1. Вперёд, в рукопашную!

Минитеория:

1. http://bdemeshev.github.io/pr201/ или http://pokrovka11.wordpress.com

2. Константы. Строчные английские буквы, a, x, z.

3. События. Заглавные английский буквы начала алфавита A, B, C, D. Вероятность $\mathbb{P}(A)$.

4. Случайные величины. Заглавные английский буквы конца алфавита X, Y, W, Z. Математическое ожидание $\mathbb{E}(X)$.

Задачи:

1.1 В вазе пять неотличимых с виду конфет. Две без ореха и три — с орехом. Маша ест конфеты выбирая их наугад до тех пор, пока не съест первую конфету с орехом. Обозначим X — число съеденных конфет. Найдите $\mathbb{P}(X = 2)$, $\mathbb{P}(X > 1)$, $\mathbb{E}(X)$

1.2 В коробке находится четыре внешне одинаковые лампочки, две из них исправны. Лампочки извлекают из коробки по одной до тех пор, пока не будут извлечены обе исправные.

1. Какова вероятность того, что опыт закончится извлечением трёх лампочек?
2. Каково ожидаемое количество извлеченных лампочек?

1.3 Маша подкидывает монетку. Если в первый раз монетка выпала орлом, то Маша подкидывает монетку ещё один раз, если решкой — то ещё два раза. Больше Маша монетку не подкидывает! Пусть X — количество выпавших орлов.

Найдите вероятности $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$, ...и ожидание $\mathbb{E}(X)$.

1.4 Две команды равной силы играют в волейбол до трёх побед одной из них, не обязательно подряд. Ничья невозможна. Из-за равенства сил будем считать, что вероятность победы каждой равна 0.5. Величина N — количество сыгранных партий.

Составьте табличку возможных значений N с их вероятностями.

Найдите $\mathbb{P}(N$ — чётное$)$, $\mathbb{E}(N)$

1.5 Какова вероятность того, что у 30 человек не будет ни одного совпадения дней рождений? Сколько человек должно собраться, чтобы вероятность совпадения дней рождения превысила $1/2$? Сколько в среднем человек должно войти в комнату, чтобы впервые произошло совпадения дней рождения?

1.6 Саша и Маша по очереди подбрасывают кубик до первой шестёрки. Посуду будет мыть тот, кто первым выбросит шестерку. Маша бросает кубик первой. Какова вероятность того, что посуду будет мыть Маша? Сколько в среднем раз они будут бросать кубик?

1.7 Неправильную монетку с вероятностью «орла» равной 0.7 подбрасывают до первого «орла». Чему равно среднее количество подбрасываний? Орлов? Решек? Какова вероятность чётного числа бросков? Как изменятся ответы, если вероятность орла будет равна p?

1.8 Вы играете в следующую игру. Кубик подкидывается неограниченное число раз. Если на кубике выпадает 1, 2 или 3, то соответствующее количество монет добавляется на кон. Если выпадает 4 или 5, то игра оканчивается и Вы получаете сумму, лежащую на кону. Если выпадает 6, то игра оканчивается, а Вы не получаете ничего. Изначально на кону лежит ноль рублей.

1. Какова вероятность того, что игра рано или поздно закончится выпадением 6-ки?
2. Какова ожидаемая продолжительность игры?
3. Чему равен ожидаемый выигрыш?
4. Чему равен ожидаемый выигрыш, если изначально на кону лежит 100 рублей?
5. Изменим изначальное условие: если выпадает 5, то сумма на кону сгорает, а игра продолжается. Чему будет равен средний выигрыш в новую игру?

1.9 Саша и Маша подкидывают монетку до тех пор, пока не выпадет последовательность РОО или ООР. Если игра закончится выпадением РОО, то выигрывает Саша, если ООР, то — Маша. Случайная величина X — общее количество подбрасываний, Y — количество выпавших решек.

1. У кого какие шансы выиграть?
2. $\mathbb{P}(X = 4)$, $\mathbb{P}(Y = 1)$, $\mathbb{E}(X)$, $\mathbb{E}(Y)$
3. Решите аналогичную задачу для ОРО и ООР.

1.10 Вася подкидывает кубик до тех пор, пока на кубике не выпадет единица, или пока он сам не скажет «Стоп». Вася получает столько рублей, сколько выпало на кубике при последнем броске. Вася хочет максимизировать свой ожидаемый выигрыш.

1. Как выглядит оптимальная стратегия? Чему равен ожидаемый выигрыш при использовании оптимальной стратегии?
2. Какова средняя продолжительность игры при использовании оптимальной стратегии?
3. Как выглядит оптимальная стратегия и чему равен ожидаемый выигрыш, если за каждое подбрасывание Вася платит 35 копеек?

1.11 Саша и Маша решили, что будут заводить новых детей до тех пор, пока в их семье не будут дети обоих полов. Обозначим X — количество детей в их семье. Найдите $\mathbb{P}(X = 4)$, $\mathbb{E}(X)$.

1.12 В каждой вершине треугольника по ёжику. Каждую минуту с вероятностью 0.5 каждый ежик независимо от других двигается по часовой стрелке, с вероятностью 0.5 — против часовой стрелки. Обозначим T — время до встречи всех ежей в одной вершине.

1. Найдите $\mathbb{P}(T = 1)$, $\mathbb{P}(T = 2)$, $\mathbb{P}(T = 3)$, $\mathbb{E}(T)$.
2. Как изменятся ответы, если вероятность движения по часовой стрелке равна p?

2. Хочу ещё задач!

2.1 Наугад из четырех тузов разных мастей выбираются два. $\mathbb{P}(\text{они будут разного цвета})$?

2.2 События A и B несовместны, то есть не могут произойти одновременно. Известны вероятности $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.4$. Найдите[1] $\mathbb{P}(A^c \cap B^c)$.

2.3 Вероятность $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.8$. В каких пределах может лежать $\mathbb{P}(A \cap B)$?

2.4 Множество исходов $\Omega = \{a, b, c\}$, $\mathbb{P}(\{a, b\}) = 0,8$, $\mathbb{P}(\{b, c\}) = 0,7$. Найдите $\mathbb{P}(\{a\})$, $\mathbb{P}(\{b\})$, $\mathbb{P}(\{c\})$

[1] Событие A^c — это событие противоположное событию A, иногда обозначается $\bar{A}$

2.5 «Amoeba». A population starts with a single amoeba. For this one and for the generations thereafter, there is a probability of 3/4 that an individual amoeba will split to create two amoebas, and a 1/4 probability that it will die out without producing offspring. Let the random variable X be the number of generations before the death of all the amoebas. Find the probabilities $\mathbb{P}(X = 2)$, $\mathbb{P}(X = 3)$, $\mathbb{P}(X = \infty)$

2.6 Вася нажимает на пульте телевизора кнопку «On-Off» 100 раз подряд. Пульт старый, поэтому в первый раз кнопка срабатывает с вероятностью $\frac{1}{2}$, затем вероятность срабатывания падает. Какова вероятность того, что после всех нажатий телевизор будет включен, если сейчас он выключен?

2.7 Suppose the probability to get a head when throwing an unfair coin is p, what's the expected number of throwings in order to get two consecutive heads? The expected number of tails?

2.8 Вам предложена следующая игра. Изначально на кону 0 рублей. Раз за разом подбрасывается правильная монетка. Если она выпадает орлом, то казино добавляет на кон 100 рублей. Если монетка выпадает решкой, то все деньги, лежащие на кону, казино забирает себе, а Вы получаете красную карточку. Игра прекращается либо когда Вы получаете третью красную карточку, либо в любой момент времени до этого по Вашему выбору. Если Вы решили остановить игру до получения трех красных карточек, то Ваш выигрыш равен сумме на кону. При получении третьей красной карточки игра заканчивается и Вы не получаете ничего. Вы заинтересованы в максимальном среднем выигрыше.

