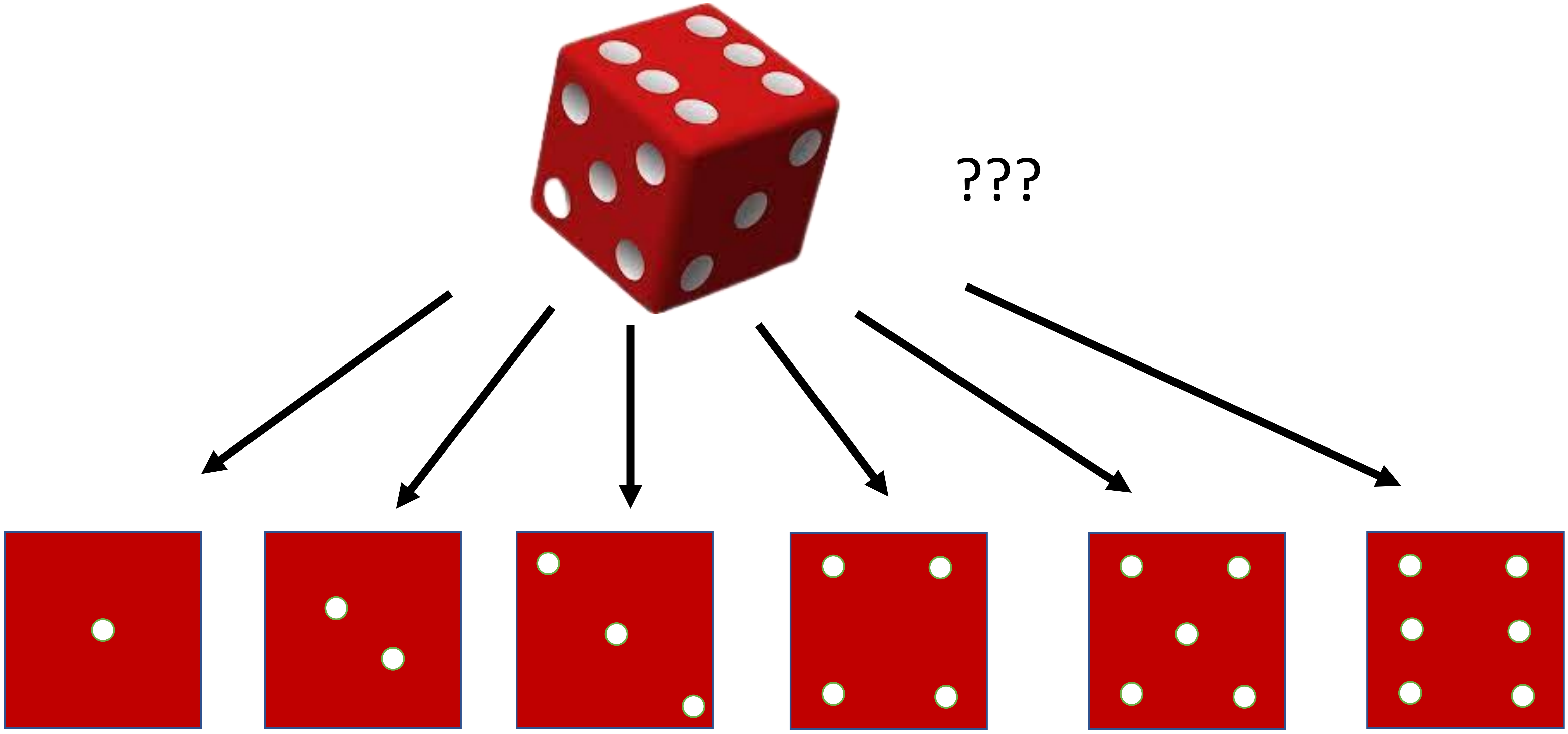
Разведывательный анализ данных с помощью языка R

Что такое случайность?

Что такое случайность?

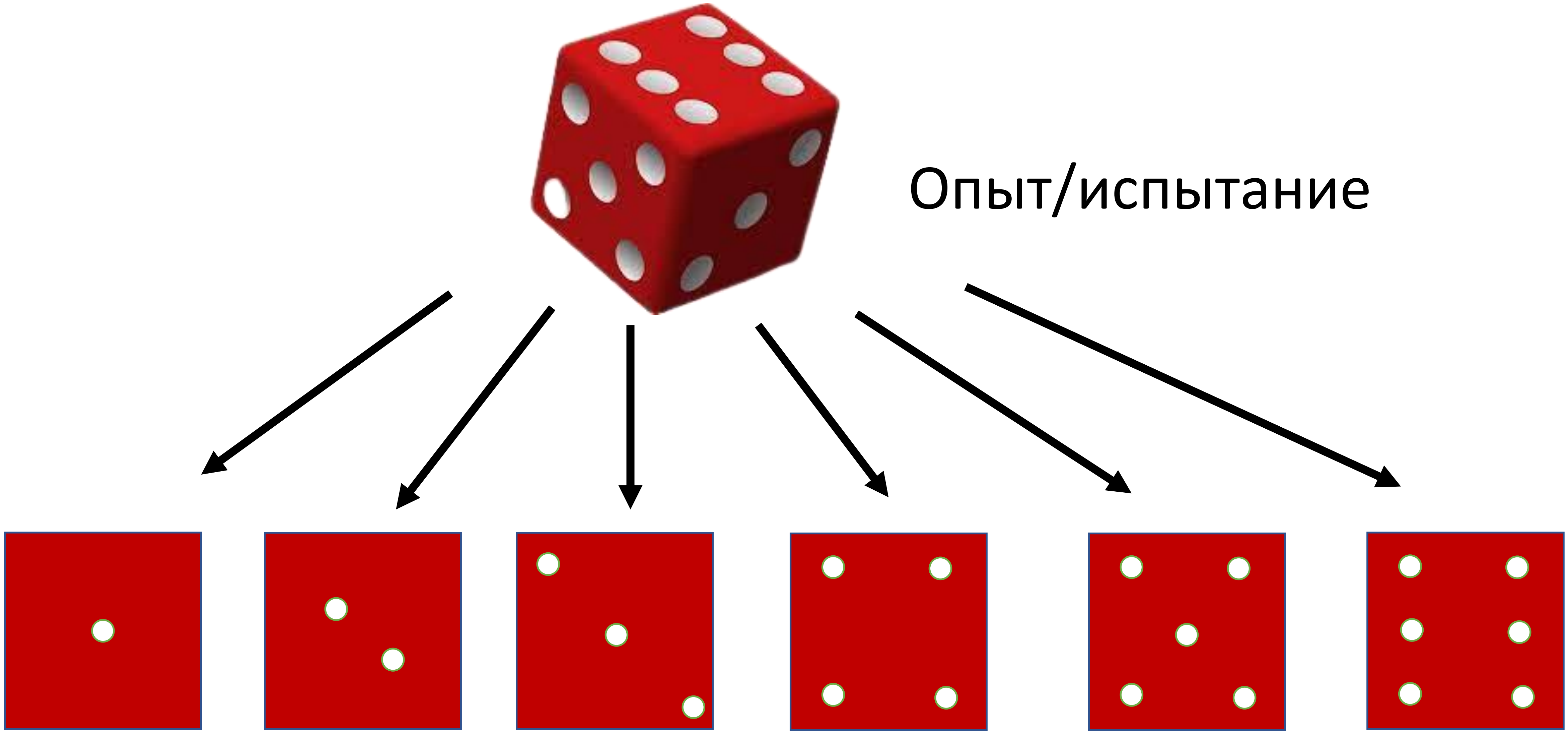
На самом деле определение дать очень непросто.

Что такое случайность?



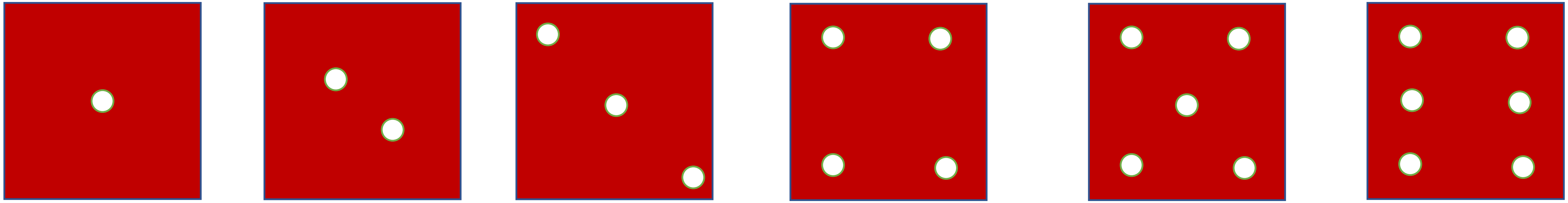
Неопределенность, недетерменированность результата нашего действия

Элементарные события



Элементарные события (исходы) – неделимы, может произойти только одно из них

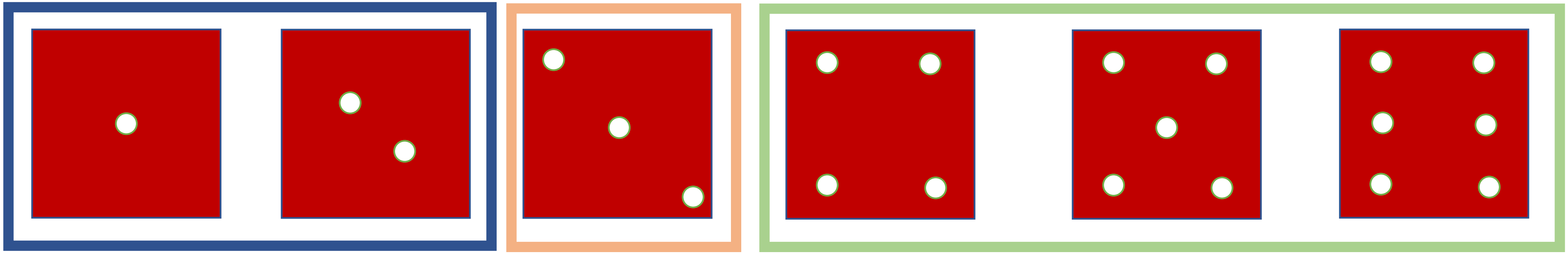
Пространство элементарных событий



Все элементарные события, которые могут в принципе произойти в ходе данного эксперимента

События

События – состоят из какого-то числа элементарных событий



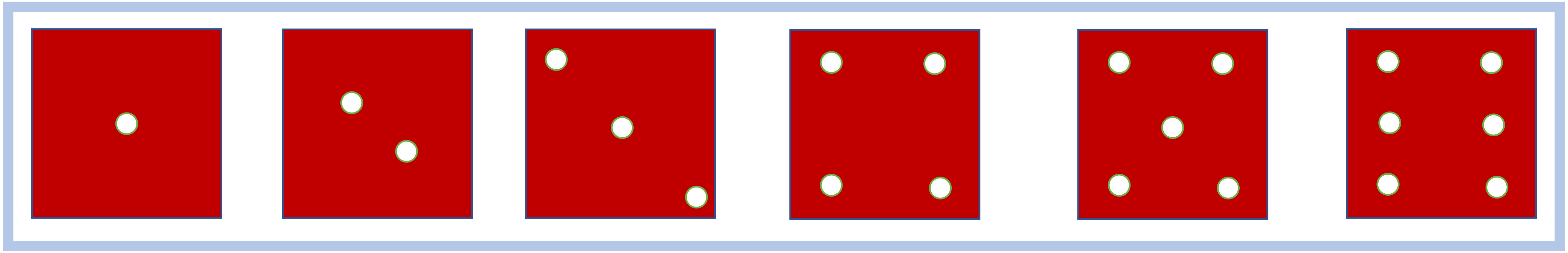
Событие A – выпало число меньше 3

Событие C – выпало число 3

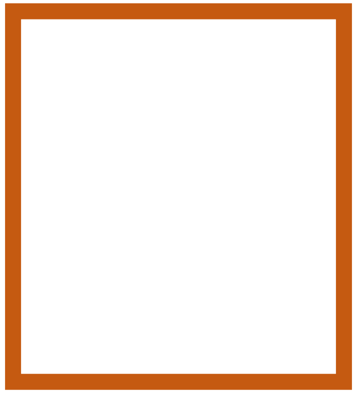
Событие B – выпало число больше 3

События

События – состоят из какого-то числа элементарных событий.
Могут состоять из 0 или всех элементарных событий



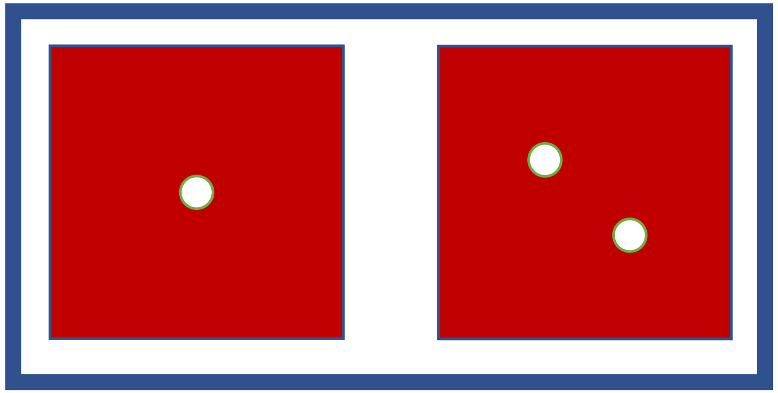
Событие D –
выпало какое-то
целое число



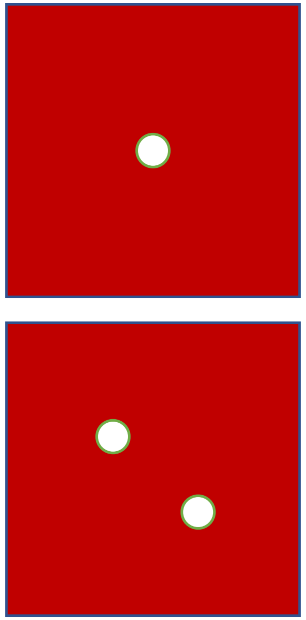
Событие E –
выпало число 7

Благоприятвующие исходы

Событиия – состоят из какого-то числа элементарных событий



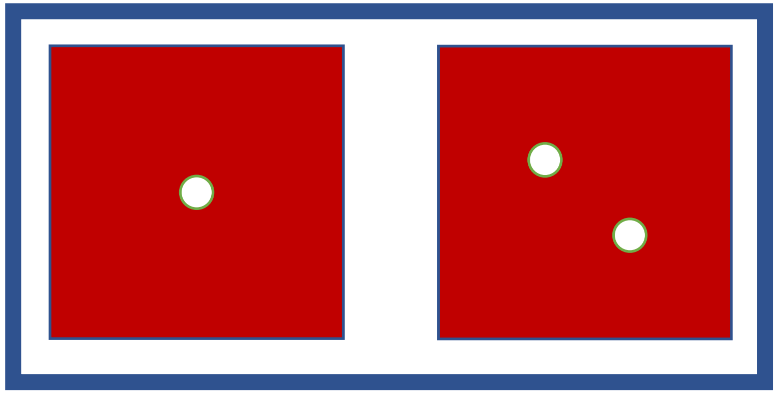
Событие А – выпало число меньше 3



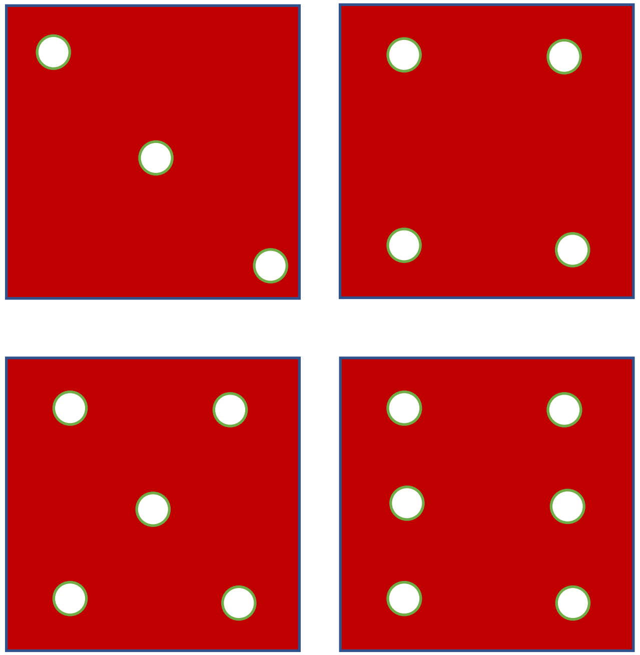
Если выпало эти две грани – событие А случилось. Эти исходы **благоприятствуют** нашему событию

Благоприятвующие исходы

События – состоят из какого-то числа элементарных событий



Событие А – выпало число меньше 3

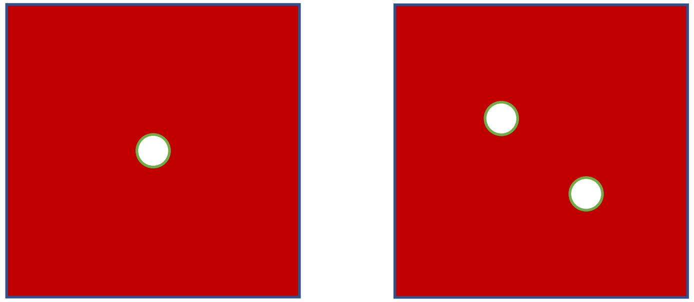


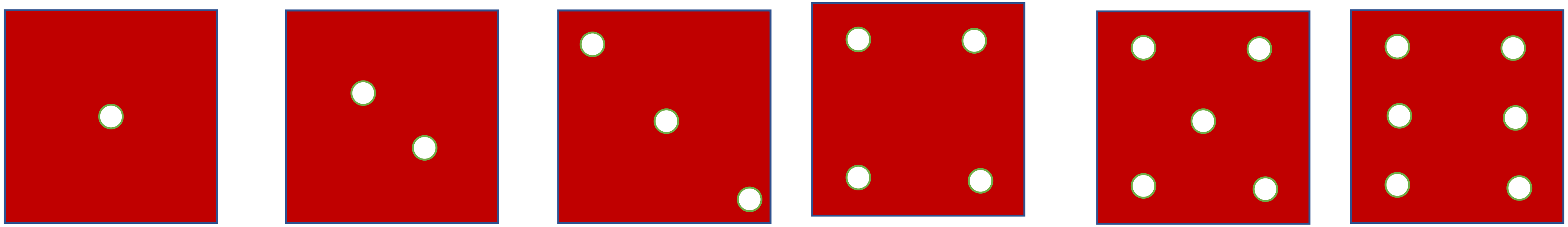
Если выпало эти четыре грани – событие А не случилось. Эти исходы **не благоприятствуют** нашему событию

Вероятность события

Какую долю всех элементарных событий включает событие А

P(Событие А – выпало число меньше 3) =

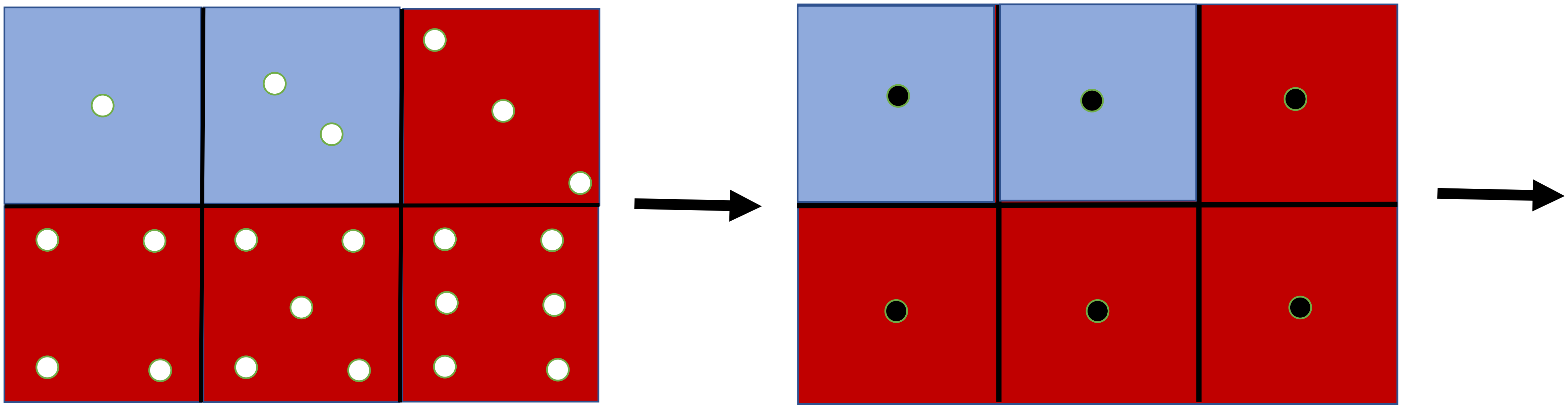
#()

#()

Вероятность события

Какую долю всех элементарных событий включает событие А

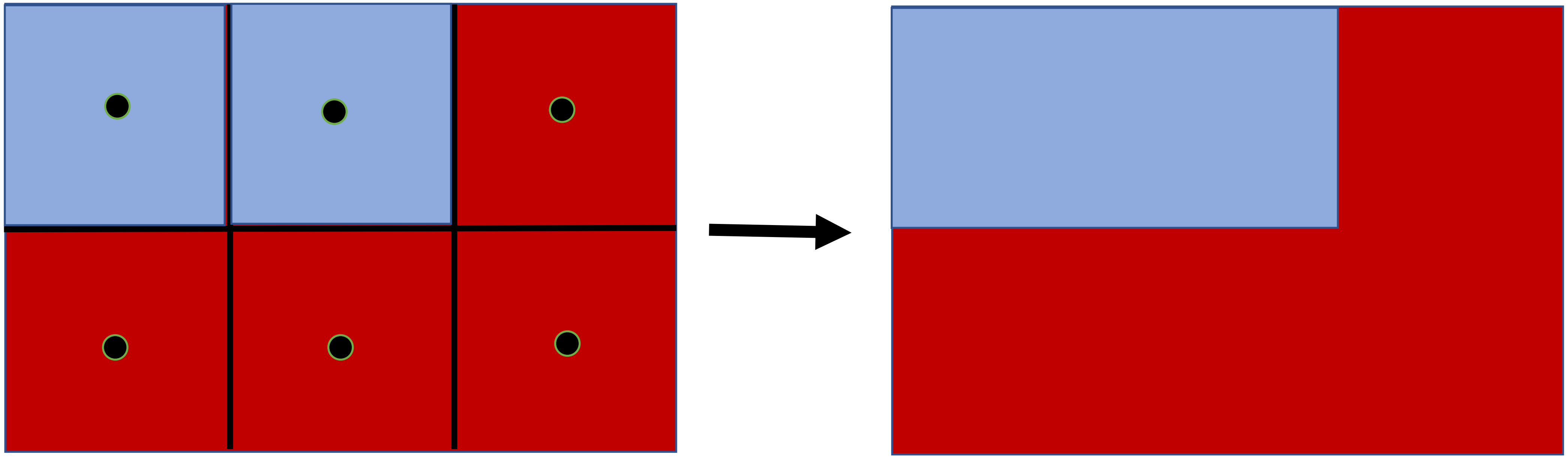
P(Событие А – выпало число меньше 3)



Вероятность события

Какую долю всех элементарных событий включает событие А

P(Событие А – выпало число меньше 3)

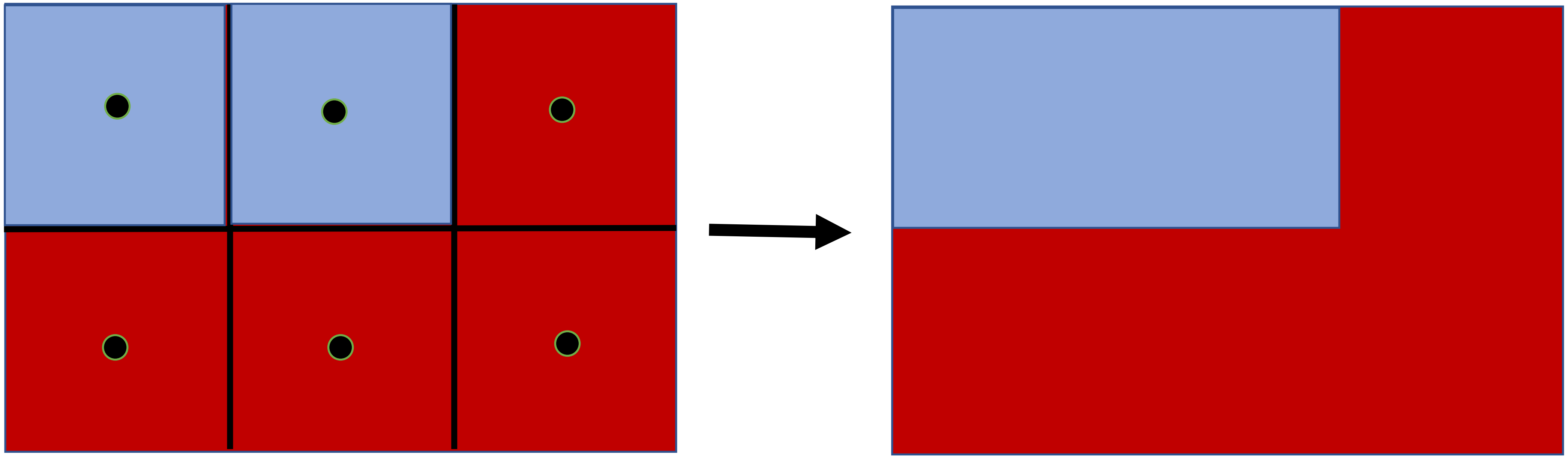


Вероятность события

Какую долю всех элементарных событий включает событие А

P(Событие А – выпало число меньше 3)

Доля площади, «занимаемой» событием А



Вероятность события

Бросаем случайно точку на прямую – какова вероятность, что попадет в первую треть?