1. Как выглядит оптимальная стратегия, если только что была получена вторая красная карточка? Чему равен средний выигрыш?
2. Как выглядит оптимальная стратегия, если только что была получена первая красная карточка?
3. Как выглядит оптимальная стратегия в исходной игре? Чему равен средний выигрыш?

2.9 Есть три комнаты. В первой из них лежит сыр. Если мышка попадает в первую комнату, то она находит сыр через одну минуту. Если мышка попадает во вторую комнату, то она ищет сыр две минуты и покидает комнату. Если мышка попадает в третью комнату, то она ищет сыр три минуты и покидает комнату. Покинув комнату, мышка выходит в коридор и выбирает новую комнату наугад, например, может зайти в одну и ту же. Сейчас мышка в коридоре. Сколько времени ей в среднем потребуется, чтобы найти сыр?

2.10 Илье Муромцу предстоит дорога к камню. От камня начинаются ещё три дороги. Каждая из тех дорог снова оканчивается камнем. И от каждого камня начинаются ещё три дороги. И каждые те три дороги оканчиваются камнем…И так далее до бесконечности. На каждой дороге живёт трёхголовый Змей Горыныч. Каждый Змей Горыныч бодрствует независимо от других с вероятностью (хм, Вы не поверите!) одна третья. У Василисы Премудрой существует Чудо-Карта, на которой видно, какие Змеи Горынычи бодрствуют, а какие — нет.

Какова вероятность того, что Василиса Премудрая сможет найти на карте бесконечный жизненный путь Ильи Муромца проходящий исключительно мимо спящих Змеев Горынычей?

2.11 У Пети — монетка, выпадающая орлом с вероятностью $p \in (0; 1)$. У Васи — с вероятностью $1/2$. Они одновременно подбрасывают свои монетки до тех пор, пока у них не окажется набранным одинаковое количество орлов. В частности, они останавливаются после первого подбрасывания, если оно дало одинаковые результаты. Сколько в среднем раз им придётся подбросить монетку?

2.12 Треугольник с вершинами $(0;0)$, $(2;0)$ и $(1;1)$. Внутри него случайным образом выбирается точка, X – абсцисса точки. Найдите $\mathbb{P}(X > 1)$, $\mathbb{P}(X \in [0.5;1])$, $\mathbb{E}(X)$

2.13 Треугольник с вершинами $(0;0)$, $(2;0)$ и $(2;1)$. Внутри него случайным образом выбирается точка, X – абсцисса точки. Найдите $\mathbb{P}(X > 1)$, $\mathbb{P}(X \in [0.5;1])$. Что больше, $\mathbb{E}(X)$ или 1?

2.14 Исследовательница Мишель подкидывает игральный кубик неограниченное количество раз и складывает выпадающие количества очков.

1. Чему примерно равна вероятность того, что однажды сумма в точности будет равна 123456789?
2. Чему точно равна указанная вероятность?

2.15 Упрямая исследовательница Мишель подбрасывает монетку до тех пор, пока количество орлов не окажется в точности равным удвоенному количеству решек. Монетка выпадает орлом с вероятностью p. Какова вероятность того, что Мишель будет подкидывать монетку вечно?

2.16 Исследовательница Мишель хочет встать утром с правой ноги с вероятностью $1/\sqrt{2}$, и с левой с вероятностью $1 - 1/\sqrt{2}$. Однако для проведения случайных экспериментов у неё есть только одна правильная монетка. Как с помощью правильной монетки ей добиться цели?

2.17 Докажите каждое равенство словами, без использования факториалов. Затем обобщите каждое равенство, добавив туда n, k и суммы.

1. $\binom{10}{7} = \binom{10}{3}$
2. $\binom{10}{3} + \binom{10}{4} = \binom{11}{4}$
3. $\binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 2^4$
4. $4\binom{10}{4} = 10\binom{9}{3}$
5. $\binom{10}{3}\binom{7}{2} = \binom{10}{2}\binom{8}{3}$
6. $1 \cdot \binom{5}{1} + 2 \cdot \binom{5}{2} + 3 \cdot \binom{5}{3} + 4 \cdot \binom{5}{4} + 5 \cdot \binom{5}{5} = 5 \cdot 2^4$
7. $\binom{2}{2} + \binom{3}{2} + \binom{4}{2} + \binom{5}{2} = \binom{6}{3}$
8. $\binom{4}{0}\binom{5}{3} + \binom{4}{1}\binom{5}{2} + \binom{4}{2}\binom{5}{1} + \binom{4}{3}\binom{5}{0} = \binom{9}{3}$
9. $\binom{4}{0}^2 + \binom{4}{1}^2 + \binom{4}{2}^2 + \binom{4}{3}^2 + \binom{4}{4}^2 = \binom{8}{4}$

2.18 Объясните каждую интерпретацию словами, без использования факториалов.

1. $\binom{10}{4}$ — количество строк из 4 единиц и 6 нулей
2. $\binom{7}{4}$ — количество строк из 4 единиц и 6 нулей, в которых единицы не идут подряд
3. $\binom{10}{4}$ — число способов раздать 4 яблока 10 людям, не дав никому больше одного
4. $\binom{13}{4}$ — число способов раздать 4 яблока 10 людям

3. К чёрту условности!

3.1 Имеется три монетки. Две «правильных» и одна — с «орлами» по обеим сторонам. Петя выбирает одну монетку наугад и подкидывает её два раза. Оба раза выпадает «орел». Какова условная вероятность того, что монетка «неправильная»?

3.2 Два охотника одновременно выстрелили в одну утку. Первый попадает с вероятностью 0.4, второй — с вероятностью 0.7 независимо от первого.

1. Какова вероятность того, что в утку попала ровно одна пуля?
2. Какова условная вероятность того, что утка была убита первым охотником, если в утку попала ровно одна пуля?

3.3 Кубик подбрасывается два раза. Найдите вероятность получить сумму равную 8, если при первом броске выпало 3.

3.4 Игрок получает 13 карт из колоды в 52 карты. Какова вероятность, что у него как минимум два туза, если известно, что у него есть хотя бы один туз? Какова вероятность того, что у него как минимум два туза, если известно, что у него есть туз пик?

3.5 В урне 7 красных, 5 желтых и 11 белых шаров. Два шара выбирают наугад. Какова вероятность, что это красный и белый, если известно, что они разного цвета?

3.6 В урне 5 белых и 11 черных шаров. Два шара извлекаются по очереди. Какова вероятность того, что второй шар будет черным? Какова вероятность того, что первый шар — белый, если известно, что второй шар — черный?

3.7 Примерно 4% коров заражены «коровьим бешенством». Имеется тест, который дает ошибочный результат с вероятностью 0,1. Судя по тесту, новая партия мяса заражена. Какова вероятность того, что она действительно заражена?

3.8 В школе три девятых класса, «А», «Б» и «В», одинаковые по численности. В «А» классе 30% обожают учителя географии, в «Б» классе — 40% и в «В» классе — 70%. Девятиклассник Петя обожает учителя географии. Какова вероятность того, что он из «Б» класса?

3.9 Ген карих глаз доминирует ген синих. Следовательно, у носителя пары bb глаза синие, а у носителя пар BB и Bb — карие. У диплоидных организмов (а мы такие :)) одна аллель наследуется от папы, а одна — от мамы. В семье у кареглазых родителей два сына — кареглазый и синеглазый. Кареглазый женился на синеглазой девушке. Какова вероятность рождения у них синеглазого ребенка?