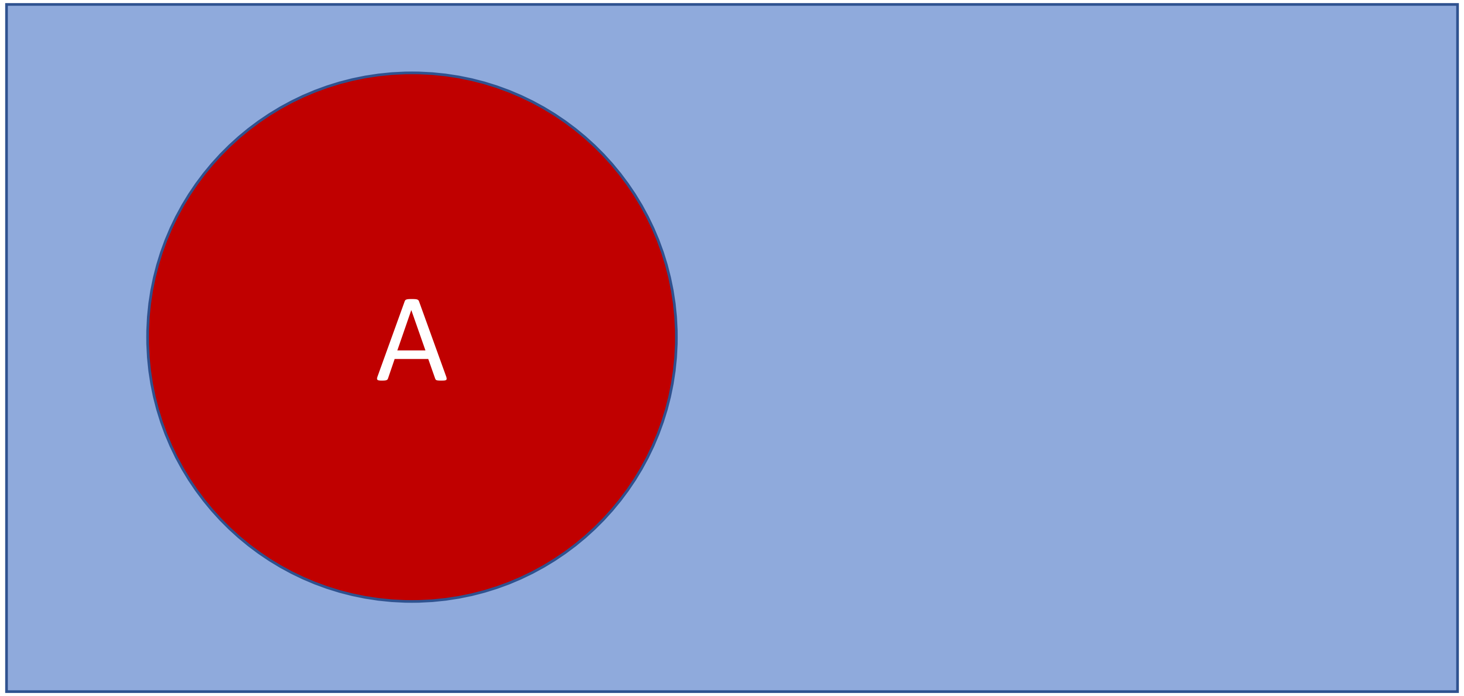
Вероятность события

Бросаем случайно точку на прямую – какова вероятность, что попадет в первую треть?

P(A) = 1/3

Вероятность события

Бросаем случайно точку в прямоугольник, какова вероятность, что попадем в окружность?



Вероятность события

Бросаем случайно точку на прямую – какова вероятность, что попадем в заданную точку

Вероятность события

Бросаем случайно точку на прямую – какова вероятность, что попадем в заданную точку?

Первый вариант ответа (теоретик) – 0. Если мы предполагаем, что умеем различать мельчайшие отклонения в положении точки – то вероятность попасть в нужное место – 0. Исходов, благоприятсвующих нашему исходу – 1. Исходов неблагоприятствующих – бесконечно много

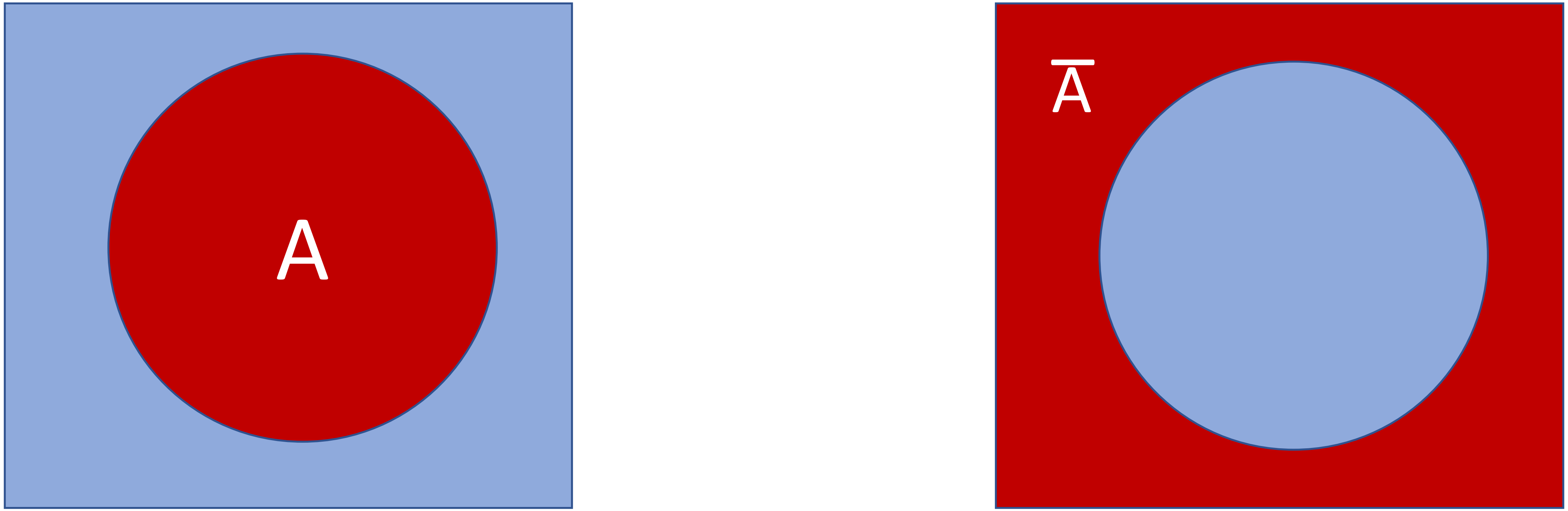
Вероятность события

Бросаем случайно точку на прямую – какова вероятность, что попадем в заданную точку?

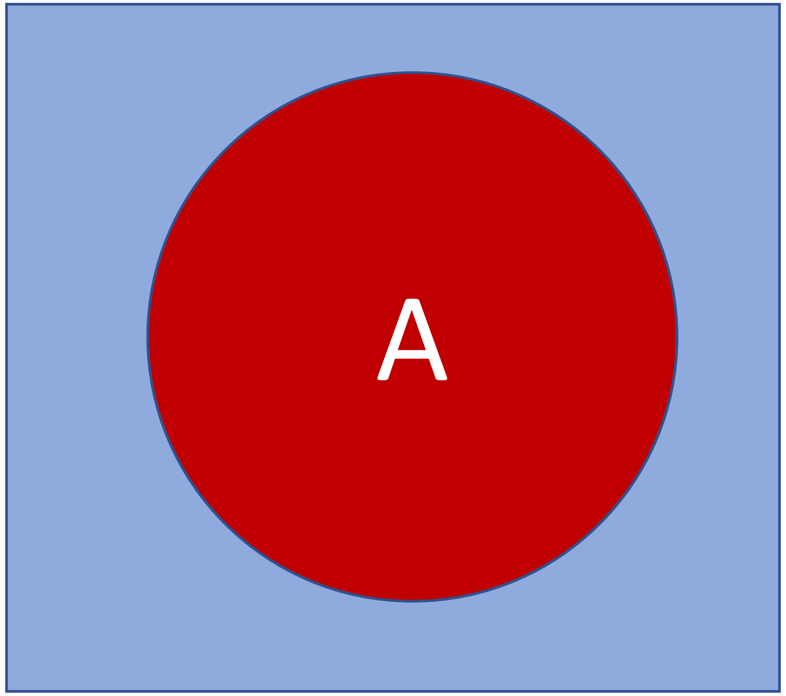
Второй вариант ответа (экспериментатор) – зависит от точности измерений. Если мы умеем различать только 10000 положений точки, то благоприятсвует нашему событию 1 исход.

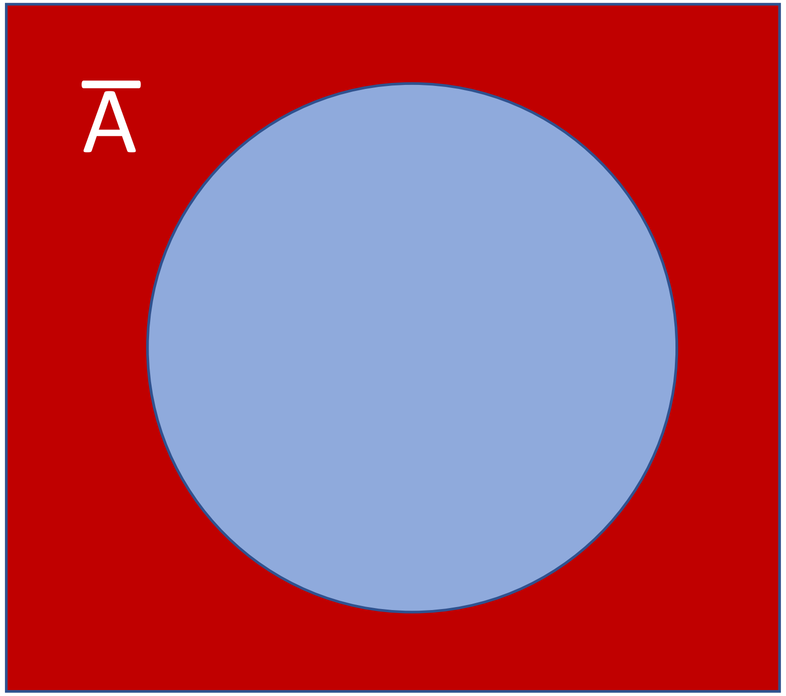
$$P(A) = 1/10000$$

Противоположное событие (отрицание)



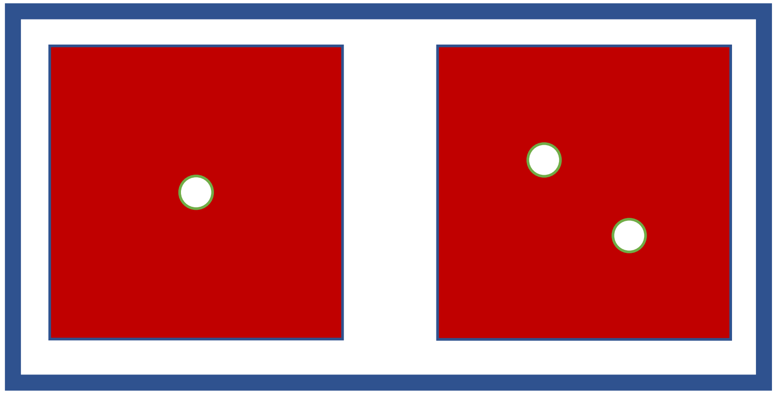
Вероятность отрицания



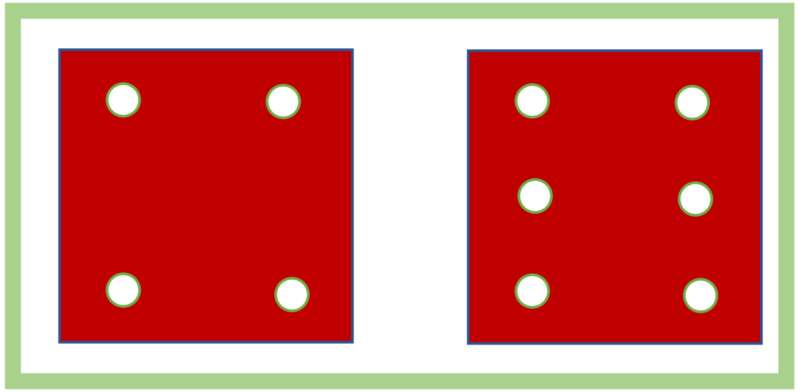


$$P(\bar{A}) = 1 - \mathrm{P}(A)$$

Несовместные события

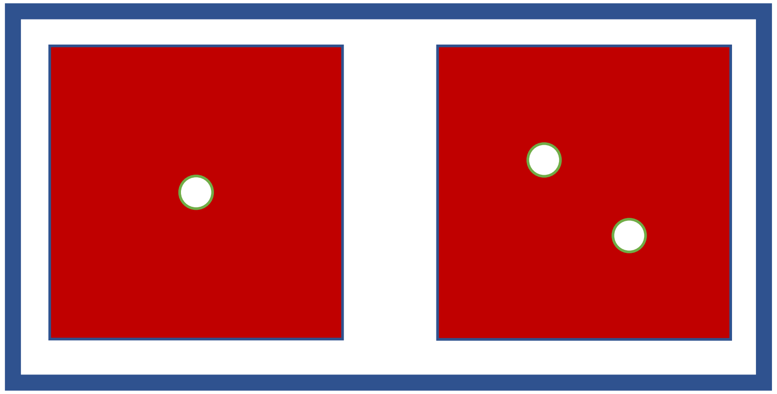


Событие А – выпало число меньше 3

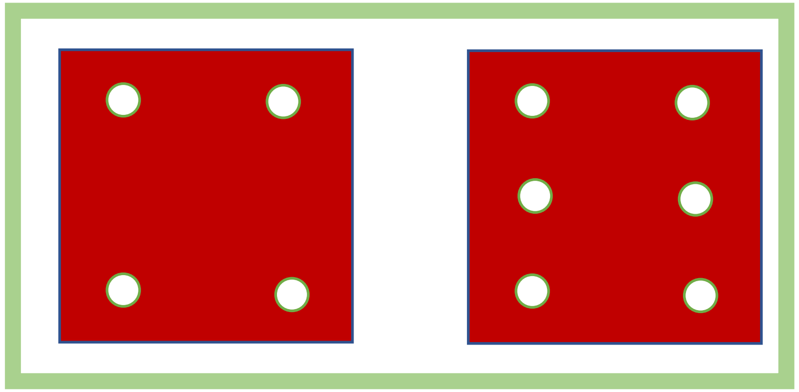


Событие В – выпало четное число, большее 3

Несовместные события



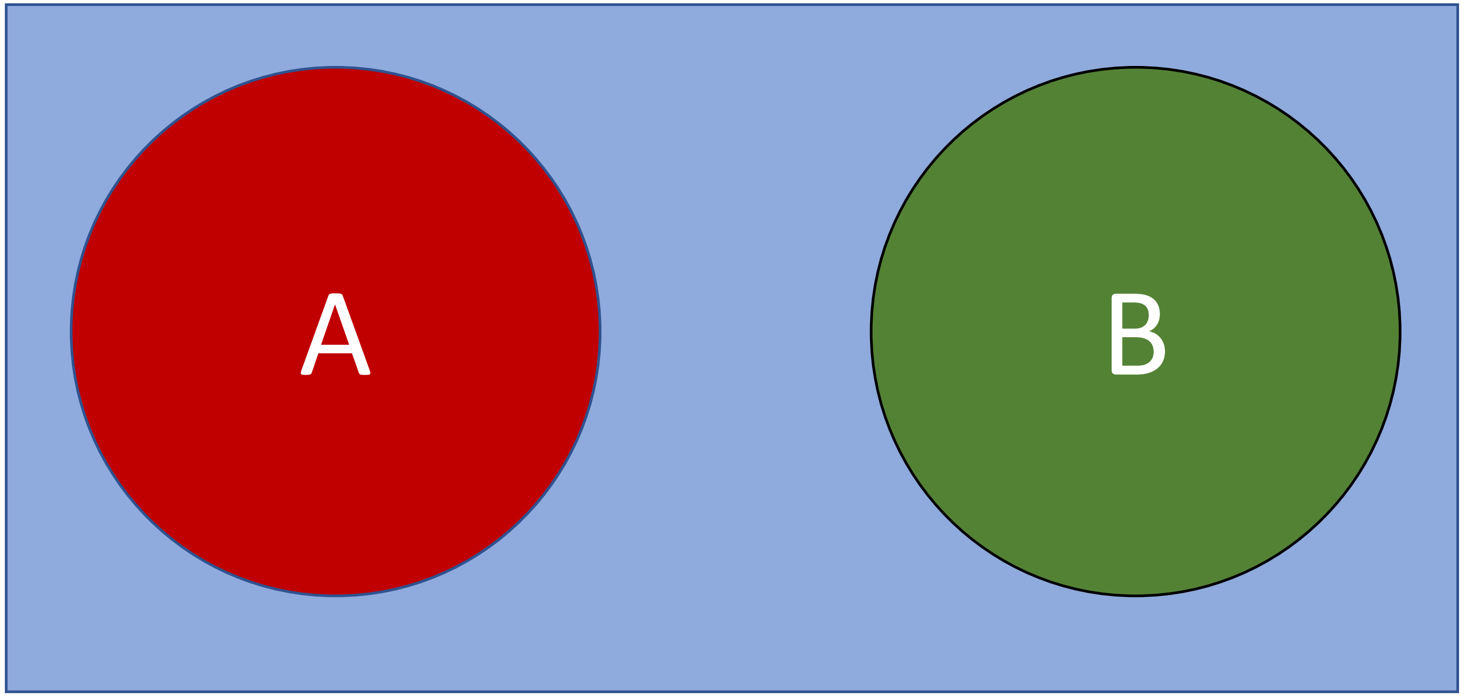
Событие А –
выпало число
меньше 3



Событие В –
выпало четное
число, большее 3

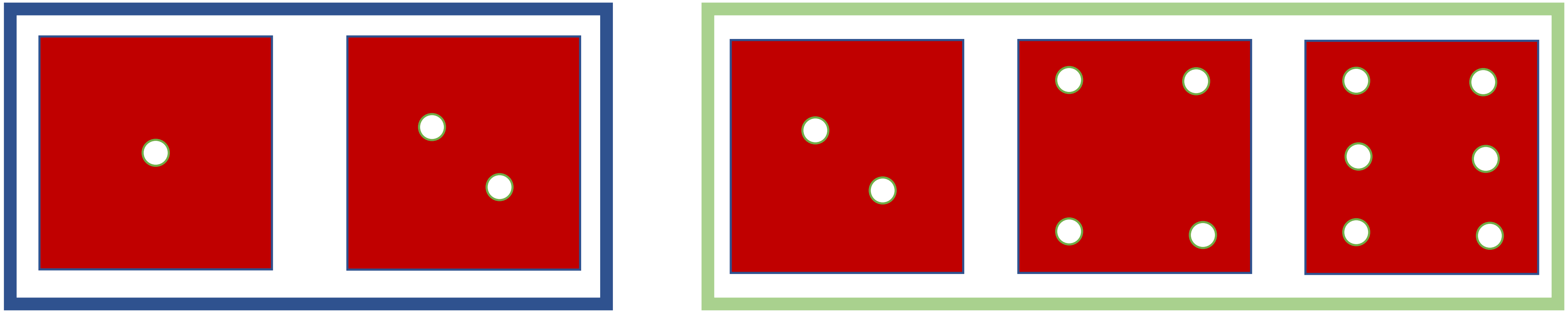
Несовместные события – это события, которые **не могут** наступить одновременно в ходе одного испытания

Несовместные события



Несовместные события – это события, которые **не могут** наступить одновременно в ходе одного испытания

Совместные события

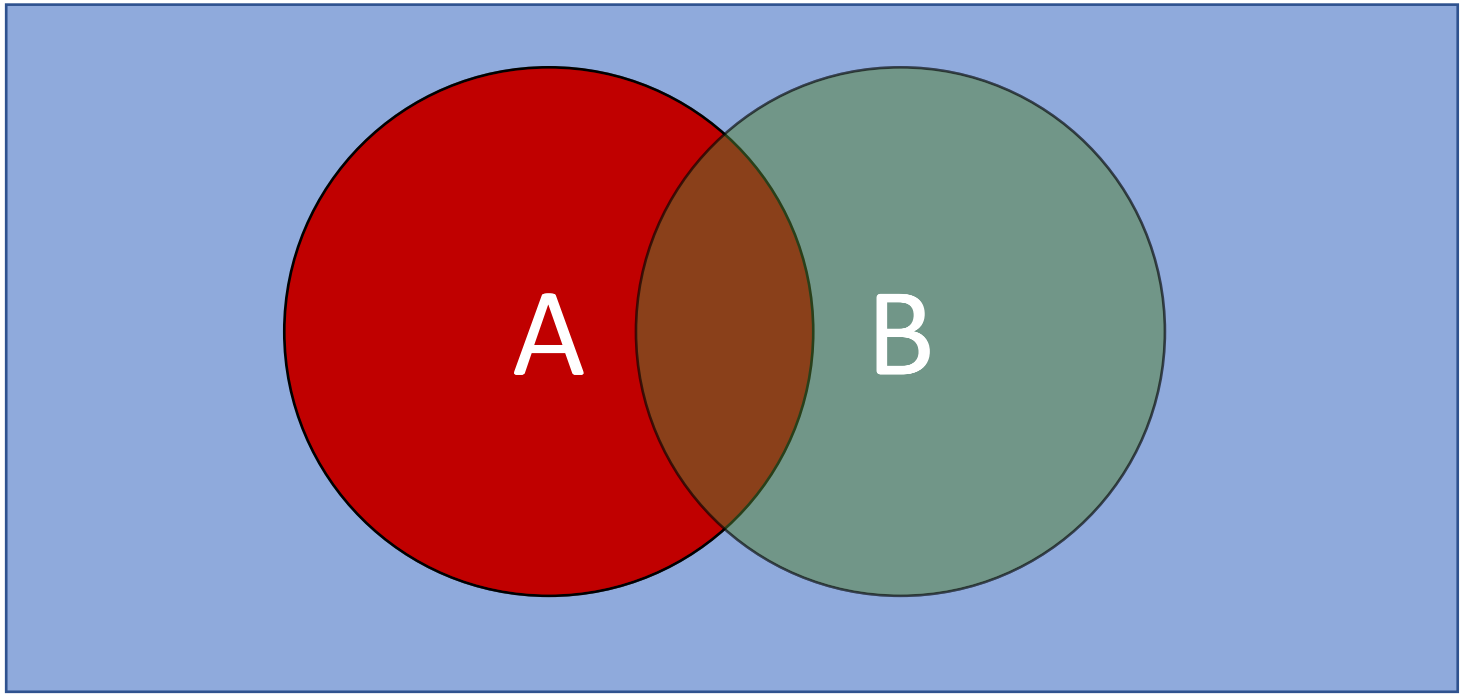


Событие А – выпало число меньше 3

Событие В – выпало четное число

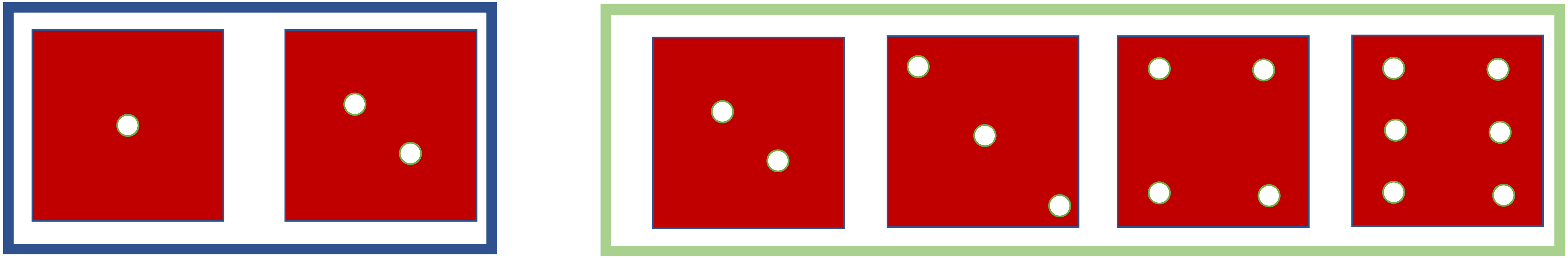
Совместные события – это события, которые **могут** наступить одновременно в ходе одного испытания

Совместные события



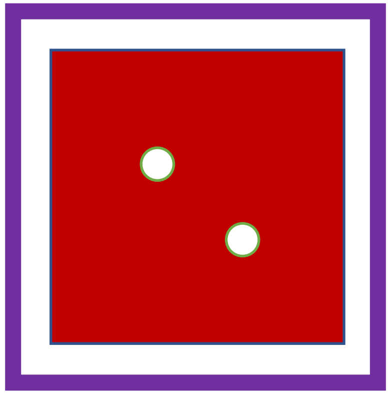
Совместные события – это события, которые **могут** наступить одновременно в ходе одного испытания

Пересечение событий



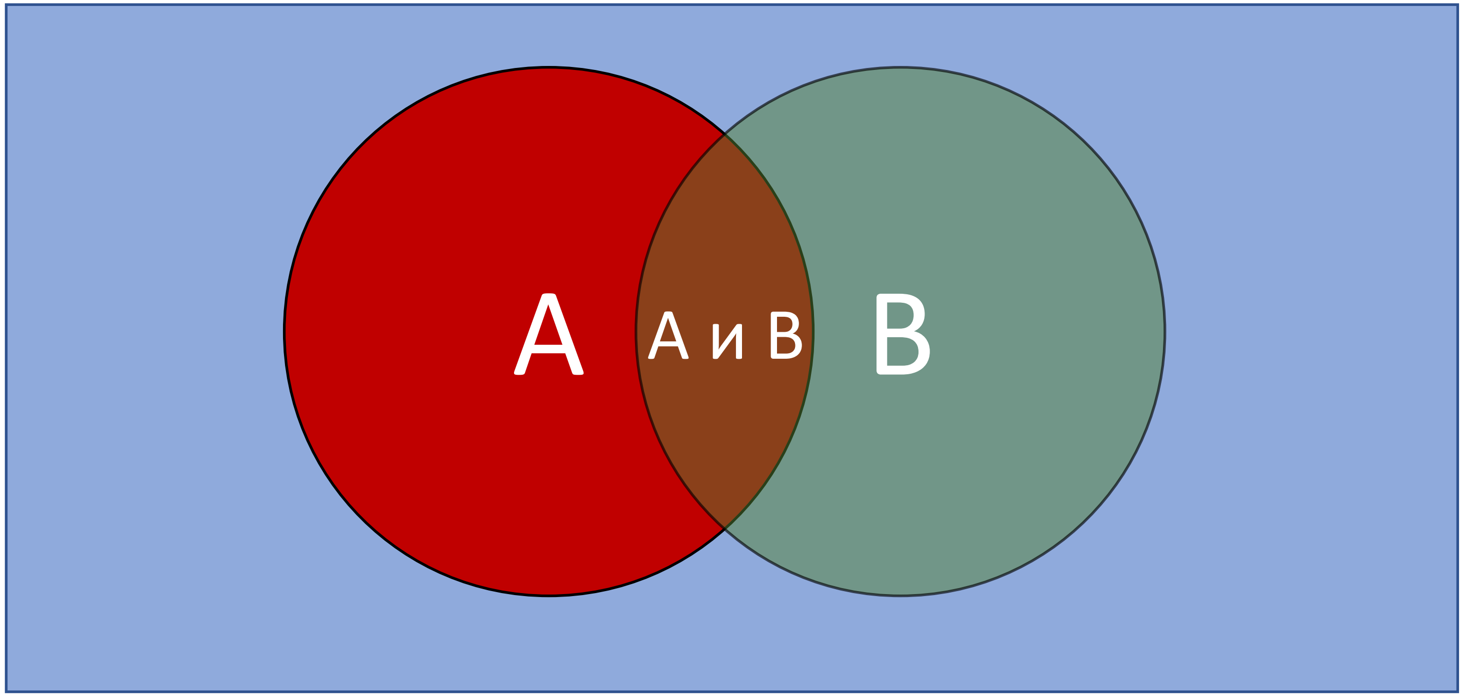
Событие А

Событие В



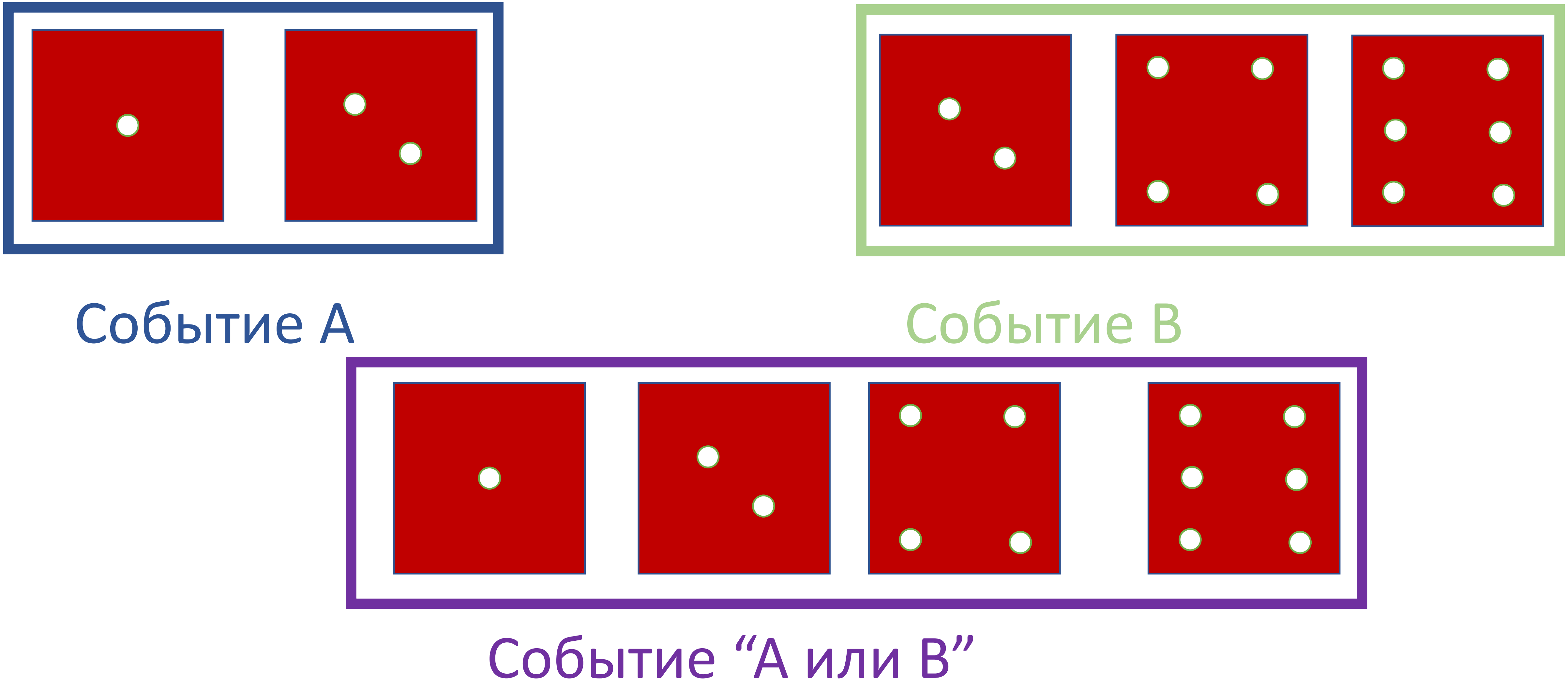
Событие “А и В”

Пересечение событий

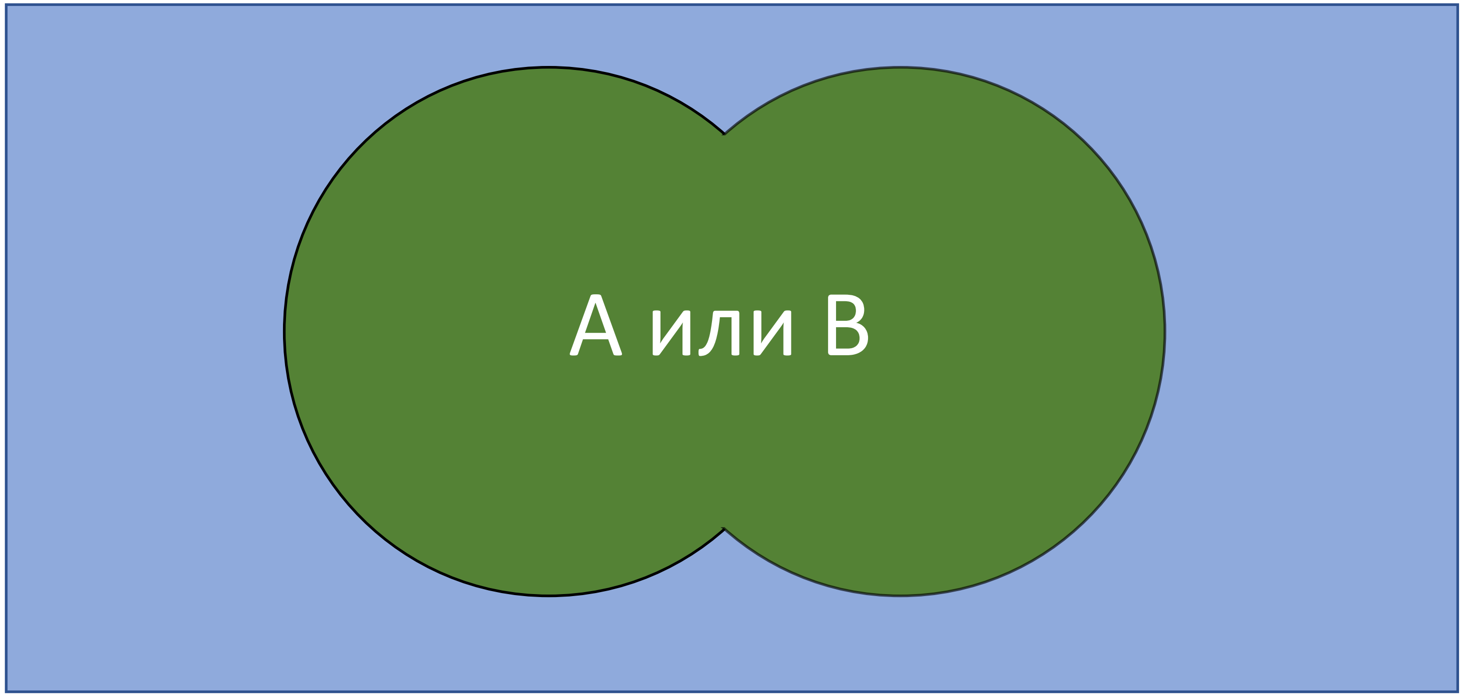


Пересечение событий A и B – событие, элементарные события которого благоприятствуют как первому, так и второму событиям

Объединение событий



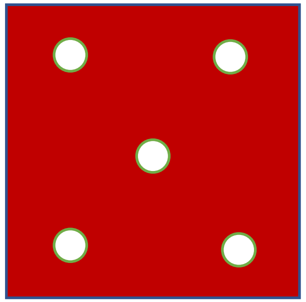
Объединение событий

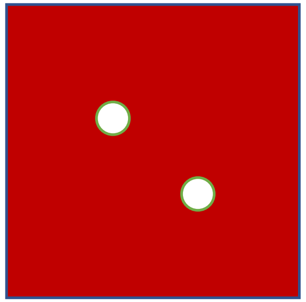


Объединение событий А и В – событие, элементарные события которого благоприятствуют или первому, или второму событиям

Независимые события

У нас есть два кубика



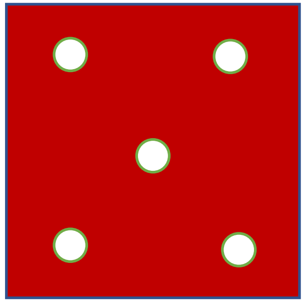


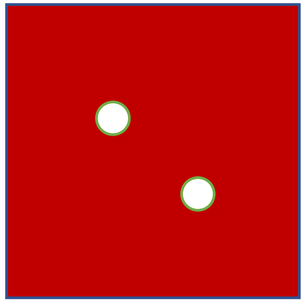
Событие А – на первом кубике выпало 5

Событие В – на втором кубике выпало 5

Независимые события

То, что произошло одно событие – никак не влияет на вероятность того, что произойдет второе





Событие А – на первом кубике выпало 5

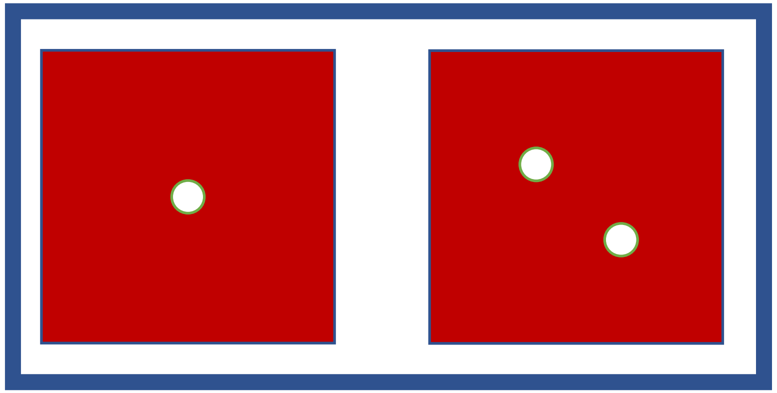
Событие В – на втором кубике выпало 2

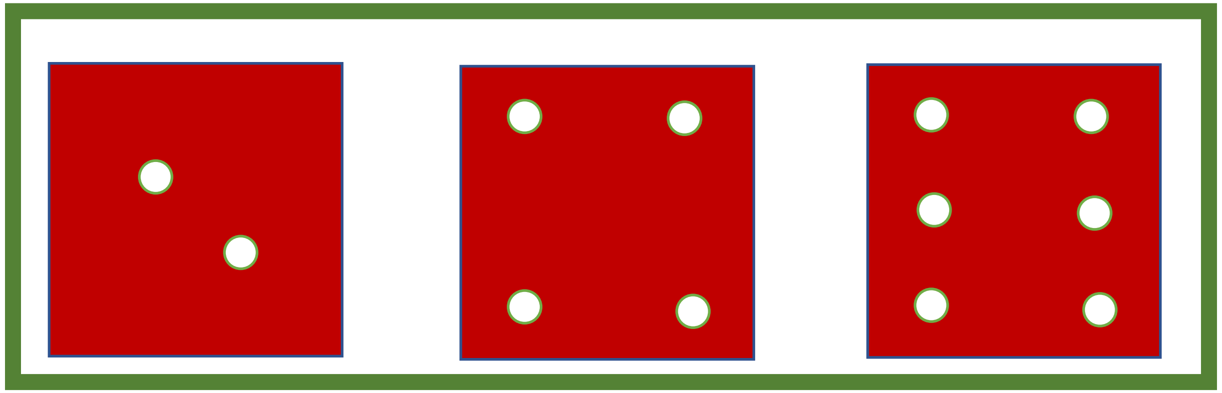
Независимые события

???

Не отобразить

Независимые события





Событие А – на кубике выпало число меньшее трех

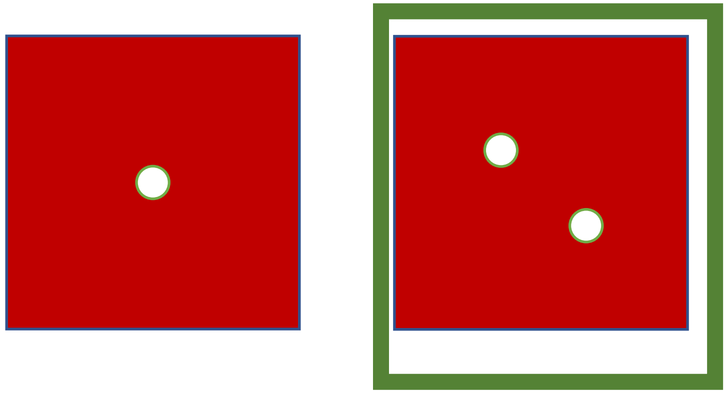
Событие В – на кубике выпало четное число

P(A) = 1/3

P(B) = 1/2

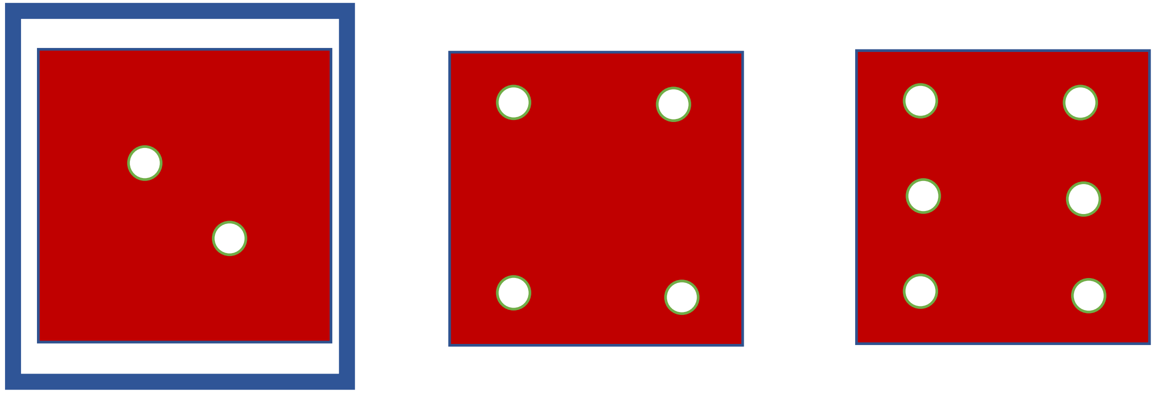
Независимые события

Если произошло A



$P(B) = 1/2$

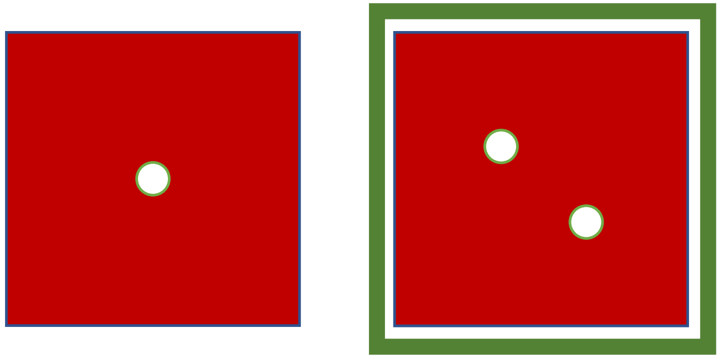
Если произошло B



$P(A) = 1/3$

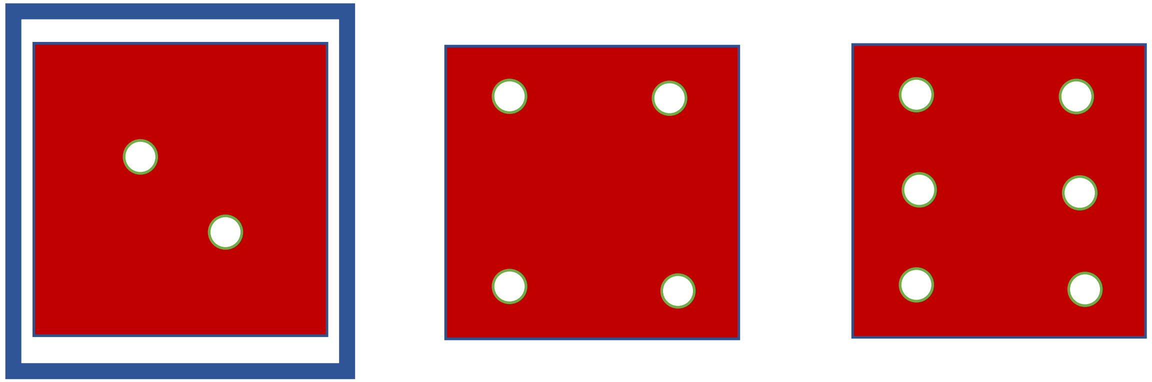
Условная вероятность

Если произошло A



P(B|A) = 1/2

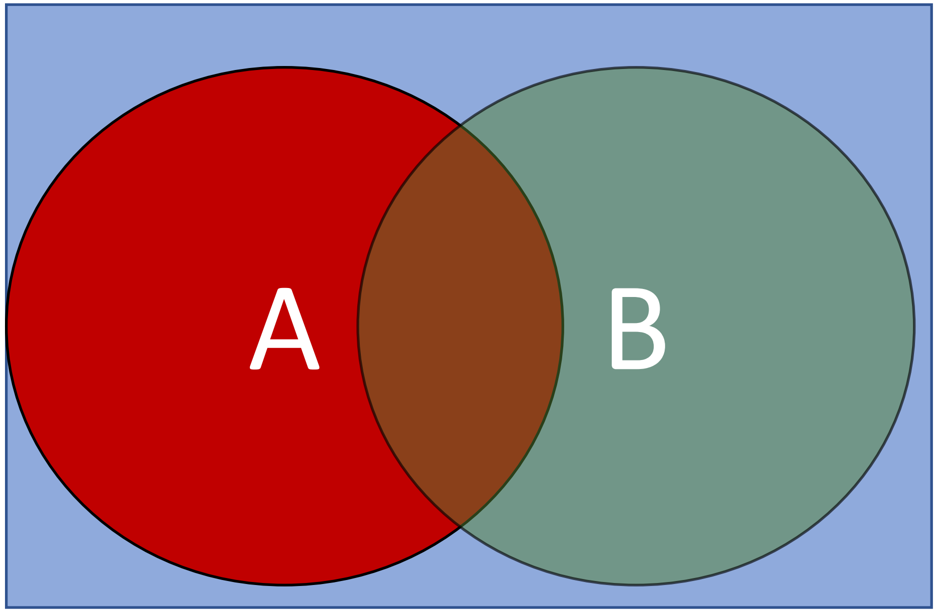
Если произошло B



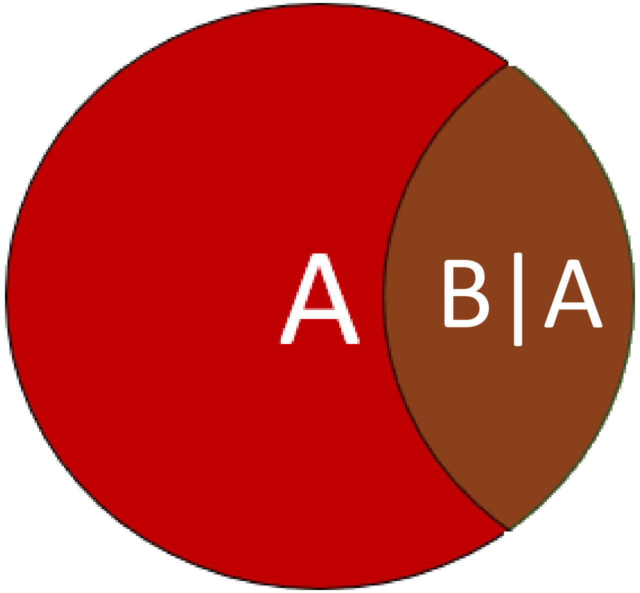
P(A|B) = 1/3

Условная вероятность - вероятность события при условии того, что произошло какое-то другое событие

Условная вероятность



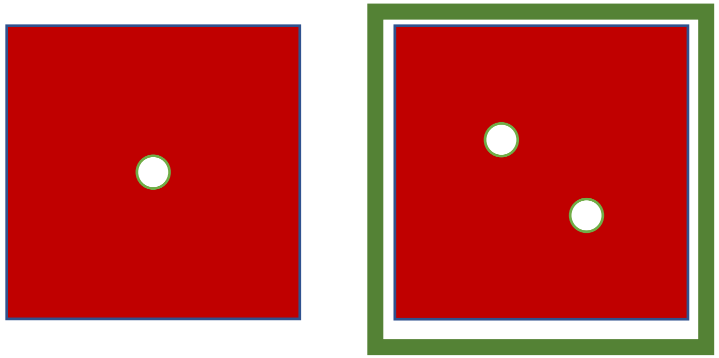
Откуда-то знаем, что произошло A



Условная вероятность - вероятность события при условии того, что произошло какое-то другое событие