3.10 Из колоды в 52 карты извлекается одна карта наугад. Являются ли события «извлечен туз» и «извлечена пика» независимыми?

3.11 Из колоды в 52 карты извлекаются по очереди две карты наугад. Являются ли события «первая карта — туз» и «вторая карта — туз» независимыми?

3.12 Известно, что $\mathbb{P}(A) = 0,3$, $\mathbb{P}(B) = 0,4$, $\mathbb{P}(C) = 0,5$. События A и B несовместны, события A и C независимы и $\mathbb{P}(B|C) = 0,1$. Найдите $\mathbb{P}(A \cup B \cup C)$.

3.13 У тети Маши — двое детей, один старше другого. Предположим, что вероятности рождения мальчика и девочки равны и не зависят от дня недели, а пол первого и второго ребенка независимы. Для каждой из четырех ситуаций найдите условную вероятность того, что у тёти Маши есть дети обоих полов.

1. Известно, что хотя бы один ребенок — мальчик.
2. Тетя Маша наугад выбирает одного своего ребенка и посылает к тете Оле, вернуть учебник по теории вероятностей. Это оказывается мальчик.
3. Известно, что старший ребенок — мальчик.
4. На вопрос: «А правда ли тетя Маша, что у вас есть сын, родившийся в пятницу?» тётя Маша ответила: «Да».

4. Use Julia/R/python/...or die!

4.1 Самая простая. Случайная величина N имеет пуассоновское распределение с $\lambda = 2$. С помощью симуляций оцените $\mathbb{E}(N^3)$, $\mathbb{P}(N \geqslant 4)$, $\mathbb{P}(N \geqslant 10 \mid N \geqslant 5)$, $\mathbb{E}(N \mid N \geqslant 5)$. Функция rpois может помочь :)

4.2 Случайные величины X_1, ..., X_5 имеют равномерное распределение на отрезке $[0; 1]$ и независимы. С помощью симуляций оцените $\mathbb{P}(\min\{X_1, \ldots, X_5\} > 0.2)$, $\mathbb{P}(\min\{X_1, \ldots, X_5\} > 0.2 \mid X_1 + X_2 < 0.5)$, $\mathbb{E}(\min\{X_1, \ldots, X_5\})$, $\mathbb{E}(\min\{X_1, \ldots, X_5\} \mid X_1 + X_2 < 0.5)$

4.3 Случайные величины X_1, X_2 независимы и обе имеют биномиальное распределение с параметрами $n = 16$, $p = 0.7$. Величина Y задана формулой $Y = X_1/(1 + X_2)$. С помощью симуляций оцените $\mathbb{P}(Y > 0.5)$, $\mathbb{E}(Y)$, $\mathbb{P}(Y > 0.5 \mid X_1 > 10)$, $\mathbb{E}(Y \mid X_1 > 10)$. Функция rbinom в помощь!

4.4 В колоде 52 карты. Мы вытаскиваем карты из колоды до первого туза, пусть X — количество вытянутых карт. С помощью симуляций оцените $\mathbb{E}(X^2)$, $\mathbb{P}(X > 10)$, $\mathbb{P}(X > 5 \mid X < 15)$, $\mathbb{E}(X^2 \mid X < 15)$

4.5 Иван Федорович Крузенштерн случайным образом с возможностью повторов выбирает 10 натуральных чисел от 1 до 100. Пусть X — минимум этих чисел, а Y — максимум. С помощью симуляций оцените $\mathbb{P}(Y > 3X)$, $\mathbb{E}(XY)$, $\mathbb{P}(Y > 3X \mid Y < X^2)$, $\mathbb{E}(XY \mid Y < X^2)$

5. Эф большое и эф малое

Минитеория:

1. $F_X(t) = \mathbb{P}(X \leqslant t)$
2. $F(t) = \int_{-\infty}^{t} f(a)\,da$, $f(t) = F'(t)$.

Задачи:

5.1 Функция плотности случайной величины X равна 5 при $x = 7$. Найдите примерно вероятность того, что X попадёт в отрезок $[7; 7.001]$.

5.2 Случайная величина X имеет функцию плотности f. С помощью $o(\Delta)$ и $f(x)$ запишите вероятность $\mathbb{P}(X \in [x; x + \Delta])$.

5.3 Величина X распределена на отрезке $[0; 1]$ и на нём имеет функцию плотности $f(t) = 3t^2$. Найдите функции плотности и функции распределения величин $Y = \ln X$, $Z = X^2$, $W = (X - 0.5)^2$.

5.4 Может ли функция плотности принимать значение больше 2015? Может ли предел $\lim_{x\to\infty} f(x)$ не равняться нулю?

5.5 Функция плотности случайной величины X имеет вид:

$$f(t) = \begin{cases} \frac{3}{16}t^2, t \in [-2; 2] \\ 0, t \notin [-2; 2] \end{cases}$$

Найдите:

1. $\mathbb{P}(X > 1)$, $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathrm{Var}(X)$, σ_X
2. $\mathbb{E}(X|X > 1)$, $\mathbb{E}(X^2|X > 1)$, $\mathrm{Var}(X|X > 1)$
3. Функцию распределения случайной величины X
4. Медиану величины X, 40%-ую квантиль величины X

5.6 Функция плотности случайной величины X имеет вид:

$$f(t) = \begin{cases} t/8, t \in [0; 4] \\ 0, t \notin [0; 4] \end{cases}$$

Найдите:

1. $\mathbb{P}(X > 1)$, $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathrm{Var}(X)$, σ_X
2. $\mathbb{E}(X|X > 1)$, $\mathbb{E}(X^2|X > 1)$, $\mathrm{Var}(X|X > 1)$
3. Функцию распределения случайной величины X
4. Медиану величины X, 40%-ую квантиль величины X

5.7 Величина X распределена на отрезке $[0; 2]$ и имеет на нём функцию распределения $F(x) = x^2/4$. Найдите $\mathbb{P}(X \in [1; 1.5])$, $\mathbb{P}(X < 1)$, $F(-5)$, $F(10)$, функцию плотности величины X

5.8 Если возможно, найдите функцию распределения и функцию плотности величины X принимающей значения 1, 2, 3 и 4 с вероятностями 0.1, 0.2, 0.3 и 0.4 соответственно.

5.9 Величина X равномерна на отрезке $[-2; 1]$, а величина Y — это расстояние от числа X до числа (-1). Найдите фукнцию плотности Y, $\mathbb{E}(Y)$.

5.10 Глафира случайным образом равномерно выбирает случайную внутри треугольника с вершинами $(0; 0)$, $(0; 2)$ и $(3; 3)$. Пусть X и Y — абсцисса и ордината выбранной точки. Найдите функцию плотности X, функцию плотности Y.

5.11 Прямой убыток от пожара в миллионах рублей равномерно распределен на $[0; 1]$. Если убыток оказывается больше 0.7, то страховая компания выплачивает компенсацию 0.7.

1. Найдите функцию распределения потерь от пожара.
2. Чему равны средние потери?

5.12 Пусть X — неотрицательная случайная величина с функцией плотности $f(t)$ и $\mathbb{E}(X) < \infty$. При каком c функция $g(t) = c \cdot t \cdot f(t)$ также будет функцией плотности?

5.13 Завтрашняя цена акции — случайная величина с функцией плотности $f(x) = \frac{3}{4} \max\{x(2-x), 0\}$.