Независимые события

Если произошло А

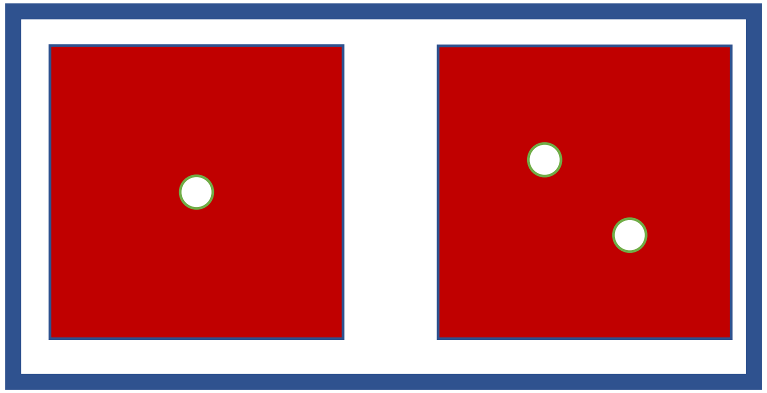


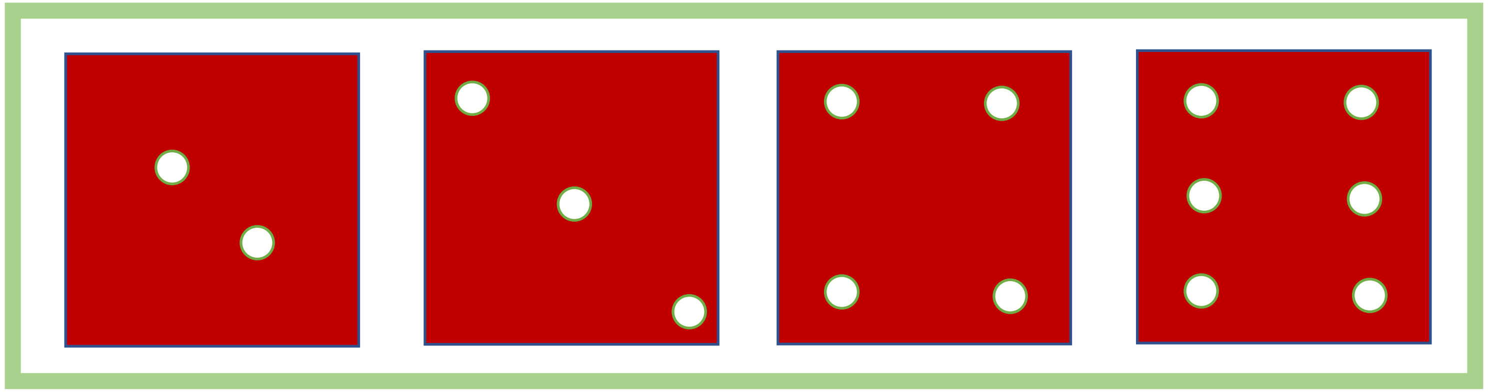
P(B|A) = 1/2

Событие А независимо от события В, если

P(A|B) = P(A)

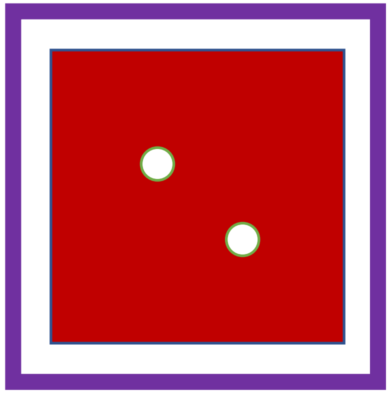
Вероятность пересечения





P(A) = 1/3

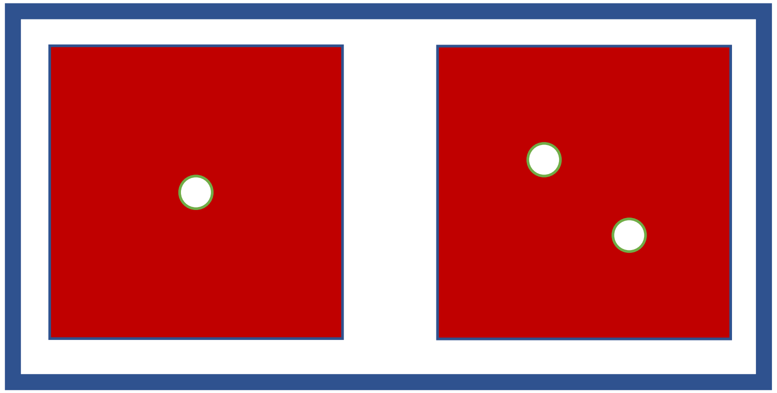
P(B) = 2/3



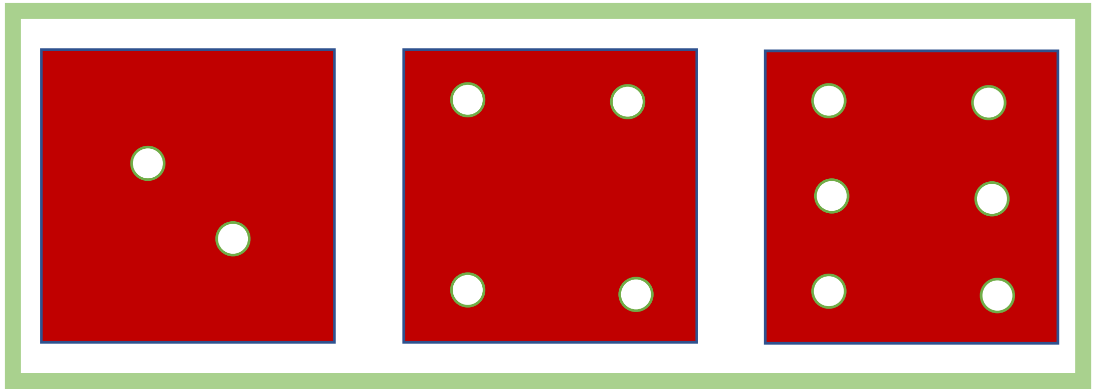
P(A и B) = 1/6

$$P(\text{"}A\text{ и }B\text{"}) = P(A)P(A|B) = P(B)P(B|A)$$

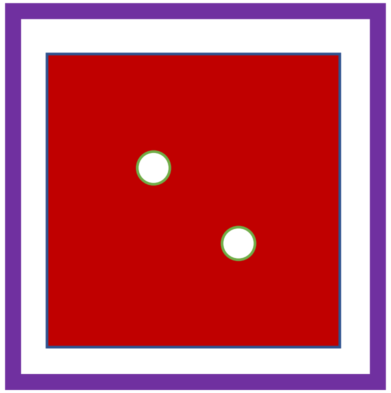
Вероятность пересечения



P(A) = 1/3



P(A) = 1/2



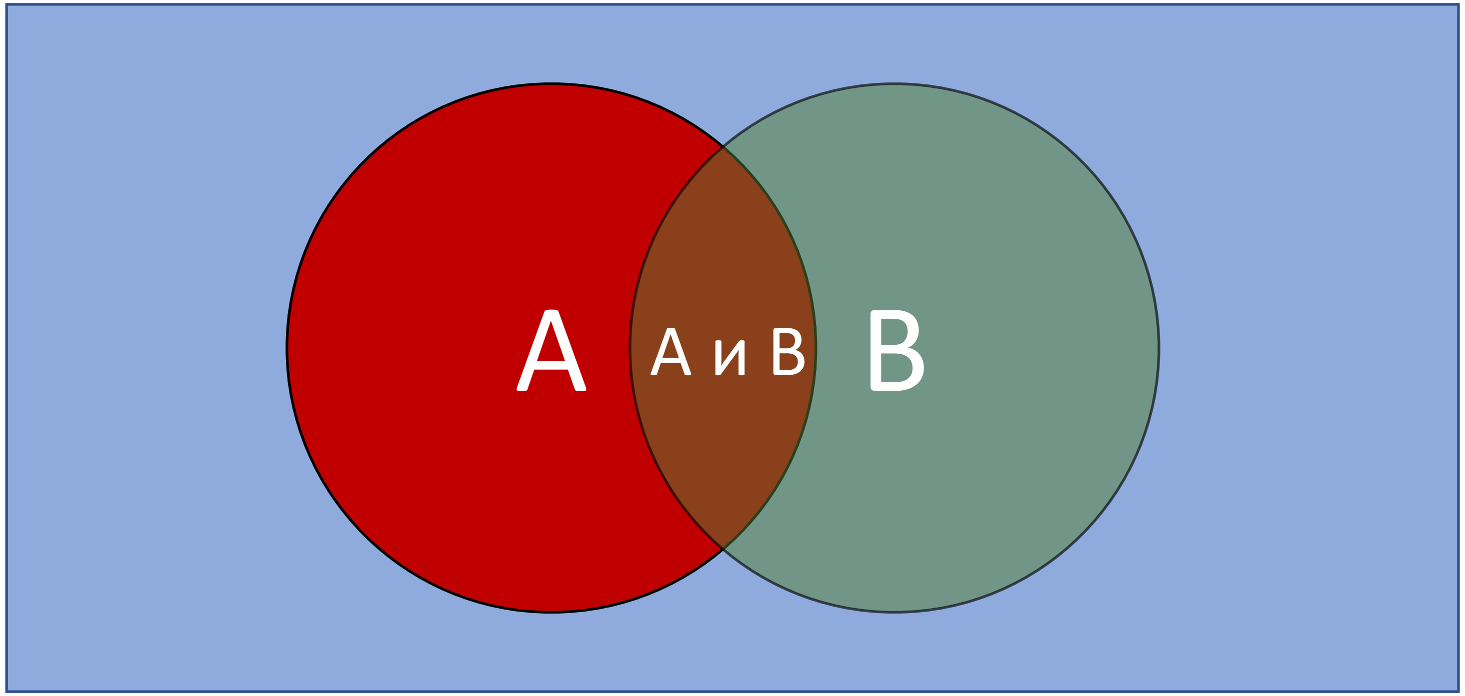
P(A и B) = 1/6

$$P(\text{"}A \text{ и } B\text{"}) = P(A)P(A|B) = P(B)P(B|A)$$

Если события независимы, то

$$P(\text{"}A \text{ и } B\text{"}) = P(A)P(B)$$

Вероятность пересечения

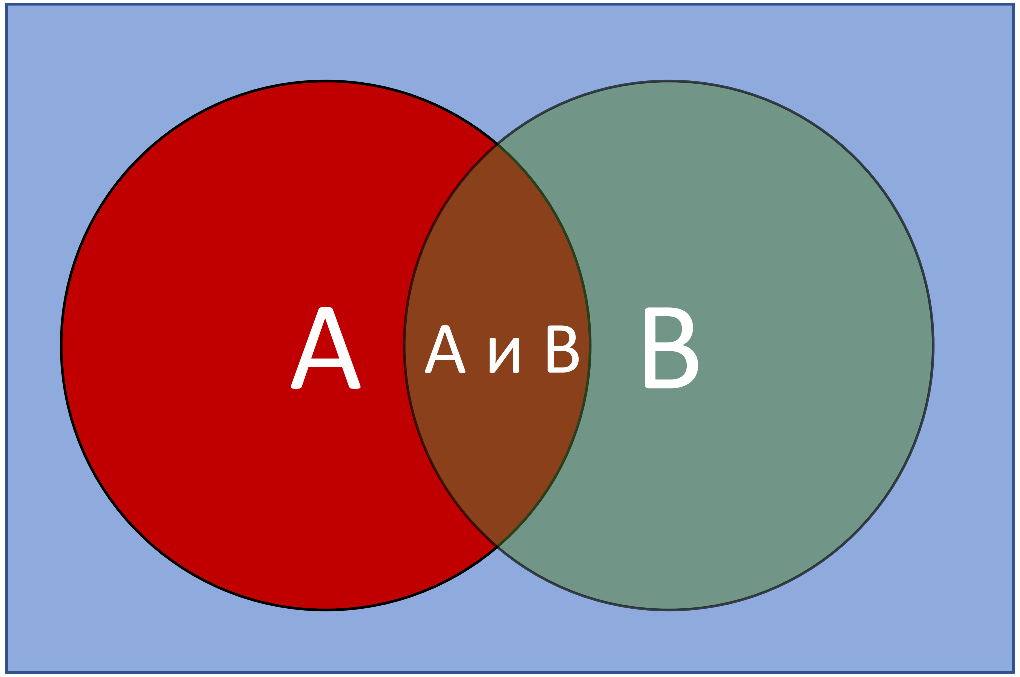


$$P(\text{"}A \text{ и } B\text{"}) = P(A)P(A|B) = P(B)P(B|A)$$

Если события независимы, то

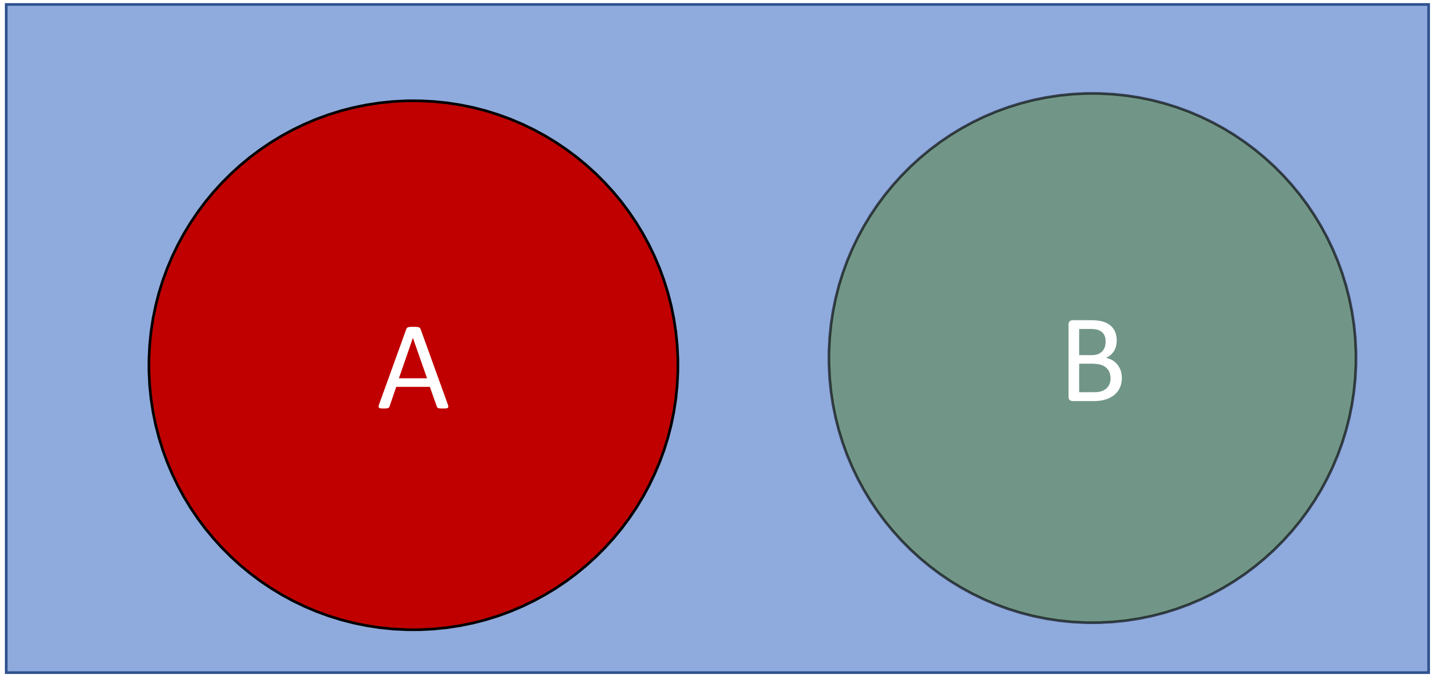
$$P(\text{"}A \text{ и } B\text{"}) = P(A)P(B)$$

Вероятность объединения



$$P(\text{"}A\text{ или }B\text{"}) = P(A) + P(B) - P(AB)$$

Вероятность объединения

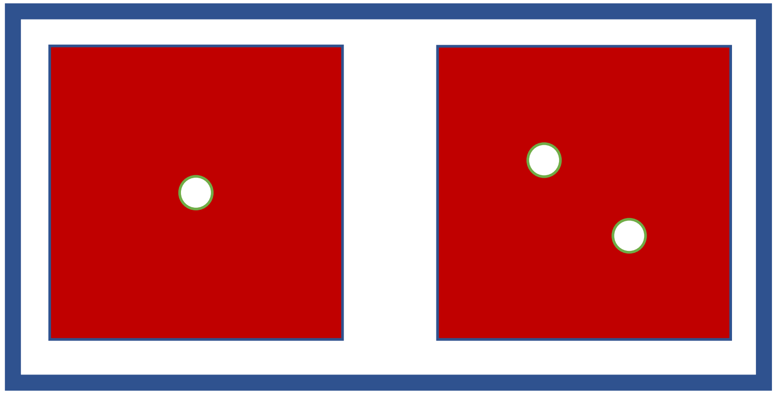


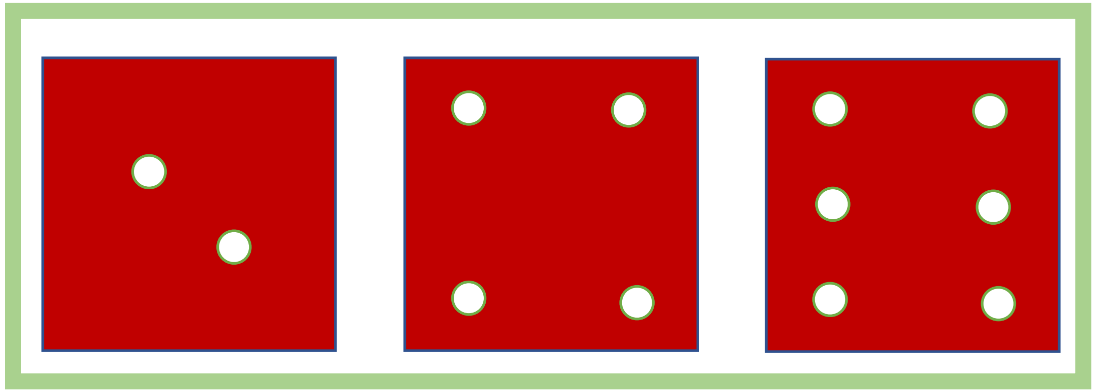
$$P(\text{"}A \text{ или } B\text{"}) = P(A) + P(B) - P(AB)$$

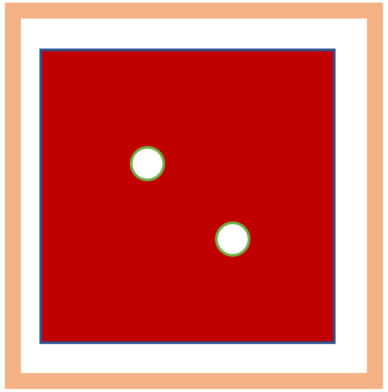
Если события несовместны, то

$$P(\text{"}A \text{ или } B\text{"}) = P(A) + P(B)$$

Вероятность объединения



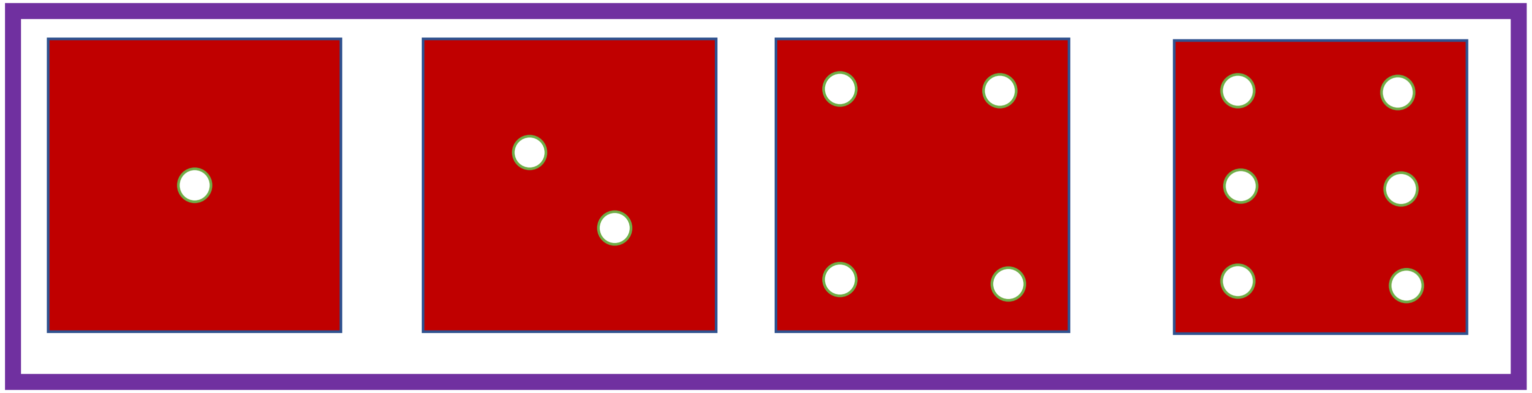




P(A) = 1/3

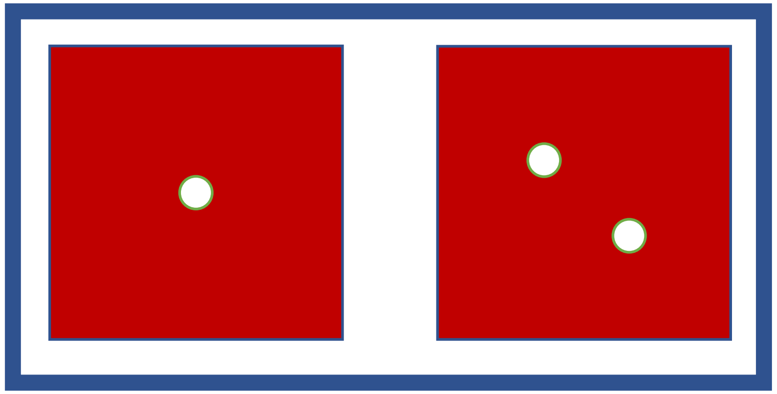
P(B) = 1/2

P(A и B) = 1/6

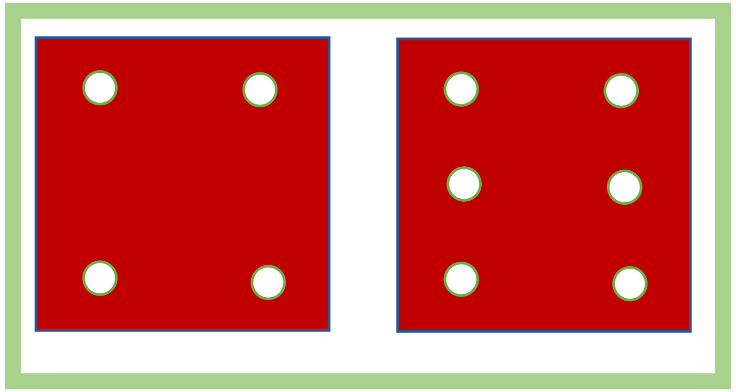


P("A или B") = 1/3 + 1/2 – 1/6 = 2/3

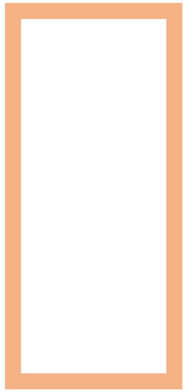
Вероятность объединения



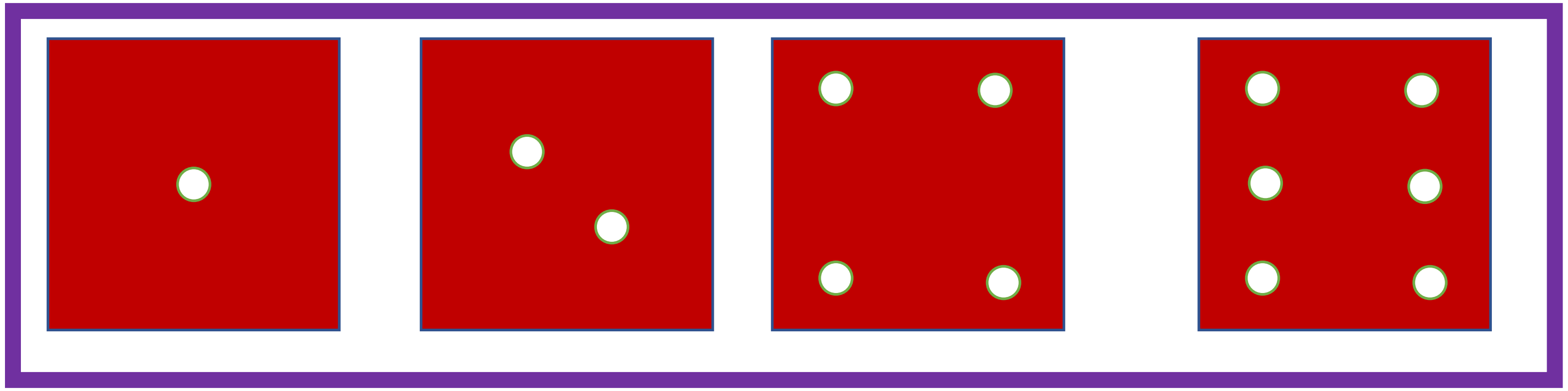
P(A) = 1/3



P(B) = 1/3



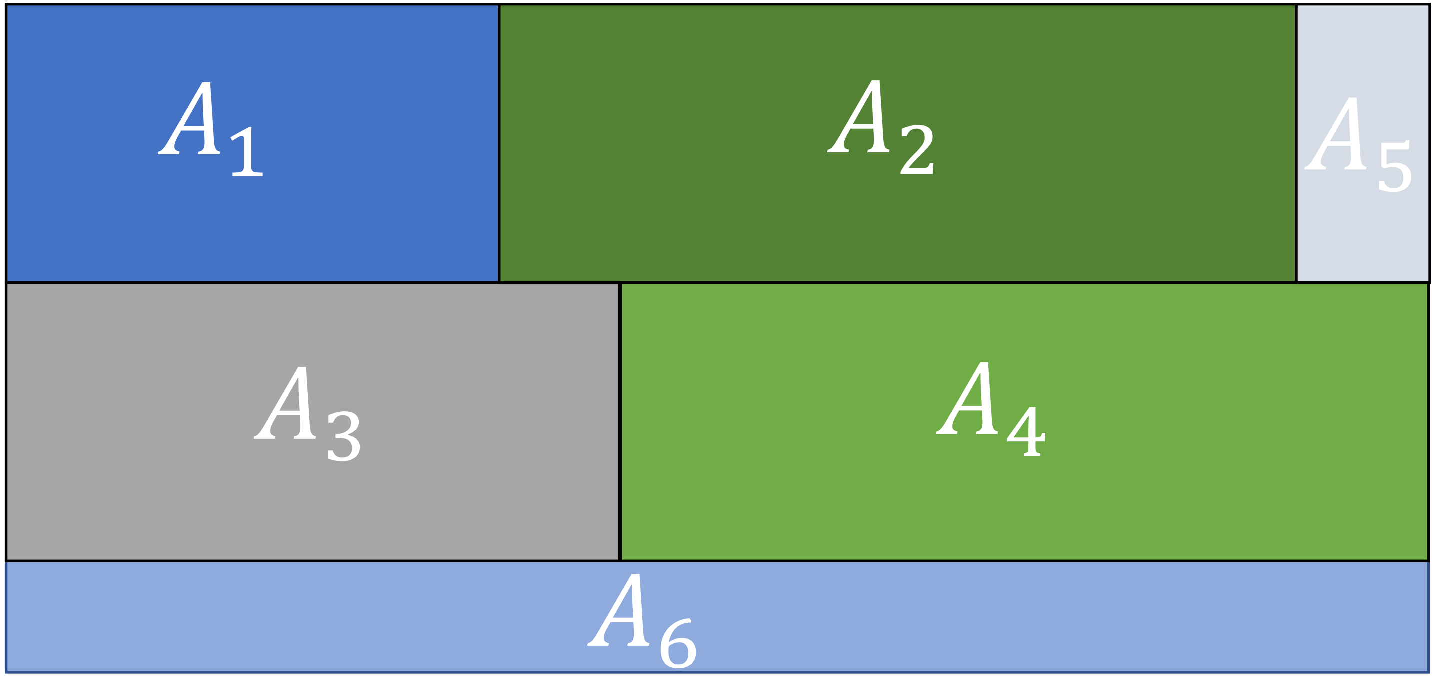
P(A и B) = 0



P(“A или B”) = 1/3 + 1/3 – 0 = 2/3

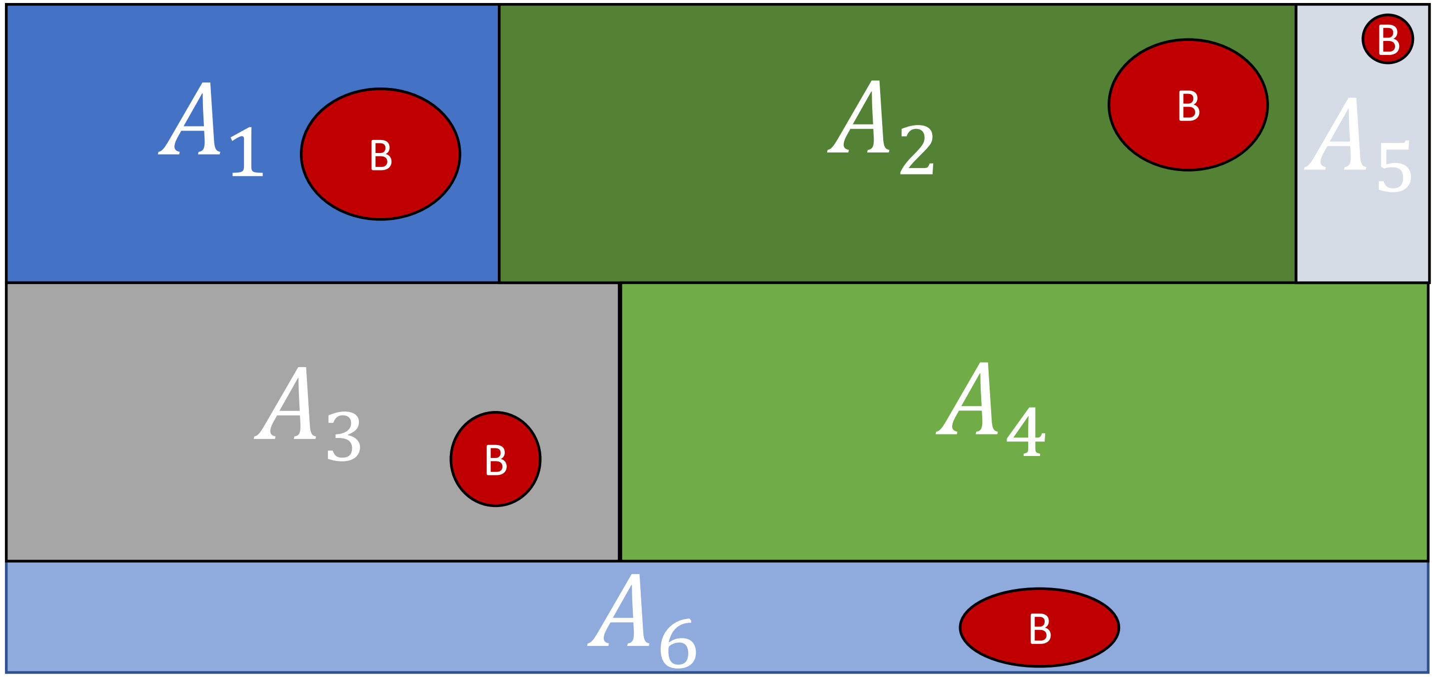
Формула полной вероятности

1) Пусть есть набор событий A_i таких, что они покрывают все вероятностное пространство. Для определенности будем считать, что каждое из этих событий – не невозможное



Формула полной вероятности

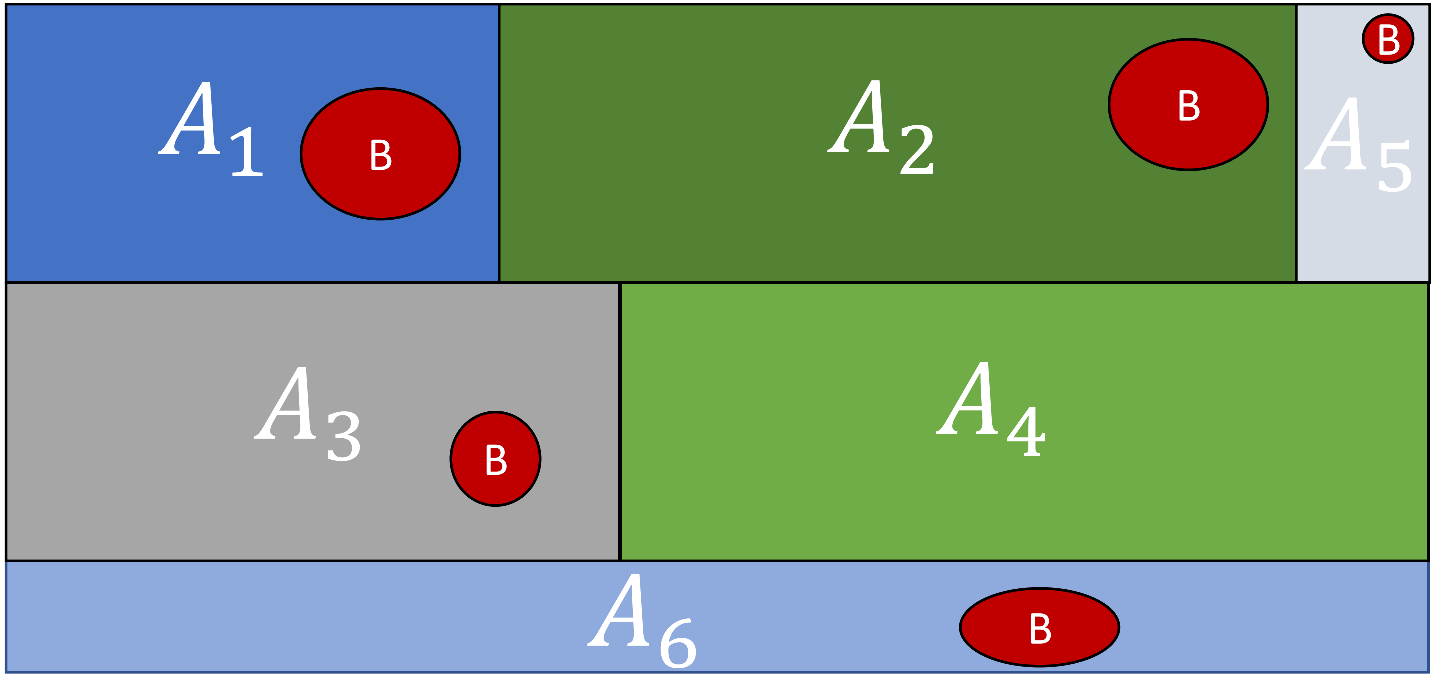
1) Пусть есть набор событий A_i таких, что они
покрывают все вероятностное пространство. Для определенности будем считать, что каждое из этих событий – не невозмножное
2) Пусть мы хотим подсчитать вероятность некоторого события B



Формула полной вероятности

1) Пусть есть набор событий A_i таких, что они
покрывают все вероятностное пространство. Для определенности будем считать, что каждое из этих событий – не невозможное
2) Пусть мы хотим подсчитать вероятность некоторого события B, при этом нам известны все $P(B|A_i)$

Тогда вероятность этого события будет вычисляться по следующей формуле:



$$P(B) = \sum_i P(A_i)P(B|A_i)$$

Теорема Байеса

P(“A и B”) = $P(A|B)P(B) = P(B|A)P(A)$

Если мы предполагаем, что P(B) – не невозможное, то можем переписать последнее равенство в следующем виде

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Терапия

Предлагается следующий тест на болезнь - в 95% данный диагноз это теста верен.

То есть в 95% для больного человека он правильно укажет, что он больной (и ошибется в 5% случаев)

И в 95% для здорового человека он укажет, что тот здоровый (и ошибется в 5% случаев)

Болезнь встречается у 1% популяции.

Согласно тесту Фред - болен. Какова вероятность того, что он действительно болен?

Терапия

Событие А – Фред болен

Событие В – тест показал положительный результат

$$P(A) = 0.01$$

$$P(\bar{A}) = \ 1 - P(A)$$

$$P(B|A) = 0.95 \qquad P(\bar{B}|A) = 0.05$$

$$P(B|\bar{A}) = 0.05 \qquad P(\bar{B}|\bar{A}) = 0.95$$

“Согласно тесту Фред - болен. Какова вероятность того, что он действительно болен?”

Что у нас спрашивают?

Терапия

Событие А – Фред болен

Событие В – тест показал положительный результат

$$P(A) = 0.01$$

$$P(\bar{A}) = \ 1 - P(A)$$

$$P(B|A) = 0.95 \qquad P(\bar{B}|A) = 0.05$$

$$P(B|\bar{A}) = 0.05 \qquad P(\bar{B}|\bar{A}) = 0.95$$

“Согласно тесту Фред - болен. Какова вероятность того, что он действительно болен?”

$P(A|B)$ - ?

Терапия

$$P(A) = 0.01$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(B|A) = 0.95 \qquad P(\bar{B}|A) = 0.05$$

$$P(B|\bar{A}) = 0.05 \qquad P(\bar{B}|\bar{A}) = 0.95$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95 \cdot 0.01}{P(B)}$$

Как найти P(B)?

Терапия

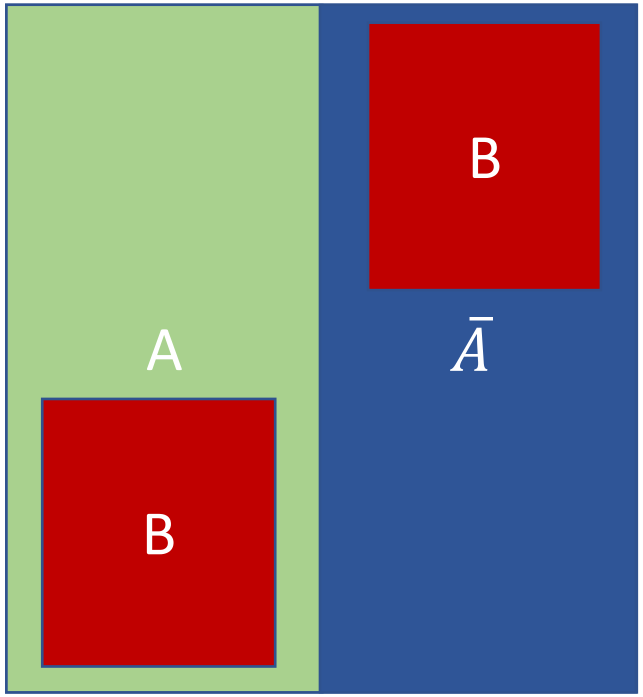
$$P(A) = 0.01$$

$$P(\bar{A}) = 1 - P(A) = 0.99$$

$$P(B|A) = 0.95 \qquad P(\bar{B}|A) = 0.05$$

$$P(B|\bar{A}) = 0.05 \qquad P(\bar{B}|\bar{A}) = 0.95$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95 \cdot 0.01}{P(B)}$$



По формуле полной вероятности

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.95 \cdot 0.01 + 0.05 \cdot 0.99$$

Терапия

$$P(A) = 0.01$$

$$P(\bar{A}) = \ 1 - P(A) = 0.99$$

$$P(B|A) = 0.95 \qquad P(\bar{B}|A) = 0.05$$

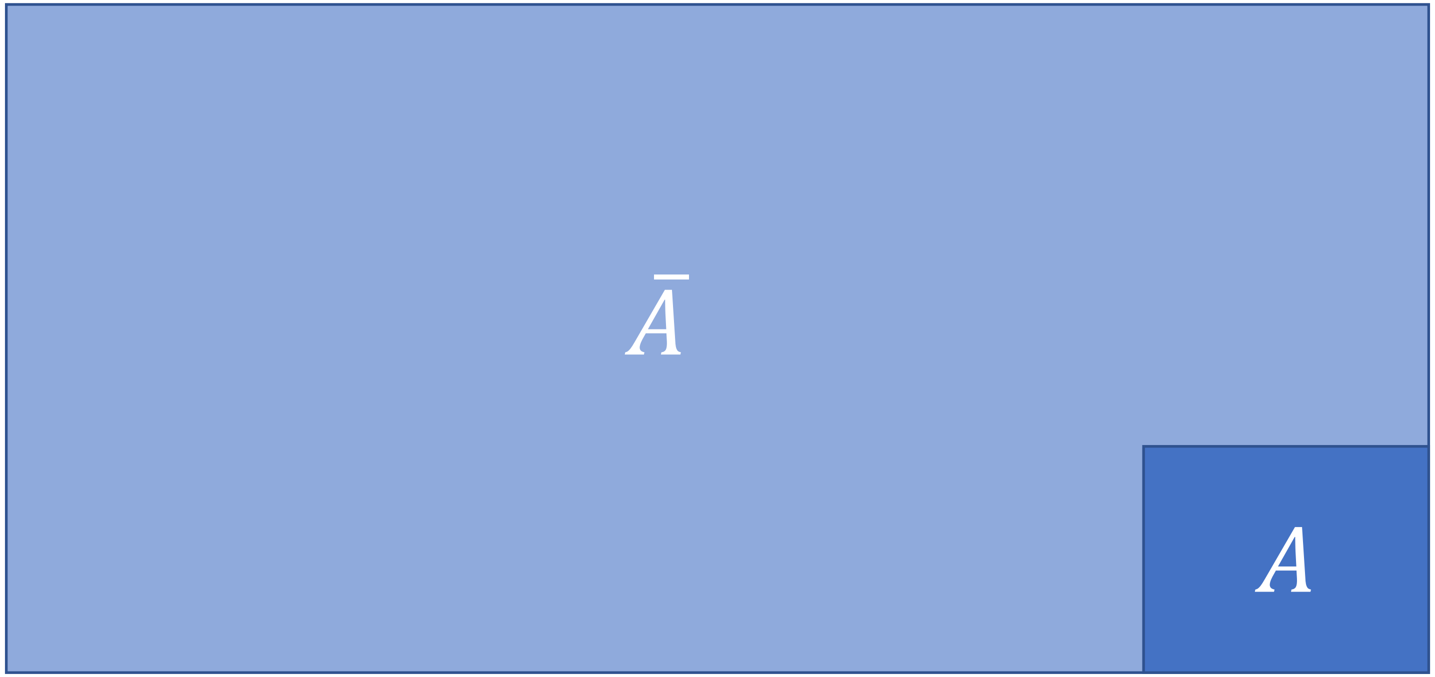
$$P(B|\bar{A}) = 0.05 \qquad P(\bar{B}|\bar{A}) = 0.95$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} A \approx 0.161$$

То есть вероятность даже меньше 50%

Терапия

Почему так? Для простоты представим, что мы взяли 2000 человек из того же места, что и Фред

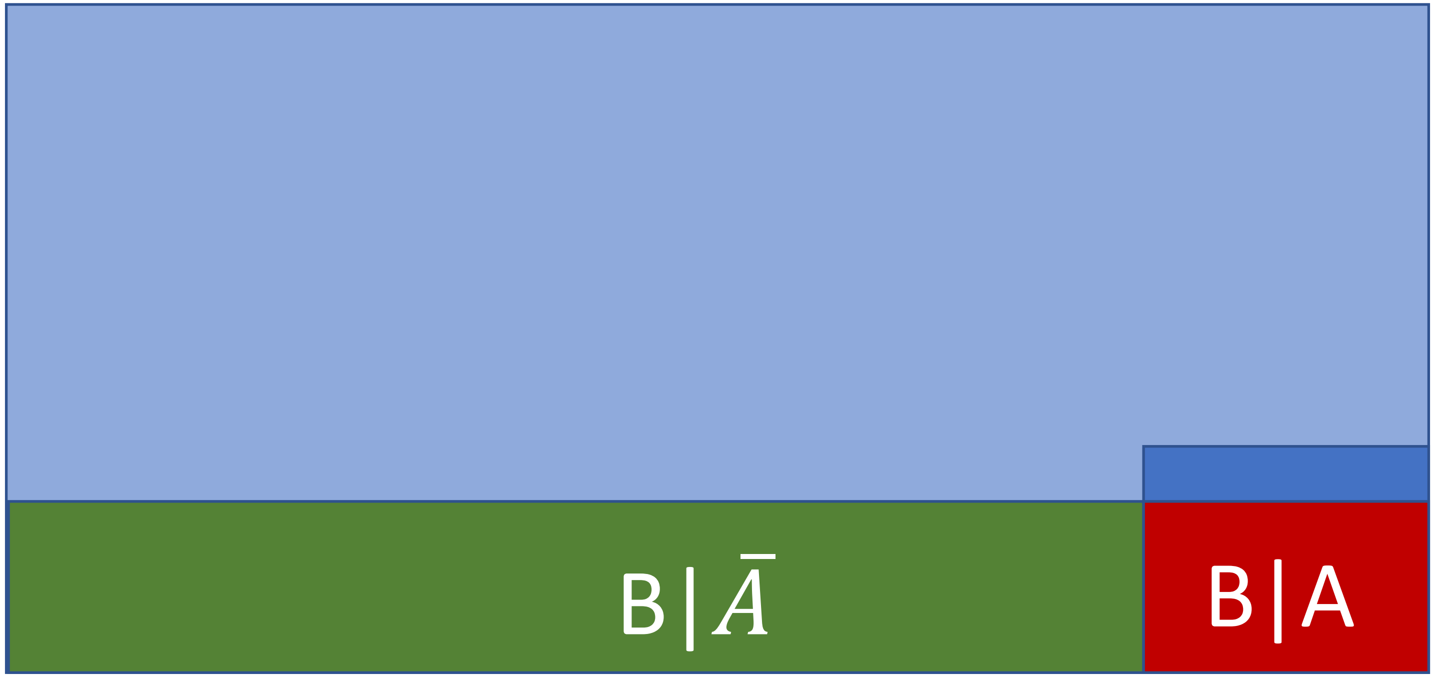


2000 * 0.01 =
20 больных

2000 * 0.99 =
1980 здоровых

Терапия

Посчитаем, сколько будет положительных диагнозов – B|A и $B|\bar{A}$

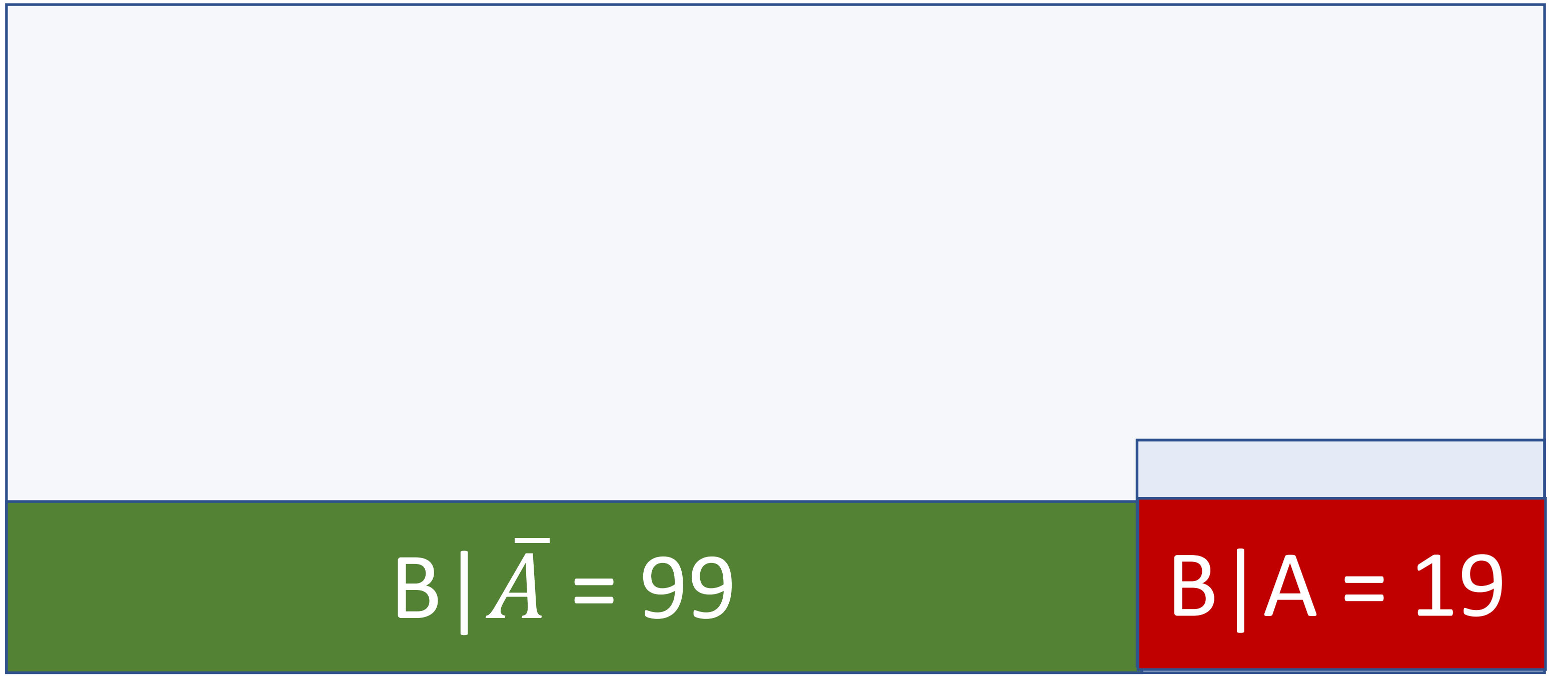


20 * 0.95 = 19 положительных диагнозов у больных

1980 * 0.05 = 99 положительных диагнозов у здоровых

Терапия

Теперь мы хотим посчитать долю болеющих среди людей с положительным тестом - $P(A|B)$.