1. Постройте график функции плотности;
2. Найдите функцию распределения Васиного дохода, средний доход и дисперсию дохода, если:
   а) У Васи есть 10 акций;
   б) У Васи нет акций, но есть опцион-пут, дающий ему право продать одну акцию по цене 1.2 рубля;
   в) У Васи нет акций, но есть опцион-колл, дающий ему право купить одну акцию по цене 1 рубль.

5.14 В соревнованиях по прыжкам в длину участвовали n спортсменов. Результаты их прыжков, величины X_i, независимы и одинаково распределены с функцией плотности f и функцией распределения F.

1. Найдите функцию распределения и функцию плотности длины наилучшего прыжка
2. Найдите функцию распределения и функцию плотности длины наихудшего прыжка
3. Найдите вероятность того, что Сидоров и Петров прыгнули меньше чем на t метров, Иванов прыгнул от t до $t + \Delta$ метров, а остальные прыгнули больше, чем на t метров.
4. Найдите функцию распределения и функцию плотности длины прыжка бронзового призёра соревнований

5.15 Светофор попеременно горит пешеходу то 40 секунд зелёным, то 60 секунд — красным. Законопослушный Вася Бубликов подходит к светофору в случайный момент времени. Пусть X — время ожидания до возможности перейти дорогу.

1. Найдите функцию распределения величины X и постройте её график
2. Найдите $\mathbb{E}(X)$ и $\mathrm{Var}(X)$

5.16 Большой Адронный Коллайдер запускают ровно в полночь. Оставшееся время до Конца Света — случайная величина X распределенная равномерно от 0 до 16 часов. Когда произойдет Конец Света, механические часы остановятся и будут показывать время Y.

1. Найдите $\mathbb{P}(Y < 2)$
2. Постройте функцию плотности величины Y
3. Найдите $\mathbb{E}(Y)$, $\mathrm{Var}(Y)$
4. Найдите $\mathrm{Cov}(X, Y)$

6. Рождение распределений

6.1 «Всякая вещь — или есть, или нет»

Как справедливо заметил исследователь Винни-Пух, всякая вещь — или есть, или нет. Рассмотрим случайную величину V, равную единице, если вещь — есть, и нулю иначе. Допустим вероятность того, что вещь есть равна 0.7.

1. Составьте табличку со значениями V и их вероятностями;
2. Найдите $\mathbb{E}(V)$ и $\mathrm{Var}(V)$;
3. Если возможно, нарисуйте функцию распределения и функцию плотности V;
4. Найдите $\mathbb{E}(V)$ и $\mathrm{Var}(V)$, если вероятность того, что вещь — есть равна параметру p.

Определение. Случайная величина V имеет распределение Бернулли с параметром p.

6.2 Винни-Пух и Пятачок играют в Пустяки (Poohsticks). Каждую партию Винни-Пух выигрывает с вероятностью $p = 0.7$. Всего они сыграли $n = 10$ партий. Пусть X — количество партий, выигранных Винни-Пухом.

1. Объясните правила игры Пустяки;
2. Найдите $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 10)$, $\mathbb{P}(X = 4)$, $\mathbb{P}(X = k)$;
3. Представьте X в виде суммы 10 случайных величин с распределением Бернулли, поясните, что означает каждое слагаемое.
4. Найдите $\mathbb{E}(X)$ и $\mathrm{Var}(X)$. Здесь без доказательства можно пользоваться тем, что для независимых величин R и S дисперсия раскладывается в сумму $\mathrm{Var}(R+S) = \mathrm{Var}(R) + \mathrm{Var}(S)$;
5. Найдите $\mathbb{P}(X = k)$, $\mathbb{E}(X)$ и $\mathrm{Var}(X)$ для произвольных p, n;

Определение. Случайная величина X имеет биномиальное распределение с параметрами n и p;

6.3 «Ну не то, чтобы совсем не попал...»

Храбрый Пятачок спешит с ружьём на помощь зависшему исследователю Винни-Пуху. При каждом выстреле Пятачок попадает в шарик с вероятностью $p = 0.7$ независимо от предыдущих выстрелов. Стреляет Пятачок до первого попадания. Пусть X — это количество выстрелов, Y — количество промахов по шарику[2].

1. Найдите $\mathbb{P}(X = 2)$, $\mathbb{P}(X = 3)$, $\mathbb{P}(X = k)$.
2. Терпения Винни-Пуха хватает на 5 промахов Пятачка. Какова вероятность того, что терпения Винни-Пуха не хватит?
3. Найдите $\mathbb{P}(Y = 2)$, $\mathbb{P}(Y = 3)$, $\mathbb{P}(Y = k)$.
4. С помощью метода первого шага найдите $\mathbb{E}(X)$ и $\mathbb{E}(X^2)$;
5. Найдите $\mathrm{Var}(X)$
6. Каким соотношением связаны X и Y?
7. Как связаны $\mathbb{E}(X)$ и $\mathbb{E}(Y)$, $\mathrm{Var}(X)$ и $\mathrm{Var}(Y)$?
8. Чему будут равны $\mathbb{P}(X = k)$, $\mathbb{E}(X)$ и $\mathrm{Var}(X)$, если при отдельном выстреле Пятачок попадает с вероятностью p?

[2] или попаданий по исследователю Винни-Пуху

Определение. Величина X имеет геометрическое распределение с параметром p.

6.4 Торопливый Пятачок

Торопливый Пятачок снова стреляет в шарик. При каждом выстреле торопливый Пятачок попадает с вероятностью p. Пятачок делает d выстрелов в минуту. Для данной задачи будем считать, что после попадания Пятачка в шарик, шарик мгновенно заменяется на новый. Пусть X — это номер выстрела первого попадания, Z — время первого попадания, а Y — количество попаданий за первую минуту.

1. Сколько раз в минуту в среднем попадает Пятачок, если вероятность попадания при отдельном выстреле равна $p = 0.7$?
2. Чему равна вероятность попадания при отдельном выстреле, если среднее количество попаданий в миниту равно $\lambda = 10$?
3. Найдите $\mathbb{P}(X \leqslant k)$ для произвольного параметра p;
4. Найдите $\mathbb{P}(Z \leqslant t)$ для произвольного параметра p;
5. Найдите $\mathbb{P}(Y = 0)$, $\mathbb{P}(Y = k)$, $\mathbb{E}(Y)$, $\mathrm{Var}(Y)$;

Торопливый Пятачок очень торопится, поэтому величина d крайне велика. Однако когда Пятачок торопится, он чаще промахивается, и оказывается, что при любом d среднее количество попаданий в минуту постоянно и равно $\lambda = 10$.

6. Чему при большом d равны $\mathbb{P}(Y = 0)$, $\mathbb{P}(Y = k)$, $\mathbb{E}(Y)$, $\mathrm{Var}(Y)$?
7. Найдите $F(t) = \lim_{d\to\infty} \mathbb{P}(Z \leqslant t)$;
8. Найдиту дифференциальную форму dF и функцию плотности $f = F'$;
9. Какова при большом d вероятность того, что Пятачок впервые попадёт в первые 5 минут? Впервые попадёт с 10-ой по 15-ую минуты, если известно, что в первые 10 минут попаданий не было?
10. Найдите $F(t)$, dF и f для произвольного λ;
11. Чему при большом d равны $\mathbb{P}(Y = 0)$, $\mathbb{P}(Y = k)$, $\mathbb{E}(Y)$, $\mathrm{Var}(Y)$ для произвольного λ?

Определение. Случайная величина Z при $d \to \infty$ имеет экспоненциальное распределение с параметром λ.

Определение. Случайная величина Y при $d \to \infty$ имеет распределение Пуассона с параметром λ.

7. Разлагай и властвуй!

7.1 Из грота ведут 10 штреков, с длинами 100м, 200м, …1000м. Самый длинный штрек оканчивается выходом на поверхность. Остальные — тупиком. Вася выбирает штреки наугад, в тупиковый штрек два раза не ходит. Какова вероятность того, что Вася посетит самый короткий штрек? Какой в среднем путь он нагуляет прежде чем выберется на поверхность?

7.2 У Маши 30 разных пар туфель. И она говорит, что мало! Пёс Шарик утащил без разбору на левые и правые 17 туфель. Какова вероятность того, что у Маши останется ровно 13 полных пар? Величина X — количество полных целых оставшихся пар, Y — количество полных пар, доставшихся Шарику. Найдите $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\operatorname{Var}(X)$, $\operatorname{Var}(Y)$.

7.3 У меня в кармане 3 рубля мелочью. Среди монет всего одна монета достоинством 50 копеек. Я извлекаю монеты по одной наугад. Я останавливаюсь после того, как извлеку монету в 50 копеек. Какую сумму в среднем я извлеку?

7.4 «Модница». В шкатулке у Маши 100 пар серёжек. Каждый день утром она выбирает одну пару наугад, носит ее, а вечером возвращает в шкатулку. Проходит год.

1. Сколько в среднем пар окажутся ни разу не надетыми?
2. Сколько в среднем пар окажутся надетыми не менее двух раз?

(c*) Как изменятся ответы, если каждый день Маша покупает себе новую пару серёжек и вечером добавляет её в шкатулку?

7.5 Вовочка получает пятерку с вероятностью 0.1, четверку — с вероятностью 0.2, тройку — с вероятностью — 0.3 и двойку с вероятностью 0.4. В этом четверти он писал 20 контрольных. Какова вероятность того, что все оценки у Вовочки одинаковые? Сколько разных оценок он в среднем получит?

7.6 «Судьба Дон Жуана» У Васи n знакомых девушек (их всех зовут по-разному). Он пишет им n писем, но, по рассеянности, раскладывает их в конверты наугад. Величина X обозначает количество девушек, получивших письма, написанные лично для них. Найдите $\mathbb{E}(X)$, $\operatorname{Var}(X)$.

7.7 Над озером взлетело 20 уток. Каждый из 10 охотников один раз стреляет в случайно выбираемую им утку. Величина Y — количество убитых уток, X — количество попавших в цель охотников. Найдите $\mathbb{E}(X)$, $\operatorname{Var}(X)$, $\mathbb{E}(Y)$, $\operatorname{Var}(Y)$, если охотники стреляют без промаха. Как изменится ответ, если вероятность попадания равна 0.7?

7.8 Вокруг новогодней ёлки танцуют хороводом 27 детей. Мы считаем, что ребенок высокий, если он выше обоих своих соседей. Величина X — количество высоких детей в хороводе. Найдите $\mathbb{E}(X)$, $\operatorname{Var}(X)$. Вероятность совпадения роста будем считать равной нулю.

7.9 По 10 коробкам наугад раскладывают 7 карандашей. Каково среднее количество пустых коробок? Дисперсия?

7.10 Внутри каждой упаковки шоколадки находится наклейка с изображением одного из 30 животных. Предположим, что все наклейки равновероятны, величина X — это количество шоколадок, которые купить, чтобы собрать полную коллекцию наклеек. Чему равны $\mathbb{E}(X)$, $\operatorname{Var}(X)$? Как это объяснить ребёнку?

7.11 Из колоды в 52 карты извлекается 5 карт. Сколько в среднем извлекается мастей? Достоинств? Тузов? Дисперсии этих величин?

7.12 За круглым столом сидят в случайном порядке n супружеских пар, всего — $2n$ человек. Величина X — число пар, где супруги оказались напротив друг друга. Найдите $\mathbb{E}(X)$ и $\mathrm{Var}(X)$

7.13 В задачнике N задач. Из них a — Вася умеет решать, а остальные не умеет. На экзамене предлагается равновероятно выбираемые n задач. Величина X — число решенных Васей задач на экзамене. Найдите $\mathbb{E}(X)$ и $\mathrm{Var}(X)$

7.14 Кубик подбрасывается n раз. Величина X_1 — число выпадений 1, а X_6 — число выпадений 6. Найдите $\mathrm{Corr}(X_1, X_6)$

Пуассоновский поток событий. Обозначим: $X[a; a+\Delta]$ — количество происшествий на интервале $[a; a+\Delta]$, $X_t = X[0; t]$ — количество происшествий за период $[0; t]$.

Если:

1. На малом интервале времени вероятность одного происшествия примерно пропорциональна длине интервала, $\mathbb{P}(X[a; a+\Delta] = 1) = \lambda\Delta + o(\Delta)$.

2. На малом интервале времени несколько происшествий происходят существенно реже одного происшествия, $\mathbb{P}(X[a; a+\Delta] \geqslant 2) = o(\Delta)$.

3. Стационарность приращений. Распределение случайной величины $X[a; a+\Delta]$, количества происшествий на интервале $[a; a+\Delta t]$, зависит только от Δ, но не от a.

4. Независимость приращений. Количество происшествий на непересекающихся интервалах времени независимы.

То:

1. Время между $(i-1)$-ым и i-ым происшествием, Y_i, имеет экспоненциальное распределение $Y_i \sim exp(\lambda)$.
$$f(y) = \begin{cases} \lambda e^{-\lambda y},\ y \geqslant 0 \\ 0,\ y < 0 \end{cases}$$
В частности, $\mathbb{E}(Y_i) = 1/\lambda$ и $\mathrm{Var}(Y_i) = 1/\lambda^2$. Величины Y_i независимы.

2. Количество происшествий за единицу времени, X, имеет пуассоновское распределение $X \sim Pois(\lambda)$
$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$
В частности, $\mathbb{E}(X) = \lambda$ и $\mathrm{Var}(X) = \lambda$

   Отсюда смысл λ — среднее количество событий за единицу времени, дисперсия количества событий за единицу времени

3. Количество событий за период времени $[0; t]$, величина X_t, имеет пуассоновское распределение $X_t \sim Pois(\lambda t)$
$$\mathbb{P}(X_t = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$
И, следовательно, $\mathbb{E}(X_t) = \lambda t$, $\mathrm{Var}(X_t) = \lambda t$.

4. Сумма двух независимых пуассоновских процессов с интенсивностями λ_1 и λ_2 — пуассоновский процесс с интенсивностью $\lambda_1 + \lambda_2$

Замена $Bin(n, p)$ на $Pois(\lambda = np)$ дает погрешность не более $\min\{p, np^2\}$

8. За время моего дежурства происшествий не было!

8.1 Маша и Саша пошли в лес по грибы. Саша собирает все грибы, а Маша — только подберезовики. Саша в среднем находит один гриб за одну минуту, Маша — один гриб за десять минут. Какова вероятность того, за 8 минут они найдут ровно 13 грибов? Какова вероятность того, что следующий гриб им попадется позже, чем через минуту, если Маша только что нашла подберезовик?

8.2 Пост майора ГИБДД Иванова И.И. в среднем ловит одного нарушителя в час. Какова вероятность того, за первые полчаса дежурства будет не меньше двух нарушителей? Какова вероятность того, что следующего нарушителя ждать еще более 40 минут, если уже целых три часа никто не превышал скорость?

8.3 Оля и Юля пишут смс Маше. Оля отправляет Маше в среднем 5 смс в час. Юля отправляет Маше в среднем 2 смс в час. Какова вероятность того, что Маша получит ровно 6 смс за час? Сколько времени в среднем проходит между смс, получаемыми Машей от подруг?