Терапия

Теперь мы хотим посчитать долю болеющих среди людей с положительным тестом - $P(A|B)$.

$B\,|\,\bar{A} = 99$	$B\,|\,A = 19$

Доля реально больных среди диагностированных $= \frac{19}{99+19} = 0.161$

Теорема Байеса

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A)$$

$P(A)$ - априорная вероятность события А

$P(B)$ - априорная вероятность события В

$P(A|B)$ - апостериорная вероятность события А

$P(B|A)$ - вероятность события А при условия события В

Теорема Байеса

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(A)}{P(B)}P(B|A)$$

$P(A)$ - априорная вероятность события А - **гипотезы**

$P(B)$ - априорная вероятность события В - **наблюдения**

$P(A|B)$ - вероятность гипотезы **после** наблюдения

$P(B|A)$ - вероятность наблюдения при условии верности гипотезы

Теорема Байеса

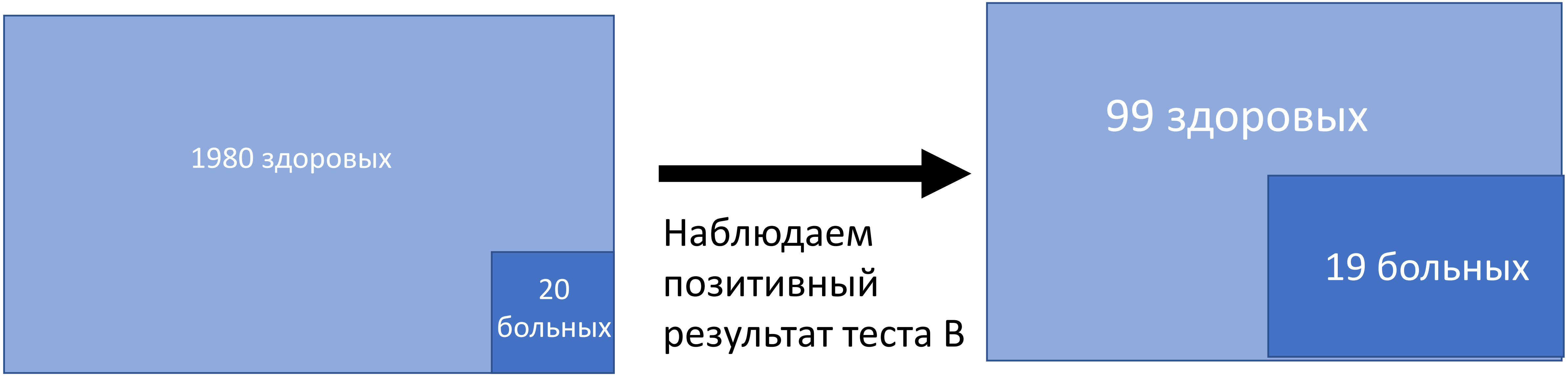
$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

Есть ли принципиальное отличие между P(A|B) и P(A)?
Является ли одно «более вероятностью» чем другое

Теорема Байеса

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

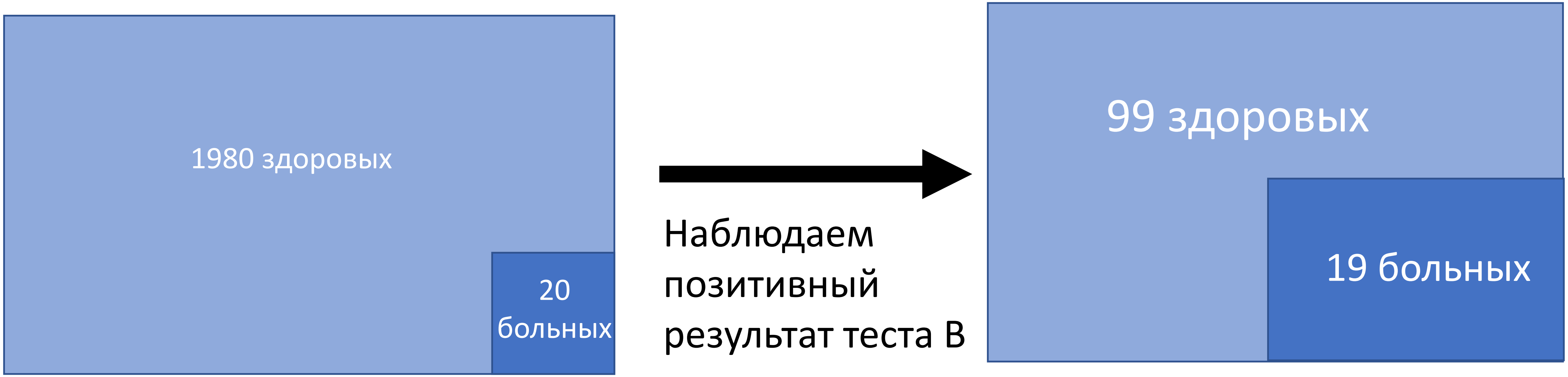
Есть ли принципиальное отличие между P(A|B) и P(A)?
Является ли одно «более вероятностью» чем другое



Теорема Байеса

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

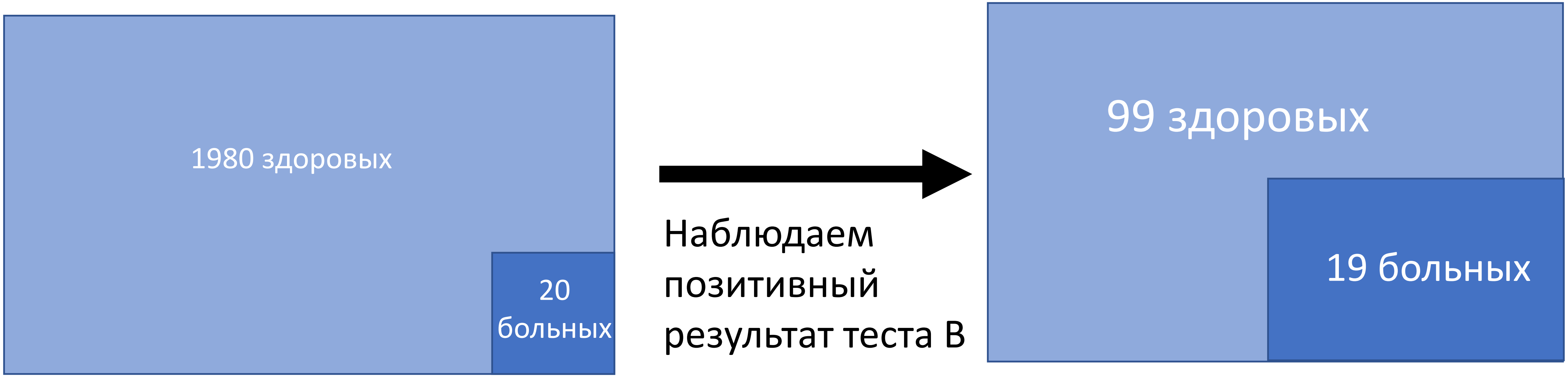
Нет. Одна вероятность посчитана в одном вероятностном пространстве, вторая – в другом, «вырезанном» из первого на основе наблюдения.



Теорема Байеса

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

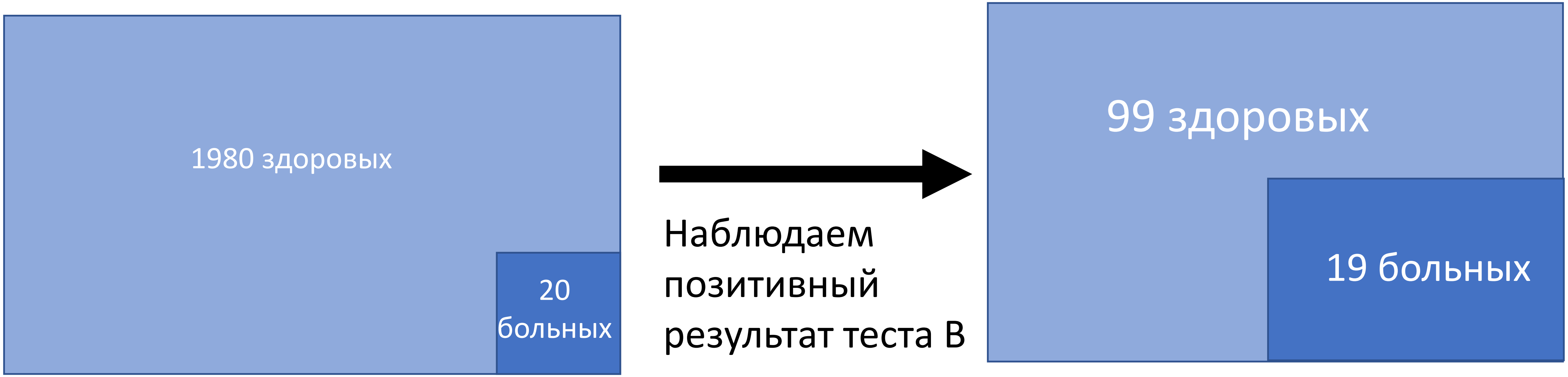
Нет. Одна вероятность посчитана в одном вероятностном пространстве, вторая – в другом, «вырезанном» из первого на основе наблюдения.



Теорема Байеса

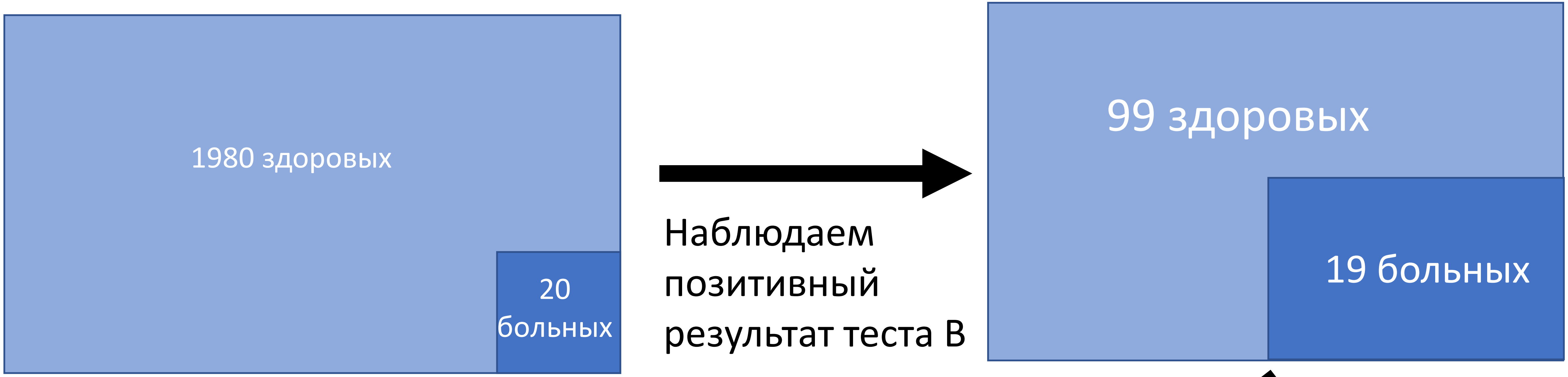
$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

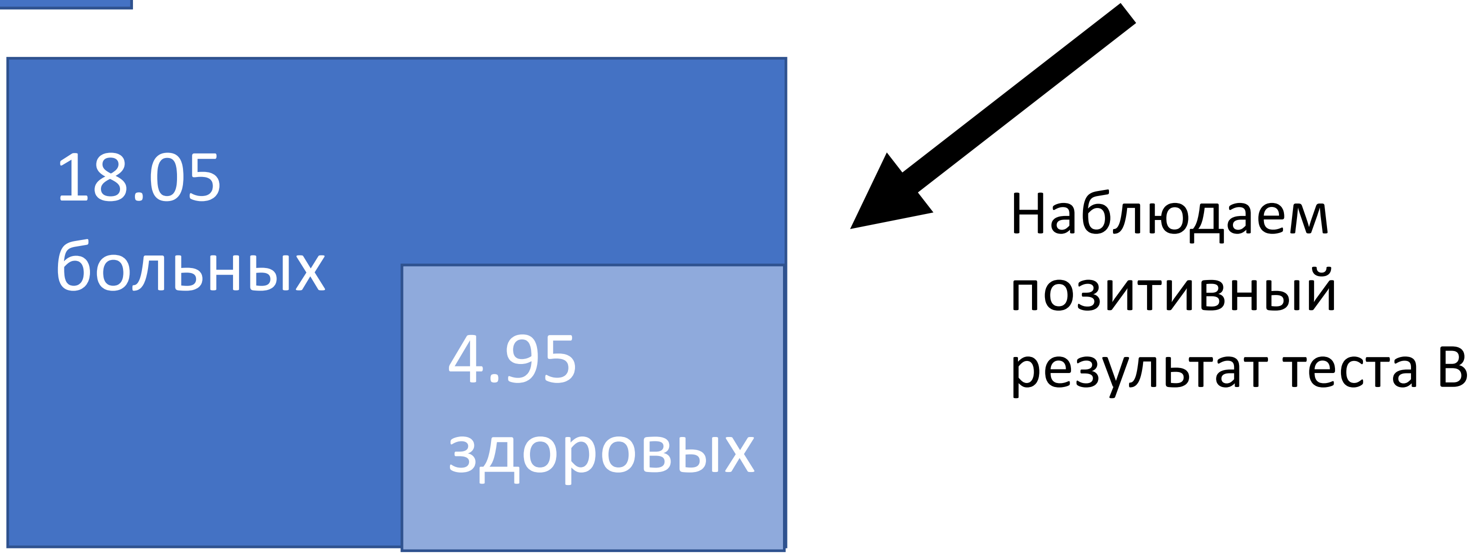
Мы можем применять вероятностью, подсчитанные на основании одного наблюдения, при трактовке нового. Допустим результаты тестов независимы. Тогда



Теорема Байеса

Допустим результаты тестов независимы. Тогда





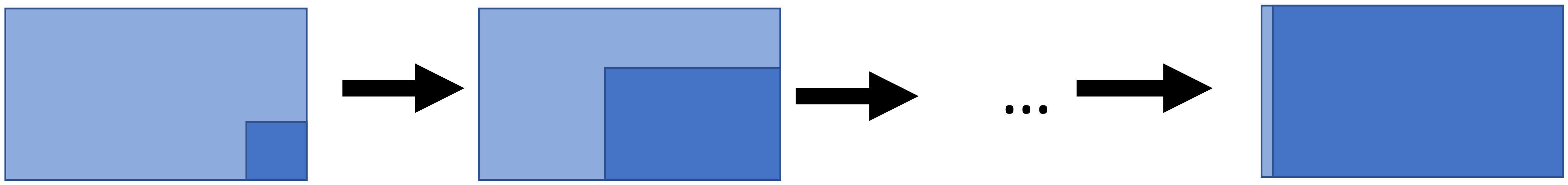
Доля реально больных среди диагностированных на основе двух тестов $=$

$$\frac{18.05}{4.95+18.05} = 0.785$$

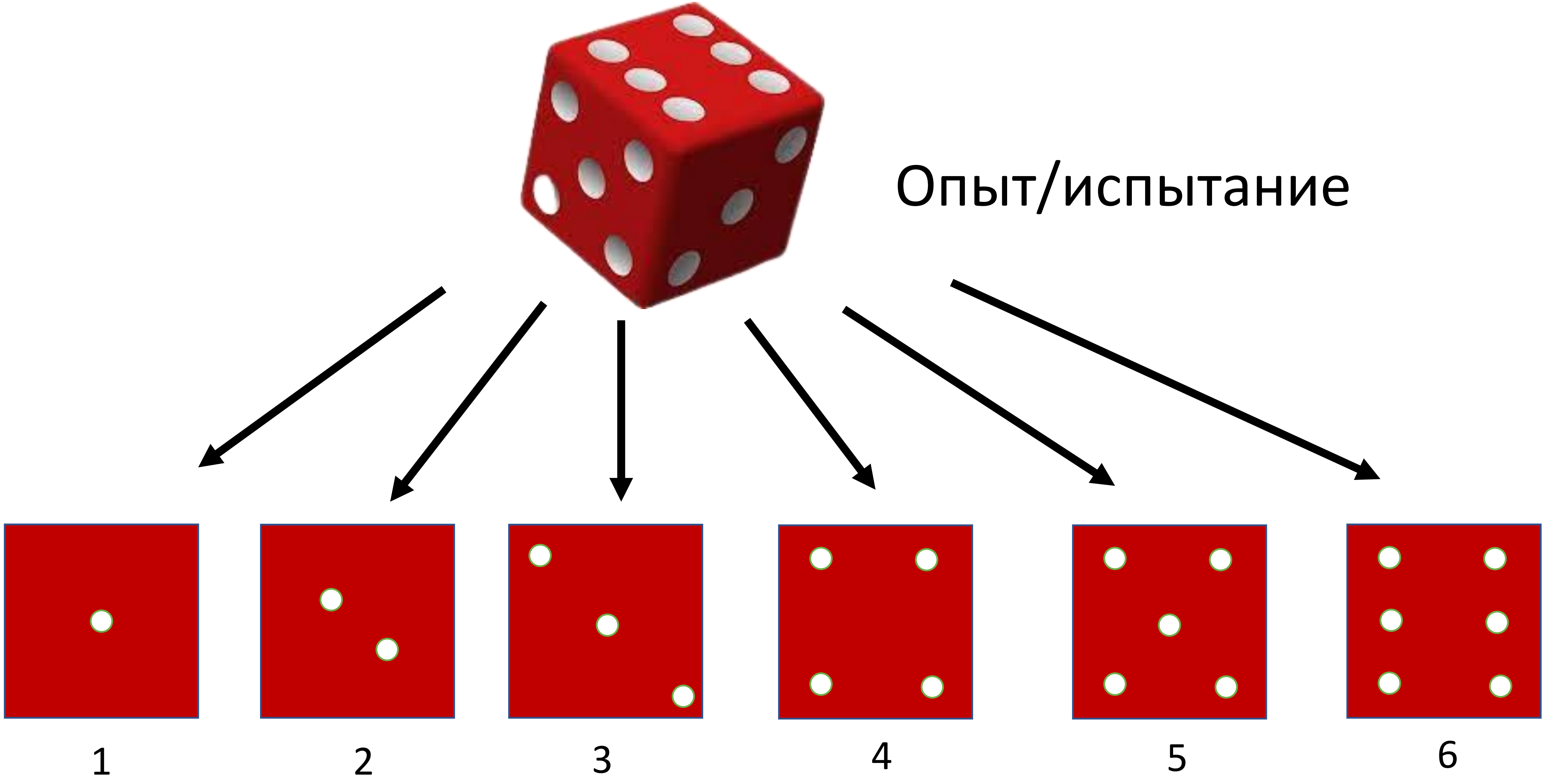
Теорема Байеса

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

Теорема Байеса – инструмент, как на основе нашей изначальной картины мира (априорной), получать все более и более точную (апостериорную) картину, включая в нее новые наблюдения.
При этом она допускает использование в качестве априорной картины мира апостериорную, которую за вас получил другой исследователь

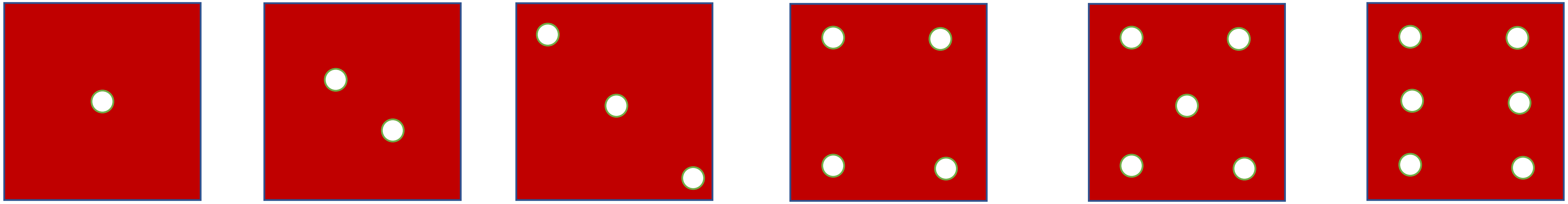


Случайная величина



Случайная величина

Случайная величина – величина, принимающая определенное в ходе эксперимента то или иное значение



1 2 3 4 5 6

Примеры случайных величин

Примеры случайных величин

1) число, выпавшее на шестигранном кубике;
2) величина, равная 0, если на моментке выпала решка и 1, если орел;
3) пол ребенка;
4) время прихода автобуса;
5) число бракованных товаров в партии;
6) время доставки еды из Яндекс Лавки;
7) курс рубля.

Дискретные случайные величины

Случайная переменная принимает не более чем **счетное** множество значений.

Такие распределения можно задавать при помощи таблицы, где каждому значению сопоставлена его вероятность. В сумме вероятности должны давать 1.

Бернуллиевская случайная величина

Значение случайной величины	1 (успех)	0 (неудача)
Вероятность значения	p	1 - p

Биномиальное распределение

Имеем последовательность из n испытаний Бернулли. Нам интересно, сколько раз произойдет успех. Не важно, в каких конкретно испытаниях он выпадет

4

4

3

Биномиальное распределение

Имеем последовательность из n испытаний Бернулли. Нам интересно, сколько раз выпадет орел. Не важно, в каких конкретно испытаниях он выпадет

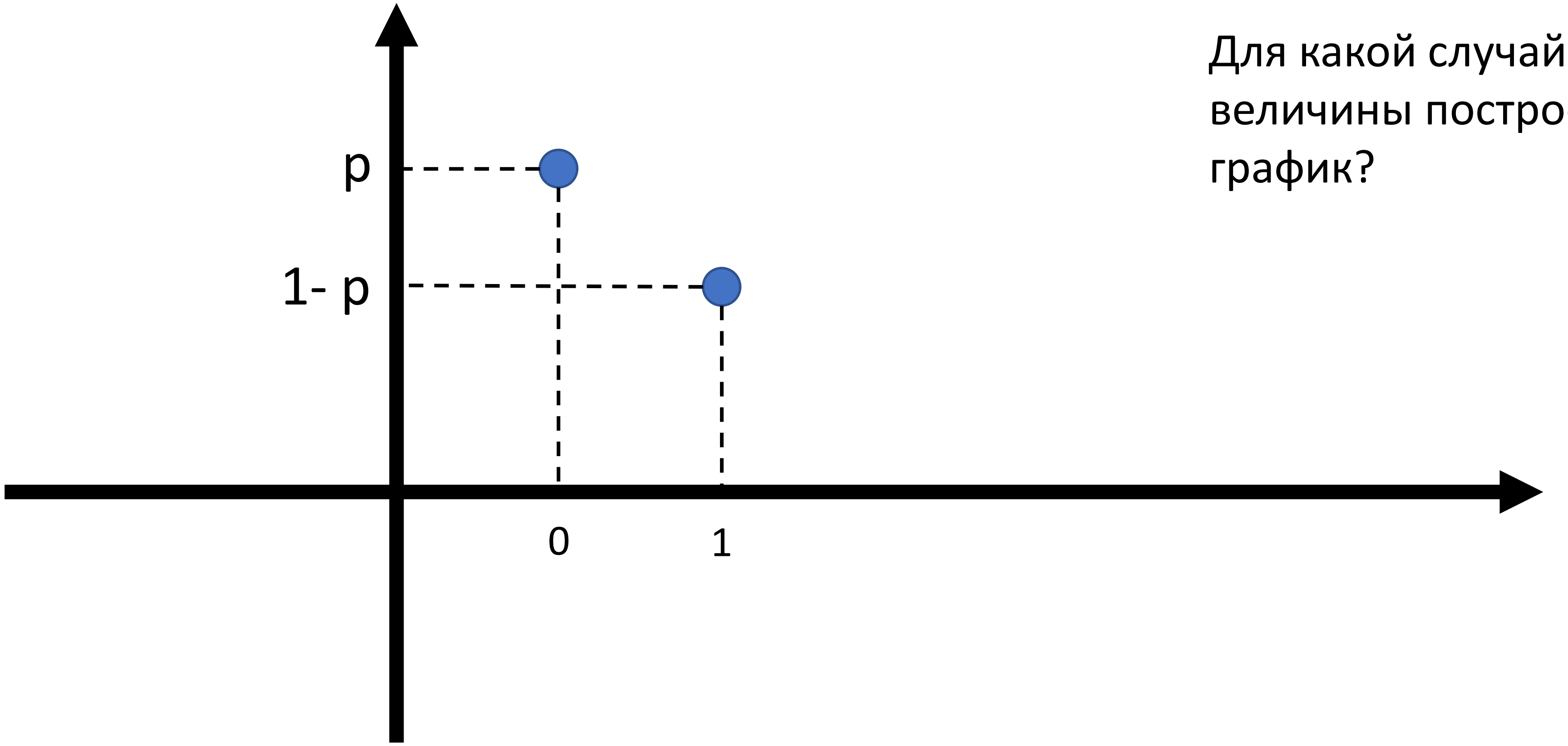
$$p(k) = C_n^k p^k (1 - p)^{n-k}$$

$$C_n^k = \frac{n!}{(n-k)!\,k!}$$

Значение случайной величины	0	1	…	n
Вероятность	$(1-p)^n$	$n(1-p)^{n-1}q$	…	p^n

Probability mass function (функция вероятности)

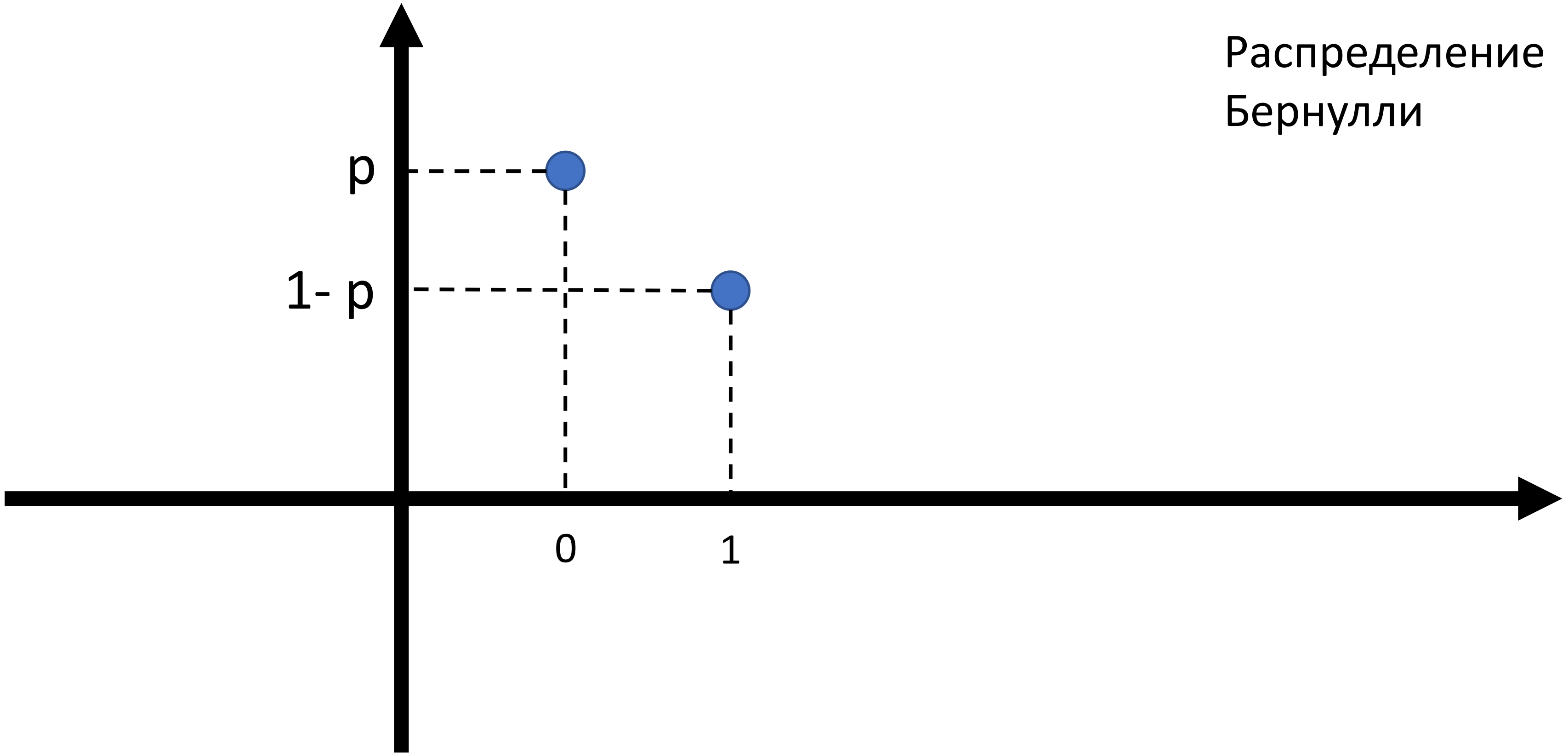
Понятно, что таблицу можно визуализировать – по оси X откладываем значения, по оси Y - вероятность



Для какой случайной величины построен график?

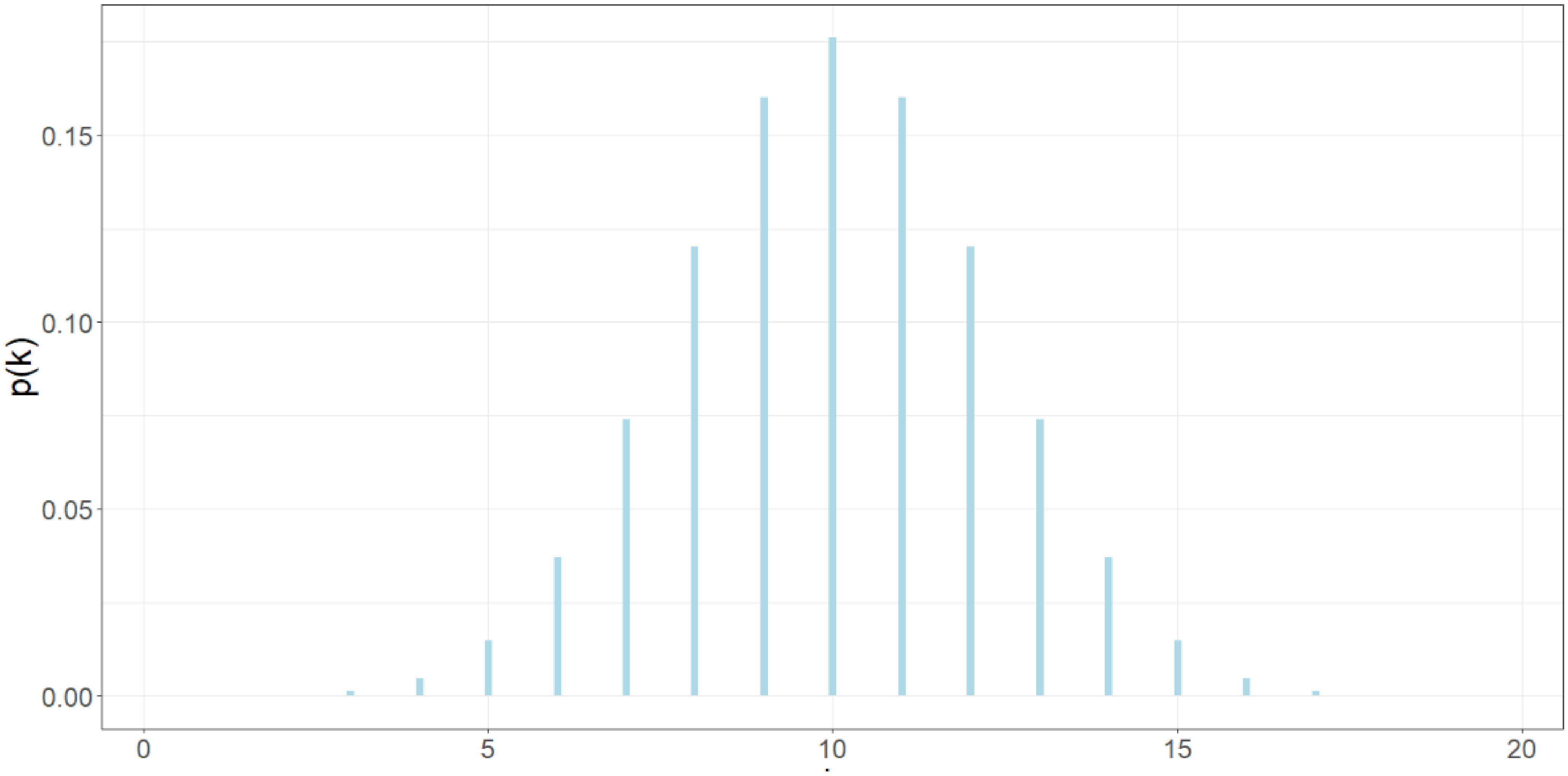
Probability mass function (функция вероятности)

Понятно, что таблицу можно визуализировать – по оси X откладываем значения, по оси Y - вероятность

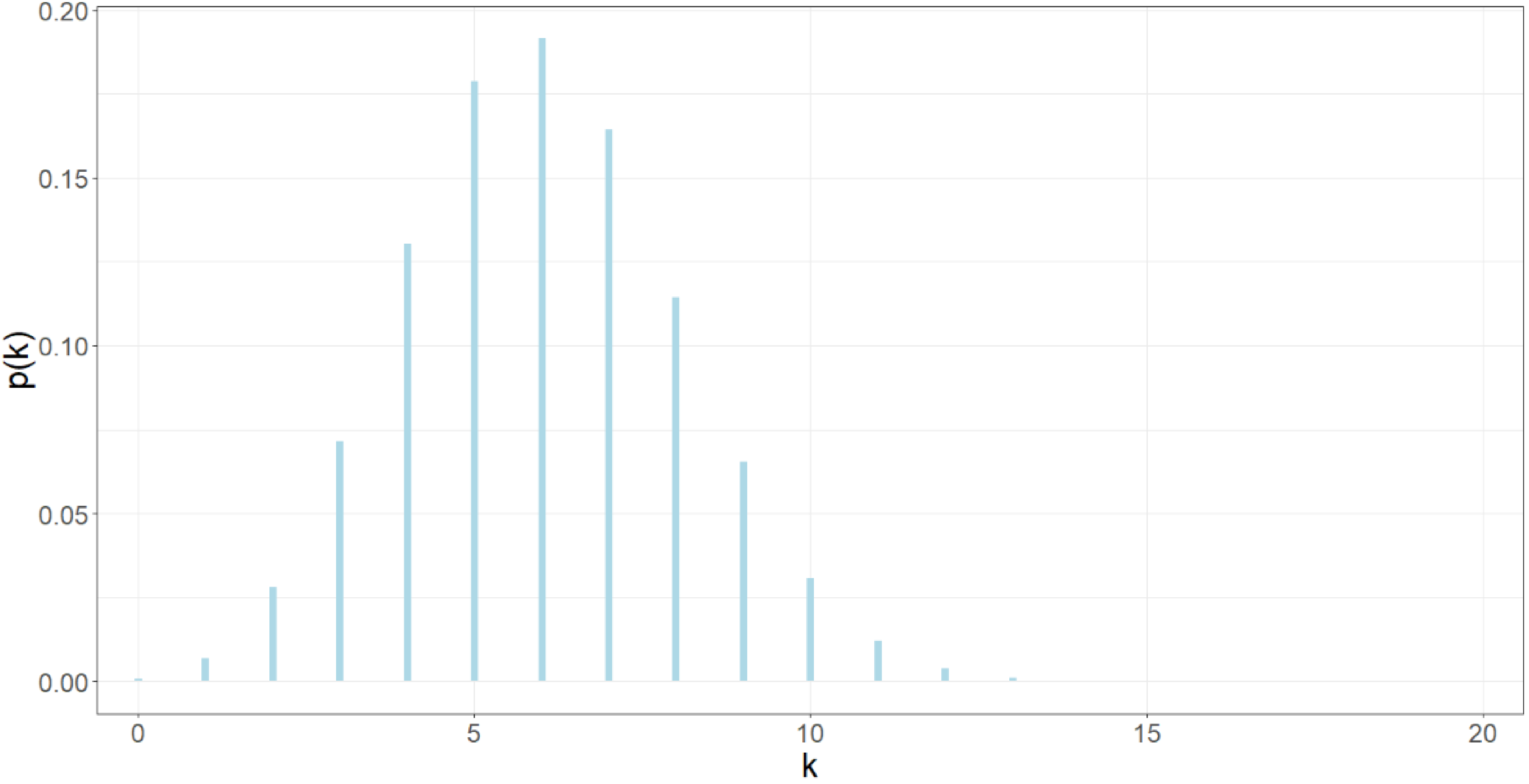


Распределение Бернулли

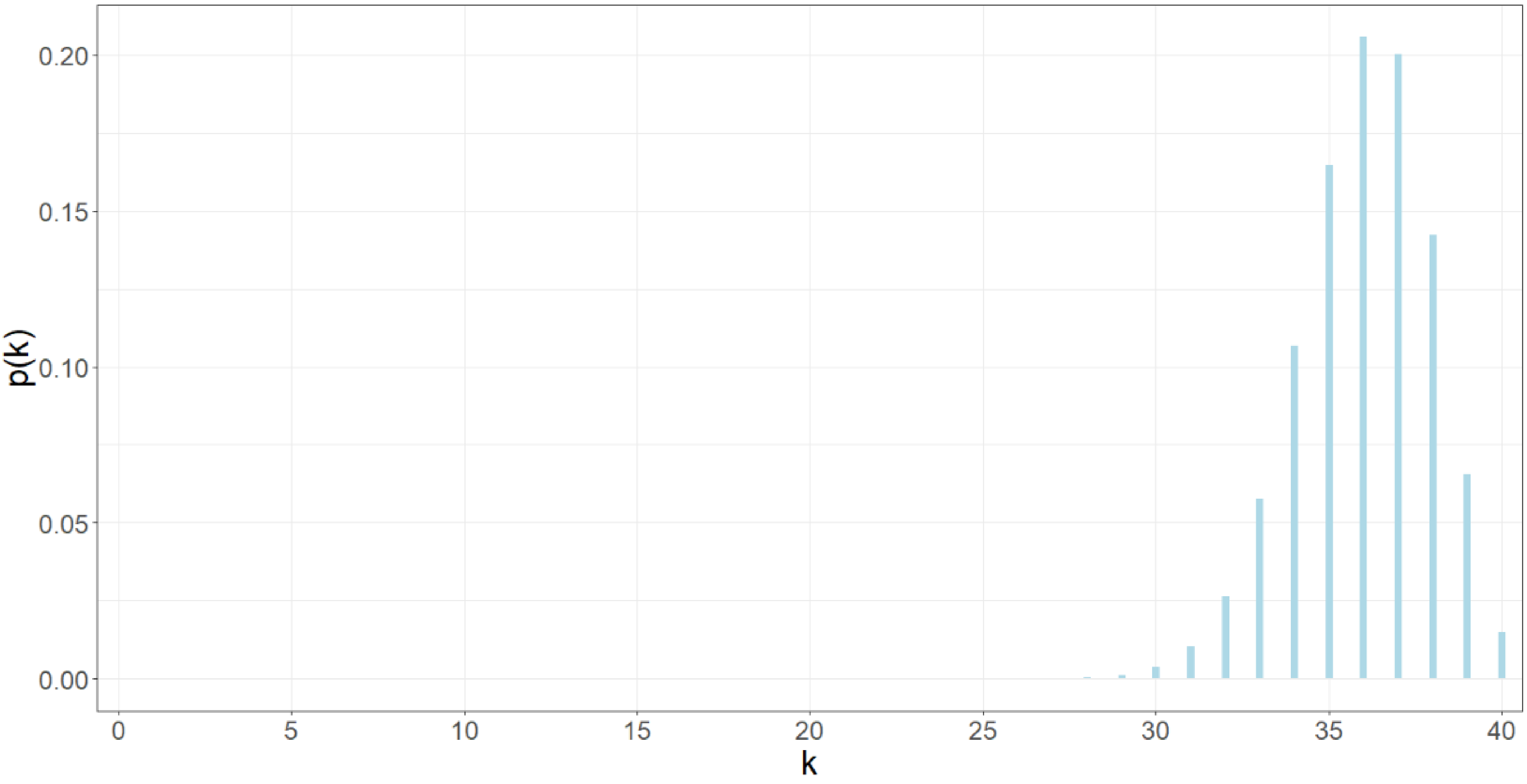
Функция вероятности для биномиального распределения



Функция вероятности для биномиального распределения



Функция вероятности для биноминального распределения



Распределение Пуассона

Лавка делает 1 кг теста для кексов

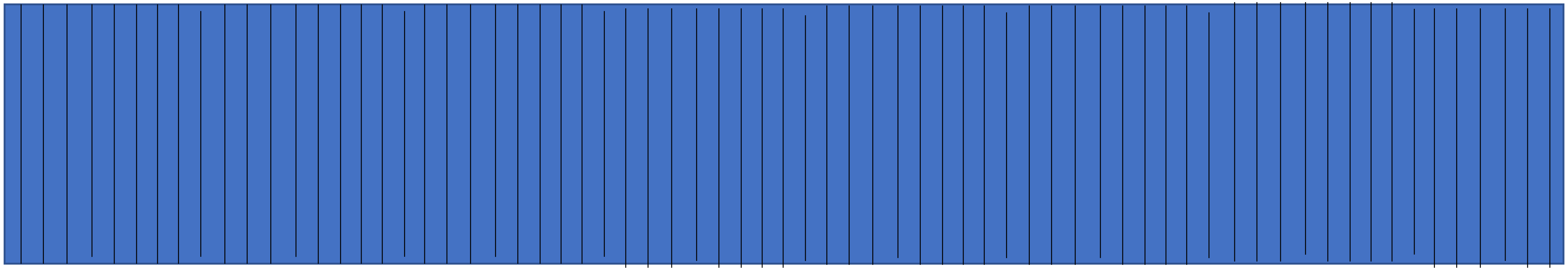
В этот кг теста в среднем попадает порядка 50 изюминок

Считая изюминки неделимыми сущностями, сколько изюминок мы ожидаем увидеть в одном кексе (100г)?

Распределение Пуассона

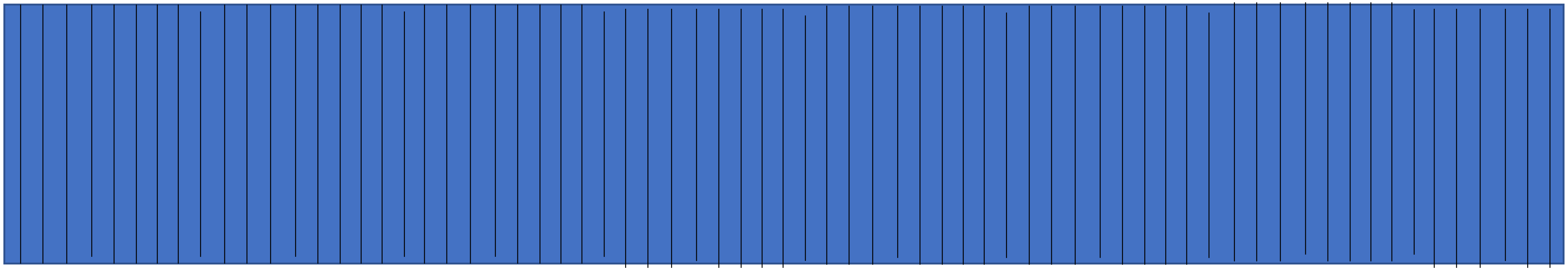
Понятно, что так как в среднем на 1 кг теста приходится 50 изюминок , то на 100г придется 5 изюминок

Но по случайным причинам в один кекс изюминки могут не попасть или попасть меньше, а в другую попасть больше. Нам нужно как-то описать это число.
Можно «разделить» кекс на настолько маленькие участки, что в каждом мы либо встретим не больше одной изюминки



Распределение Пуассона

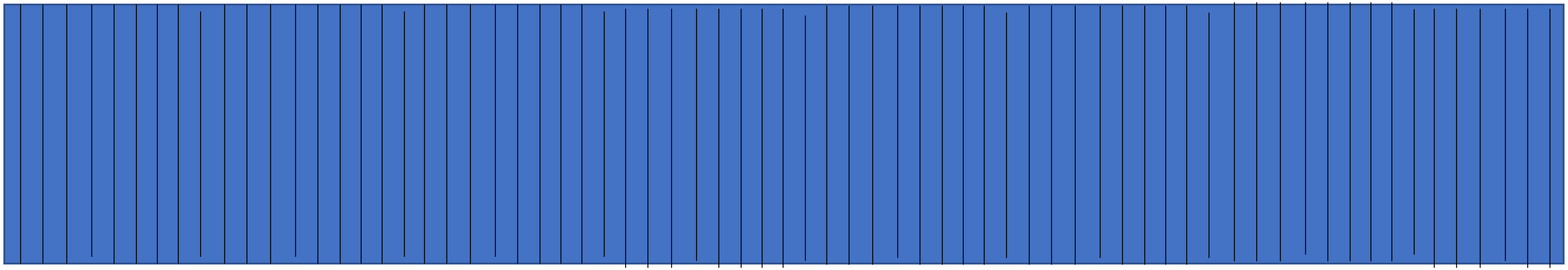
Можно «разделить» кекс на настолько маленькие участки, что в каждом мы либо встретим не больше одной изюминки



Всего кусков получится n (>>5), а вероятность попадания изюминки в кусок – $\frac{5}{n}$.
Нужно понять, в какое число кусков попала изюминка. Какое распределение?

Распределение Пуассона

Можно «разделить» кекс на настолько маленькие участки, что в каждом мы либо встретим не больше одной изюминки

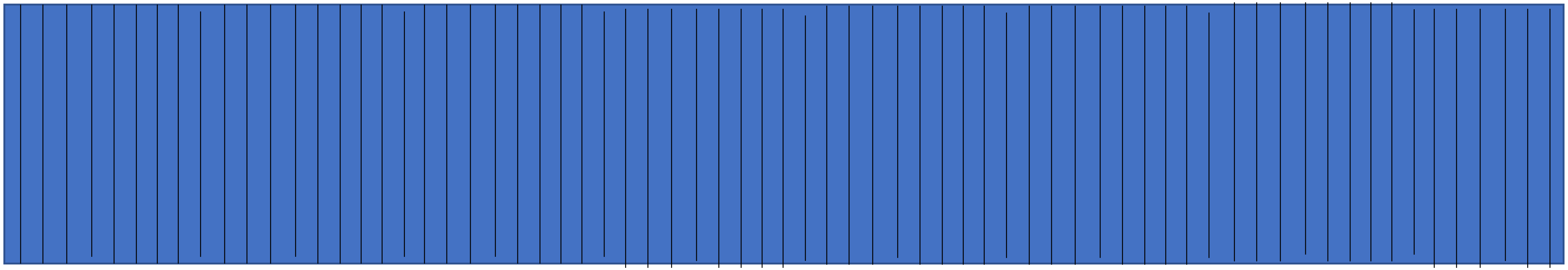


Всего кусков получится n (>>5), а вероятность попадания изюминки в кусок – $\frac{5}{n}$.
Нужно понять, в какое число кусков попала изюминка. Какое распределение?

Биномиальное.

Распределение Пуассона

Можно «разделить» кекс на настолько маленькие участки, что в каждом мы либо встретим не больше одной изюминки



Всего кусков получится n (>>5), а вероятность попадания изюминки в кусок – $\frac{5}{n}$. Нужно понять, в какое число кусков попала изюминка. Какое распределение?

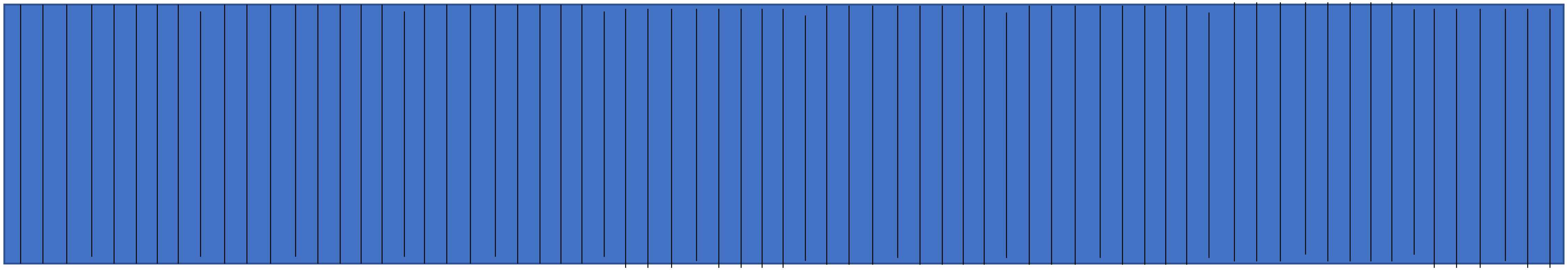
Биномиальное.