8.4 Кузнечики на большой поляне распределены по пуассоновскому закону, в среднем 3 кузнечика на квадратный метр. Какой следует взять сторону квадрата, чтобы вероятность найти в нем хотя бы одного кузнечика была равна $0,8$?

8.5 В магазине две кассирши (ах, да! две хозяйки кассы). Допустим, что время обслуживания клиента распределено экспоненциально. Тетя Зина обслуживает в среднем 5 клиентов в час, а тетя Маша - 7. Два клиента подошли к кассам одновременно.

1. Какова вероятность того, что тетя Зина обслужит клиента быстрее?
2. Как распределено время обслуживания того клиента, который освободится быстрее?
3. Каково условное среднее время обслуживания клиента тетей Зиной, если известно, что она обслужила клиента быстрее тети Маши?

8.6 Время между приходами студентов в столовую распределено экспоненциально; в среднем за 10 минут приходит 5 студентов. Время обслуживания имеет экспоненциальное распределение; в среднем за 10 минут столовая может обслужить 6 студентов. Столовая находится в динамическом равновесии, то есть закон распределения длины очереди стабилен (это не означает, что длина очереди не меняется).

1. Какова вероятность того, что в очереди ровно n студентов?
2. Какова средняя длина очереди?

Подсказка: если сейчас в очереди n человек, то через малый промежуток времени Δ...

8.7 The arrival of buses at a given bus stop follows Poisson law with rate 2. The arrival of taxis at the same bus stop is also Poisson, with rate 3. What is the probability that next time I'll go to the bus stop I'll see at least two taxis arriving before a bus? Exactly two taxis?

8.8 Время, которое хорошо обученная свинья тратит на поиск трюфеля — экспоненциальная случайная величина со средним в 10 минут. Какова вероятность того, что свинья за 20 минут не найдет ни одного трюфеля?

8.9 Величина X распределена экспоненциально с параметром λ, а константа $a > 0$. Как распределена величина $Y = aX$?

8.10 В гирлянде 25 лампочек. Вероятность брака для отдельной лампочки равна 0,01. Какова вероятность того, что гирлянда полностью исправна? Оцените точность ответа при использовании распределения Пуассона.

8.11 По некоему предмету незачет получило всего 2% студентов. Какова вероятность того, что в группе из 50 студентов будет ровно 1 человек с незачетом? Оцените точность ответа при использовании распределения Пуассона.

8.12 Вася испек 40 булочек. В каждую из них он кладет изюминку с $p = 0,02$. Какова вероятность того, что всего окажется 3 булочки с изюмом? Оцените точность ответа при использовании распределения Пуассона.

8.13 В офисе два телефона — зеленый и красный. Входящие звонки на красный — Пуассоновский поток событий с интенсивностью $\lambda_1 = 4$ звонка в час, входящие на зеленый – с интенсивностью $\lambda_2 = 5$ звонка в час. Секретарша Василиса Премудрая одна в офисе. Перед началом рабочего дня она подбрасывает монетку и отключает один из телефонов, зеленый — если выпала решка, красный — если орел. Обозначим Y_1 время от начала дня до первого звонка.

1. Найдите функцию плотности Y_1
2. Верно ли, что процесс количества звонков, которые услышит Василиса, имеет независимые приращения?

8.14 Владелец салуна «Огненная зебра» закрывает заведение, если в течение 5 минут никто не заказывает виски. Посетители заказывают в среднем один виски в минуту. Заказы представляют собой пуассоновский поток. Пусть X — время от открытия до закрытия таверны. Найдите $\mathbb{E}(X)$, $\mathrm{Var}(X)$.

8.15 Рассмотрим определение пуассоновского процесса, а именно: вероятность ровно одного события за интервал времени $[a; a + \Delta]$ есть $\lambda\Delta + o(\Delta)$, вероятность не менее двух событий за тот же интервал времени есть $o(\Delta)$. Докажите, что время между событиями имеет экспоненциальное распределение.

8.16 Количество трапперов, заходящих в салун «Огненная зебра», — пуассоновский поток с единичной интенсивностью. Какова вероятность того, что через время t от момента открытия в салун зайдёт чётное количество трапперов?

8.17 На Краю Вселенной давным-давно работает парикмахерская. В ней счётное количество занумерованных по порядку парикмахеров. Каждый парикмахер независимо от других обслуживает клиента за экспоненциально распределенное время с параметром λ. Клиенты приходят в парикмахерскую пуассоновским потоком с интенсивностью μ. Клиент всегда выбирает свободного парикмахера с наименьшим номером.

Какую долю времени будет в среднем занят парикмахер номер n?

8.18 Множество A состоит из n элементов. На пленэр Маэстро захватил с собой случайное количество красок, X, распределённое по Пуассону с параметром $\lambda = 1$. Маэстро работает в новаторской методике и никогда не смешивает краски!

1. Сколькими способами Маэстро может раскрасить множество A, если можно любой элемент красить в любой цвет?
2. Маэстро разрезал множество A на k подмножеств, «я так вижу!». Сколькими способами Маэстро может раскрасить множество A, если каждое подмножество он хочет раскрасить в свой оригинальный цвет?

   Это количество обозначается $(X)_k$ и называется *убывающим факториалом* :)

3. Найдите $\mathbb{E}((X)_k)$.
4. Как связаны между собой $\mathbb{E}(X^n)$ и число Белла B_n, количество различных разбиений множества из n элементов?

Результат называется формулой Добинского и был опубликован в 1877 году :)

9. Пиастры, пиастры!

9.1 Функция SHA256 превращает произвольное текстовое сообщение в последовательность длины 256 из нулей и единиц. Например, SHA256("Люблю вероятности")= 000011101001001.... Результат вычисления функции называется хэшем. Функция детерминистическая, но её внутренне устройство настолько сложно, что явное обращение функции невозможно. Другими словами, если я хочу найти фразу, которую функция SHA256 превращает в последовательность 010101010101..., то никакого способа кроме перебора всех фраз у меня нет[3]. Никаких простых закономерностей в результате вычисления функции нет.

Для внесения блока из нескольких сделок в блокчейн биткойна и многих других криптовалют необходимо добавить после блока сделок бессмысленный текст так, чтобы функция SHA256 от всего текста начиналась с заданного количества нулей.

1. Какова вероятность того, что дописав произвольный текст к блоку сделок майнер Мария получит хэш, начинающийся с 30 нулей?
2. Какова вероятность того, что перебрав 1000 бессмысленных текстовых прибавок, майнер Мария получит хэш, начинающийся с 30 нулей?
3. Почему хэш можно трактовать как равномерную на отрезке $[0;1]$ случайную величину?

Видео про блокчейн от 3blue1brown, https://youtu.be/bBC-nXj3Ng4. Статья, https://bitcoin.org/bitcoin.pdf.

9.2 Компьютер майнера Марии может перебирать 10^5 хэшей SHA256 в секунду. Других майнеров в сети нет. Обозначим время до получения первого хэша, начинающегося с 30 нулей, буквой T. Подписав блок сделок, Мария сразу переходит к подписанию очередного блока.

5. Как распределена величина T?
6. Чему равны $\mathbb{E}(T)$, $\mathrm{Var}(T)$?
7. Какова вероятность того, что майнер Мария получит нужный хэш быстрее, чем за полчаса?
8. Сколько блоков майнер Мария в среднем может подписать за сутки? Какова дисперсия этой величины?
9. Сколько нулей должно требоваться, чтобы на подписание одного блока майнер Мария тратила больше суток с вероятностью 0.95?