Но число кусков, на которые бить – наш произвол.

Чем больше кусков – тем точнее аппроксимация

Распределение Пуассона

Можно «разделить» кекс на настолько маленькие участки, что в каждом мы либо встретим не больше одной изюминки



Но число кусков, на которые бить – наш произвол.

Чем больше кусков – тем точнее аппроксимация.

При $n \to \infty$ (соответственно, $p \to 0$) получим распределение Пуассона.

У этого распределения будет один параметр – λ равный как раз среднему числу изюминок в кексе

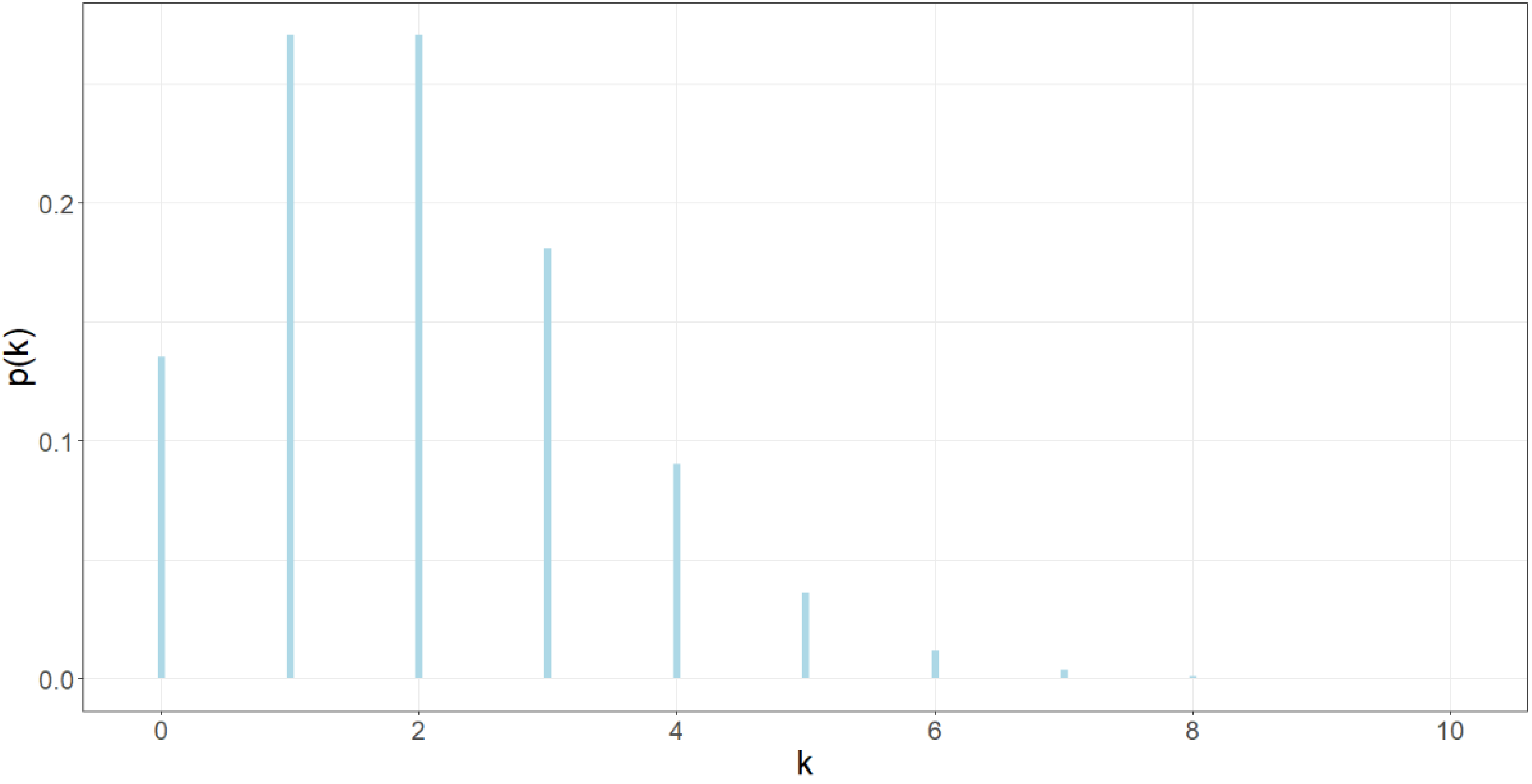
Распределение Пуассона

Число успехов, произошедших за фиксированный отрезок времени или пространства.
При этом мы предполагаем, есть некое фиксированное число успехов в единицу времени и успехи происходят независимо

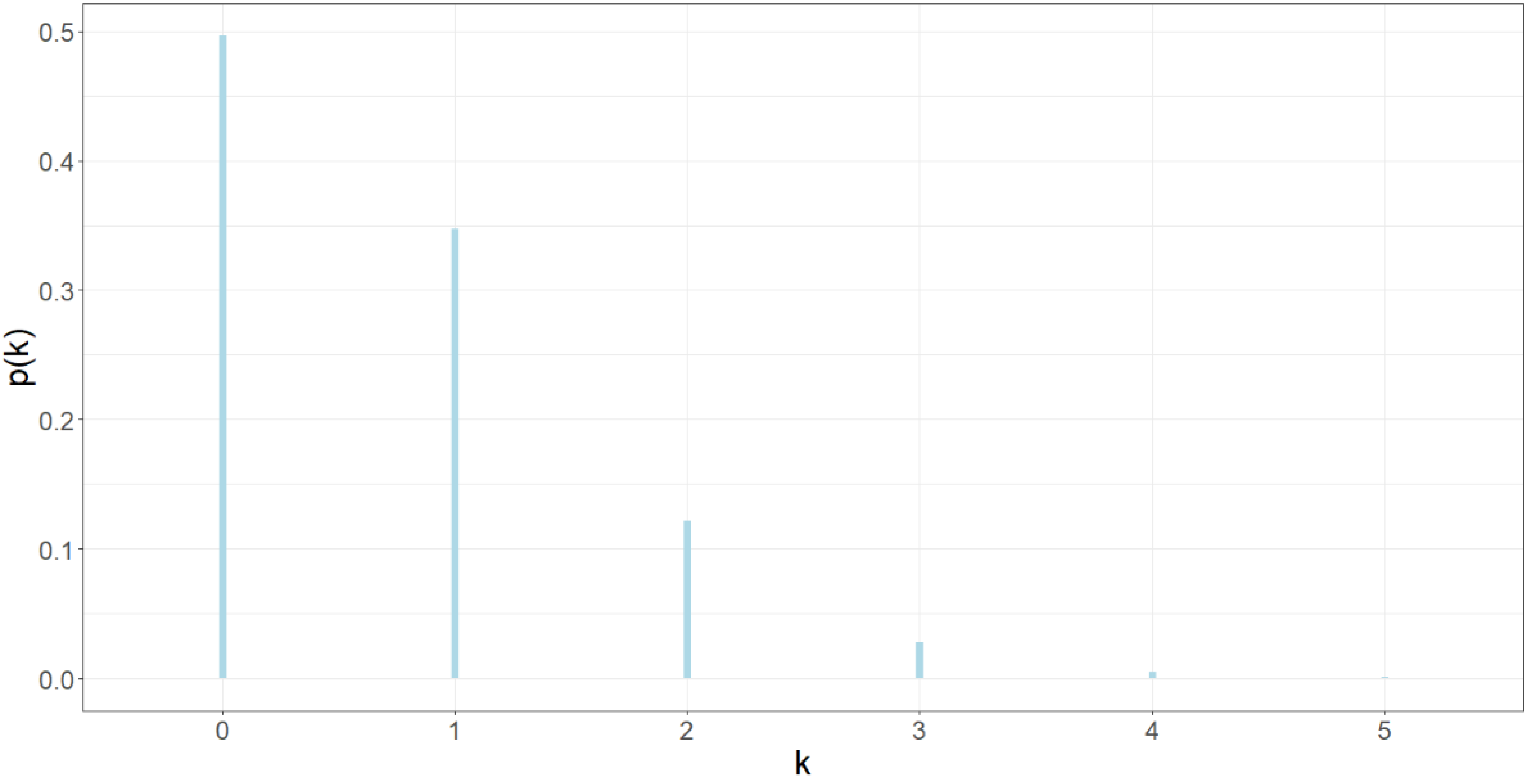
$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

1) Гипотетически – число успехов не ограничено. Фактически – сильно отклоняющиеся от значения λ быстро становятся маловероятны;
2) Может использоваться как аппроксимация биномиального распределения при $np < 5$, $n(1-p) < 5$;

Распределение Пуассона ($\lambda = 2$)

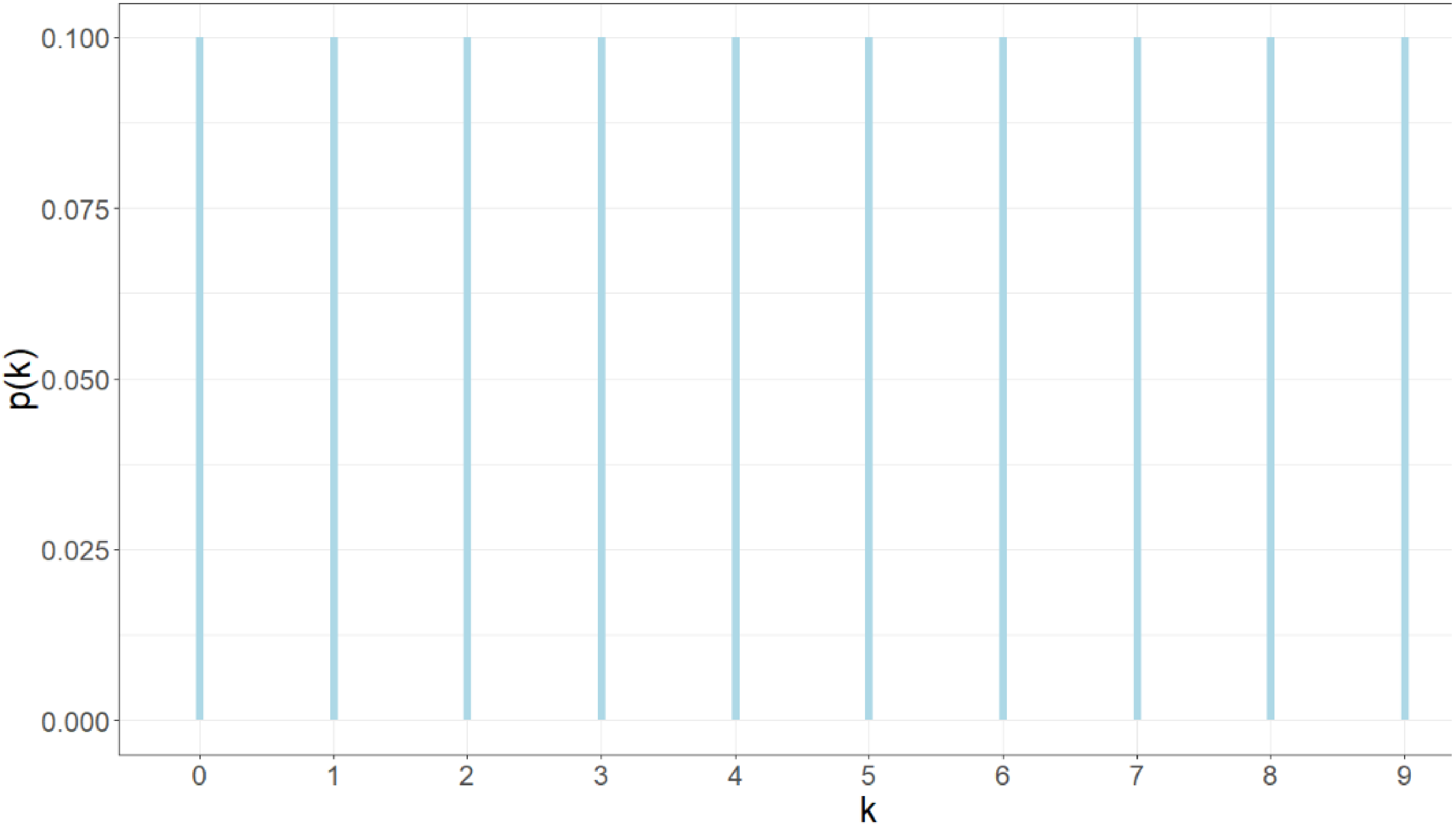


Распределение Пуассона ($\lambda = 0.7$)



Равномерное дискретное распределение

На определенном интервале [n, m] равновероятно генерируем **целые** числа



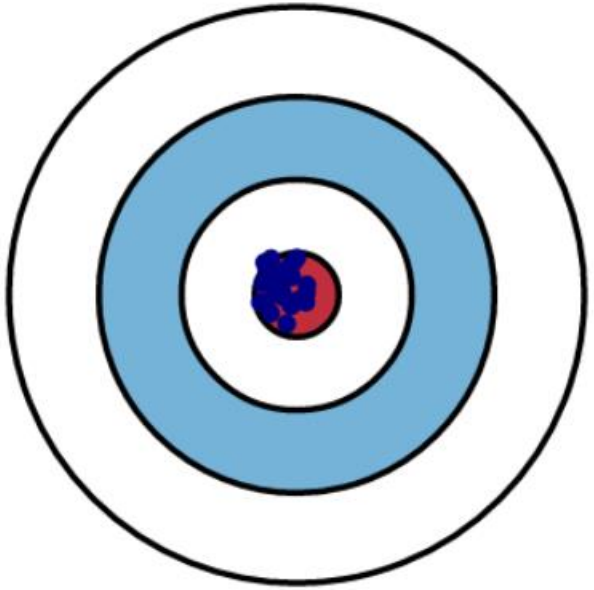
Константа

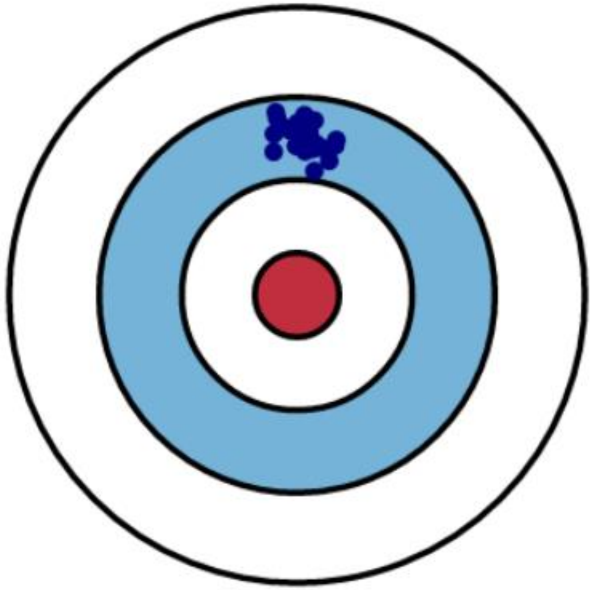
В принципе, тоже дискретная случайная величина. Просто очень скучная((

Среднее дискретной случайной величины

$$E(X) = \sum_i x_i p_i$$

Куда мы попадаем в среднем





Среднее дискретной случайной величины

$$E(X) = \sum_i x_i p_i$$

Чему равно матожидание распределения Бернулли?

Среднее дискретной случайной величины

$$E(X) = \sum_i x_i p_i$$

Чему равно матожидание распределения Бернулли?

$$E(X) = \sum_i x_i p_i = 0 \ * \ (1 - p) + 1 \ * p = p$$

Свойства среднего (верны и в общем случае)

Если у нас есть две случайные величины X и Y, и константа c, то

$$E(c) = c$$

$$E(cX) = cE(X)$$

$$E(X + Y) = E(X) + E(Y)$$

Если у нас есть две **независимые** случайные величины X и Y, и константа c, то

$$E(XY) = E(X) \cdot E(Y)$$

Свойства среднего (корректны и для непрерывных распределений)

Если у нас есть две случайные величины X и Y, и константа c, то

$$E(c) = c$$

$$E(cX) = cE(X)$$

$$E(X + Y) = E(X) + E(Y)$$

Если у нас есть две **независимые** случайные величины X и Y, то

$$E(XY) = E(X) \cdot E(Y)$$

Свойства среднего

Чему равно матожидание биномиального распределения?

Свойства среднего

Чему равно матожидание биномиального распределения?

$$E(\sum_i X_i) = \sum_i E(X_i) = n\, E(X_1) = np$$

Чему равно матожидание распределения Пуассона?

Свойства среднего

Чему равно матожидание биномиального распределения?

$$E(\sum_i X_i) = \sum_i E(X_i) = n\,E(X_1) = np$$

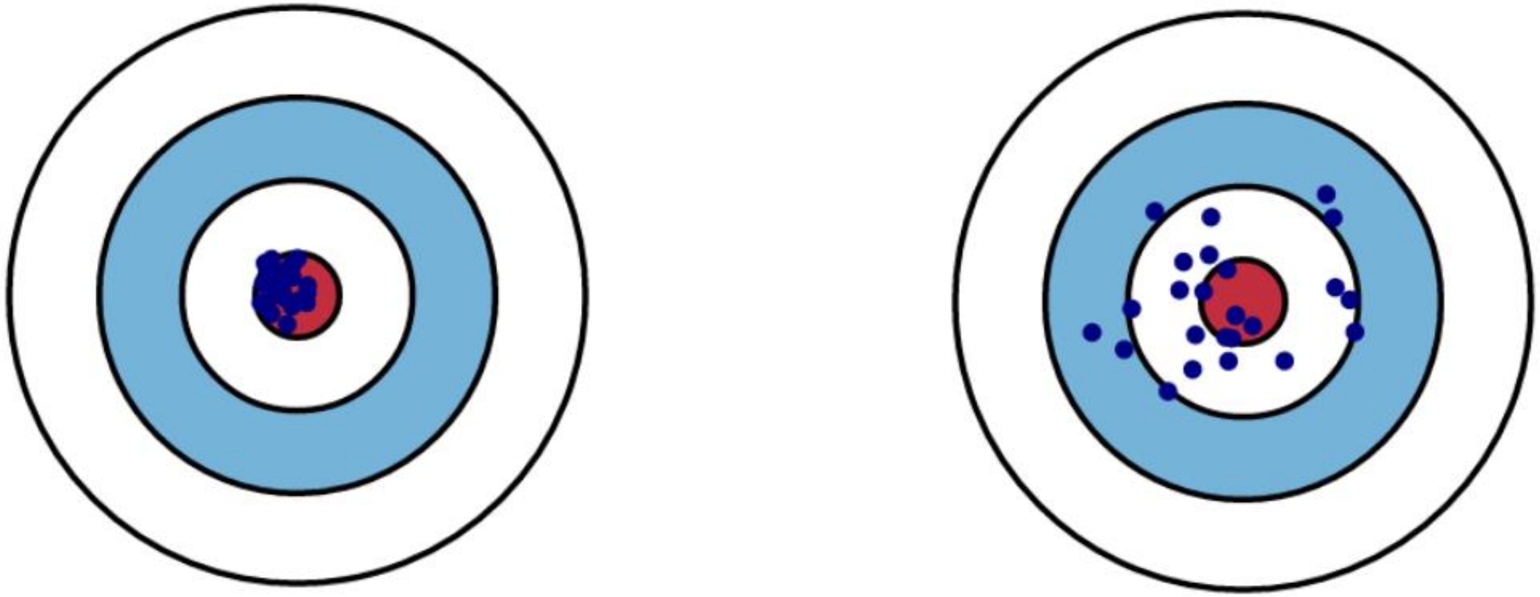
Чему равно матожидание распределения Пуассона?

$$E\big(pois(\lambda)\big) = \lambda$$

Дисперсия случайной величины

$$\mathrm{D}(X) = E(X - EX)^2$$

Характеризует, насколько сильно величина склонна отклоняться от среднего



Анекдот

Три аналитика данных на охоте.
Видят большого оленя.
Один целится, стреляет, промахнулся - на метр влево.
Втрой - целится, стреляет, не попал – на метр вправо.
Третий аналитик, не стреляя: «Ну, в среднем- мы его убили!»

Свойства дисперсии (корректны и для непрерывных распределений)

Если у нас есть две **независимые** случайные величины X и Y, и константа c, то

$$\mathrm{D}(X) = E(X^2) - (EX)^2$$

$$\mathrm{D}(X+Y) = D(X) + D(Y)$$

$$\mathrm{D}(\mathrm{c}X) = c^2 DX$$

Даже для **независимых** случайных величин в общем случае:

$$D(XY) \neq D(X)D(Y)$$

Дисперсия случайной величины

Какая дисперсия распределения Бернулли:

Дисперсия случайной величины

Дисперсия распределения Бернулли:

$$\mathrm{D}(X) = \mathrm{p} - p^2 = (1 - p)p$$

Какая дисперсия у биномиального распределения?

Дисперсия случайной величины

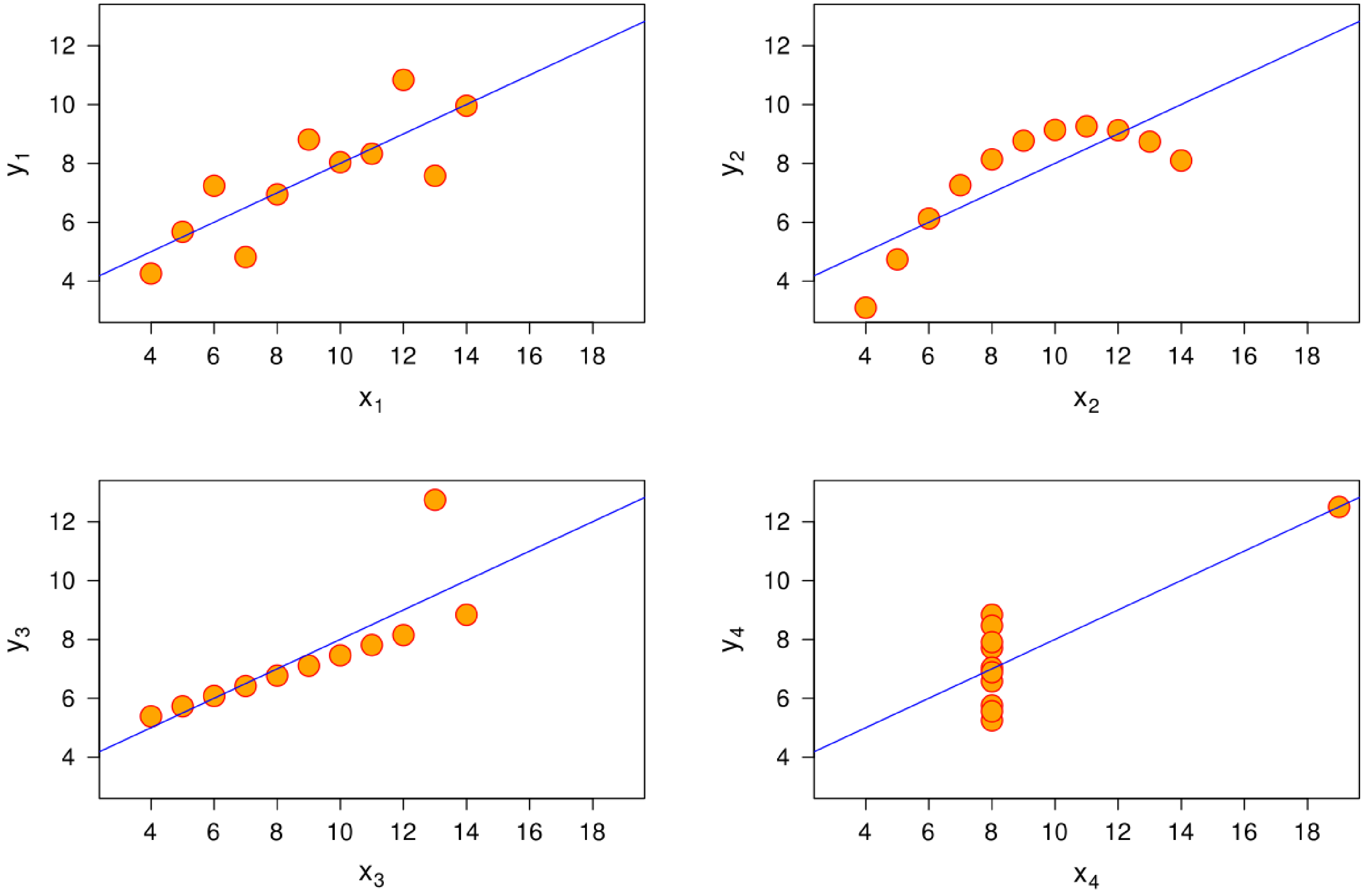
Дисперсия распределения Бернулли:

$$\mathrm{D}(X) = (1 - p)p$$

Какая дисперсия у биномиального распределения?

$$\mathrm{D}(X) = np(1 - p)$$

Квартет Энскомба



Все 4 набора данных имеют одинаковую дисперсию и матожидание