9.3 Для подписания блока в блокчейне требуется хэш SHA256, начинающийся с 30 нулей. Компьютер майнера Виктории может перебирать 10^5 хэшей в секунду, а компьютер майнера Кристины — $2 \cdot 10^5$ хэшей в секунду. Других майнеров в сети нет. Виктория и Кристина одновременно и независимо друг от друга приступили к подписанию первого блока. После того, как кто-то подпишет блок, Виктория и Кристина одновременно приступают к подписанию одного и того же нового блока. Обозначим B_V и B_K — количество блоков, подписанных за час работы Викторией и Кристиной соответственно.

[3] Возможно он есть, но вряд ли, пока лучше домашку по вероятностям делать :)

1. Какова вероятность того, что Виктория подпишет первый блок раньше Кристины?
2. Как распределена величина B_V? Найдите $\mathbb{E}(B_V)$ и $\text{Var}(B_V)$;
3. Найдите $\text{Cov}(B_V, B_K)$;

9.4 Для подписания блока в блокчейне требуется хэш SHA256, начинающийся с 30 нулей. Компьютер майнера Виктории может перебирать 10^5 хэшей в секунду, а компьютер майнера Кристины — $2 \cdot 10^5$ хэшей в секунду. Кроме Виктории и Кристины в сети находятся прочие майнеры с совокупной мощностью 10^7 хэшей в секунду. Все майнеры одновременно и независимо друг от друга работают над одним и тем же блоком. Когда блок подписан, все майнеры начинают работу над одним и тем же новым блоком. За один блок майнер получает награду в 10 тугрикойнов. Обозначим S_V и S_K — заработок Виктории и Кристины за месяц, а $S = S_V + S_K$.

1. Найдите $\mathbb{E}(S_V)$, $\text{Var}(S_V)$, $\mathbb{E}(S_K)$, $\text{Var}(S_V)$, $\mathbb{E}(S)$, $\text{Var}(S)$;
2. Как изменятся все найденные величины, если Виктория и Кристина объединятся в пул? При объединении в пул награду за каждый подписанный блок Виктория и Кристина будут делить пропорционально мощностям своих компьютеров.

9.5 Полученные в блоке деньги не сразу считаются подтвержденными: их можно тратить, только если за данным блоком в блокчейне подписано 50 других блоков. Красивая, умная и состоятельная мошенница Василиса сосредоточила в своих руках 40% имеющихся вычислительных мощностей сети. Хитрая Василиса решила дважды потратить свой миллион тугрикойнов.

Она создала сделку в которой миллион тугрикойнов уходят с её счета продавцу А. Сразу после того, как блок с этой сделкой только что был подписан, хитрая Василиса в тайне начала рассчитывать ложную параллельную ветку блоков. В ложной ветке миллион тугрикойнов уходят другому продавцу Б за другой товар. Все майнеры кроме Василисы подписывают реально существующие сделки. Только что сделка с продавцом А была признана подтверждённой.

1. Как изменилось ожидаемое количество честных блоков подписанных сетью за единицу времени во время атаки Василисы?
2. Какова вероятность того, что Василиса сможет рано или поздно предъявить ложную последовательность блоков более длинную чем честная?

Для упрощения можно предполагать, что на подсчёт честной цепочки блоков к моменту подтверждения сделки с продавцом А ушло ожидаемое для этого время.

9.6 Совокупные вычислительные возможности сети — 10^7 хэшей в секунду. Будем трактовать каждый хэш как равномерную на отрезке $[0; 1]$ случайную величину. Блок сделок считается подписанным, если полученный хэш оказался меньше константы p. Какой должна быть константа p, чтобы новый блок подписывался в среднем каждые 10 минут?

10. Совместная плотность и вероятность

Минитеория:

1. $f(x|y) = \frac{f(x,y)}{f(y)}$
2. $f(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$

3. $\mathbb{E}(g(X,Y)) = \int\int g(x,y)\cdot f(x,y)\,dx\,dy$
4. Если $f(x,y) = f(x)\cdot f(y)$, то величины X и Y независимы

10.1. Совместная вероятность

10.1 Эксперимент может окончиться одним из шести исходов:

	$X=-1$	$X=0$	$X=2$
$Y=0$	0.1	0.2	0.3
$Y=4$	0.2	0.1	0.1

Найдите: $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $\mathrm{Cov}(X,Y)$, $\mathrm{Corr}(X,Y)$, $\mathrm{Cov}(X,Y|X\geqslant 0)$

10.2 Кубик подбрасывается два раза, X — сумма очков, Y — разность очков, число при первом броске минус число при втором. Найдите $\mathbb{E}(XY)$, $\mathrm{Cov}(X,Y)$, $\mathrm{Corr}(X,Y)$

10.3 Пусть X равновероятно принимает значения -1, 0, +1, а $Y = X^2$

1. Найдите $\mathrm{Cov}(X,Y)$;
2. Верно ли, что X и Y независимы?

10.4 Паук сидит в начале координат. Равновероятно он может сместиться на единицу вверх, вниз, влево или вправо (по диагонали паук не ползает). Пусть X и Y — это абсцисса и ордината паука после первого шага.

1. Найдите $\mathrm{Cov}(X,Y)$?
2. Верно ли, что X и Y независимы?

10.5 Кубик подбрасывается n раз. Пусть X_1 — число выпадений 1, а X_6 — число выпадений 6.

Найдите $\mathrm{Corr}(X_1, X_6)$.

Подсказка: $\mathrm{Cov}(X_1, X_1 + \ldots + X_6) = \ldots$

10.6 Вероятность дождя в субботу 0.5, вероятность дождя в воскресенье 0.3. Корреляция между наличием дождя в субботу и наличием дождя в воскресенье равна r.

Какова вероятность того, что в выходные вообще не будет дождя?

10.7 Пусть $\mathbb{P}(A|B) > \mathbb{P}(A)$. Можно ли определить знак $\mathrm{Cov}(1_A, 1_B)$?

10.2. Свойства ковариаций и корреляций

10.8 Известно, что $Y = 2X - 3$, а $Z = 6 - 3X$. Найдите $\mathrm{Corr}(X,Y)$, $\mathrm{Corr}(X,Z)$

10.9 Пусть X и Y независимы.

1. Найдите $\mathrm{Cov}(X,Y)$, $\mathrm{Cov}(X^3, Y^2 - 5Y)$, $\mathrm{Corr}(X,Y)$
2. Выразите $\mathrm{Var}(X+Y)$ и $\mathrm{Var}(X-Y)$ через $\mathrm{Var}(X)$ и $\mathrm{Var}(Y)$

10.10 Вася наблюдает значение X, но не наблюдает значение Y; при этом он знает, что $\mathrm{Var}(X) = 3$, $\mathrm{Var}(Y) = 8$, $\mathrm{Cov}(X,Y) = -3$, $\mathbb{E}(Y) = 3$, $\mathbb{E}(X) = 2$. Задача Васи — спрогнозировать Y с помощью линейной функции от X, т.е. построить $\hat{Y} = aX + b$.

Васю штрафуют за неправильный прогноз на сумму $(Y - \hat{Y})^2$. Найдите a и b

10.11 Случайные величины X и Y зависимы, случайные величины Y и Z зависимы. Верно ли, что случайные величины X и Z зависимы?

10.12 Пусть $\mathrm{Cov}(X,Y) > 0$, $\mathrm{Cov}(Y,Z) > 0$. Верно ли, что $\mathrm{Cov}(X,Z) > 0$? $\mathrm{Cov}(X,Z) \geqslant 0$?

10.3. Совместная плотность

10.13 Величины X_1, X_2, X_3, X_4 — независимы и равномерны $U[0;1]$. Величины Y_1, Y_2, Y_3 и Y_4 — это величины $\{X_i\}$, упорядоченные по возрастанию. Например, $Y_1 = \min\{X_1, X_2, X_3, X_4\}$.

1. Найдите $\mathbb{P}(X_1 \in [x; x + dx])$.
2. Найдите вероятность $\mathbb{P}(Y_1 \in [y_1; y_1 + dy_1])$ с точностью до $o(dy_1)$. Найдите функцию плотности Y_1;
3. Найдите примерные вероятности попадания в отрезок малой длины и функции плотности для Y_2, Y_3 и Y_4.
4. Найдите $\mathbb{P}(Y_1 \in [y_1; y_1 + dy_1]), Y_3 \in [y_3; y_3 + dy_3]$. Найдите совместную функцию плотности пары Y_1, Y_3;
5. Найдите совместную функцию плотности пары Y_1, Y_4;
6. Найдите совместную функцию плотности пары Y_2, Y_4;
7. Найдите условные плотности $f(y_1|y_4)$, $f(y_1|y_3)$, $f(y_2|y_4)$.

10.14 На первом шаге значение X выбирается случайно равномерно на отрезке $[0;1]$. На втором шаге значение Y выбирается случайно и равномерно от 0 до получившегося X.

1. Найдите функции плотности $f(y|x)$, $f(x)$, $f(x, y)$, $f(x|y)$, $f(y)$
2. Найдите $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{E}(X^2)$, $\mathbb{E}(Y^2)$
3. Найдите $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $\mathrm{Cov}(X, Y)$, $\mathrm{Corr}(X, Y)$
4. Найдите $\mathbb{E}(X|Y)$, $\mathbb{E}(Y|X)$, $\mathbb{E}(X^2|Y)$, $\mathbb{E}(Y^2|X)$
5. Найдите $\mathrm{Var}(Y|X)$, $\mathrm{Var}(X|Y)$
6. Найдите $\mathbb{P}(Y > 0.2|X = 0.5)$, $\mathbb{P}(Y > 0.2|X < 0.5)$

10.15 Величины Y_1 и Y_2 независимы и экспоненциально распределены с параметром $\lambda = 5$, а $S = Y_1 + Y_2$ и $R = Y_1/S$.

1. Выпишите и нарисуйте функцию плотности $f(y_1)$
2. Найдите совместную функцию плотности $f(y_1, y_2)$
3. Найдите совместную функцию плотности R и S.
4. Правда ли, что R и S независимы?
5. Найдите плотности $f(y_1, s)$, $f(y_1|s)$.
6. Правда ли, что Y_1 и S независимы?
7. Прокомментируйте простыми словами вид функции $f(y_1|s)$

10.16 Величины X и Y имеют совместную функцию плотности

$$f(x, y) = \begin{cases} 4xy, \ 0 < x < 1, \ 0 < y < 1 \\ 0, \ \text{иначе} \end{cases}$$

1. Найдите $\mathbb{P}(X > 0.5)$, $\mathbb{P}(X + Y > 0.5)$, $\mathbb{P}(X = Y + 0.2)$, $\mathbb{P}(X \leqslant Y)$, $\mathbb{P}(Y > 0.5|X > 0.5)$
2. Найдите $\mathbb{E}(X)$, $\mathbb{E}(XY)$, $\mathrm{Cov}(X, Y)$, $\mathrm{Corr}(X, Y)$
3. Найдите $\mathbb{E}(Y|X)$, $\mathbb{E}(Y^2|X)$, $\mathrm{Var}(Y|X)$

4. Зависимы ли величины X и Y?
5. Найдите совместную функцию распределения $F(x,y)$

10.17 Величины X и Y имеют совместную функцию распределения

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, \ x > 0, \ y > 0 \\ 0, \ \text{иначе} \end{cases}$$

1. Найдите $\mathbb{P}(X > 0.5)$, $\mathbb{P}(X = Y + 0.2)$, $\mathbb{P}(Y > 0.5 | X > 0.5)$;
2. Найдите частные функции распределения $F(x)$, $F(y)$
3. Найдите совместную функцию плотности $f(x,y)$;
4. Найдите $\mathbb{P}(X + Y > 0.5)$, $\mathbb{P}(X \leqslant Y)$;
5. Найдите $\mathbb{E}(X)$, $\mathbb{E}(XY)$, $\mathrm{Cov}(X,Y)$, $\mathrm{Corr}(X,Y)$
6. Зависимы ли величины X и Y?

10.18 Точка случайно равномерно выбирается внутри треугольника с вершинами в $(0;0)$, $(3;3)$ и $(1;2)$. Пусть X и Y — абсцисса и ордината этой точки.

1. Найдите совместную функцию плотности пары (X,Y);
2. Найдите $f(x|y)$, $f(y|x)$.

10.19 Приведите пример пары X и Y у которой нет совместной функции плотности, однако X и Y равномерно распределены на отрезке $[0;1]$

10.20 Пусть X равномерно на $[-1;1]$, $Y = X^2$

1. Найдите $\mathrm{Cov}(X,Y)$;
2. Верно ли, что X и Y независимы?

10.21 Кольцо задается системой неравенств: $x^2 + y^2 \geqslant 1$ и $x^2 + y^2 \leqslant 4$. Случайным образом, равномерно на этом кольце, выбирается точка, X и Y — её координаты.

Чему равна корреляция X и Y? Зависимы ли X и Y?

10.22 Совместная функция плотности величин X и Y имеет вид

$$f(x,y) = \begin{cases} 4xy, \ \text{если } x, y \in [0;1]; \\ 0, \ \text{иначе} \end{cases}$$

1. Выпишите дифференциальную форму вероятностей для пары X и Y
2. Найдите дифференциальную форму для $S = X + Y$ и $R = X/(X+Y)$.
3. Найдите совместную функцию плотности для $S = X + Y$ и $R = X/(X+Y)$.

10.23 Величины U_1 и U_2 независимы и равномерны $U[0;1]$. Рассмотрим пару величин $Y_1 = R \cdot \cos\alpha$, $Y_2 = R \cdot \sin\alpha$, где $R = \sqrt{-2\ln U_1}$, а $\alpha = 2\pi U_2$.

1. Выпишите дифференциальную форму для пары U_1, U_2
2. Выпишите дифференциальную форму для пары Y_1, Y_2
3. Найдите совместный закон распределения Y_1 и Y_2;

4. Верно ли, что Y_1 и Y_2 независимы?
5. Как распределены Y_1 и Y_2 по отдельности?

10.24 В Пуассоновском потоке с интенсивностью λ время наступления k-го события распределено согласно гамма-распределению $Gamma(k, \lambda)$ и имеет вероятностную дифференциальную форму

$$\mathbb{P}(Y \in [y; y+dy]) = const \cdot y^{k-1} \cdot \exp(-\lambda y), \text{ при } y \geqslant 0;$$

До прихода Ёжика Медвежонок сидел на крылечке один и насчитал k_1 падающую звезду. А после прихода Ёжика Медвежонок насчитал ещё k_2 падающих звёзд. Пусть Y_1 и Y_2 — время, которое которое наблюдал за звёздами Медвежонок до и после прихода Ёжика.

1. Выпишите совместную дифференциальную форму для Y_1 и Y_2
2. Найдите совместную дифференциальную форму для $S = Y_1 + Y_2$ и $Z = Y_1/S$.
3. С точностью до сомножителя выпишите дифференциальную форму для доли времени, в течение которого Медвежонок наблюдал звёзды один.
4. Выпишите функцию плотности бета-распределения

10.25 Случайным образом на сфере $x^2 + y^2 + z^2 = 1$ равномерно выбирается точка. Её координаты — случайные величины X, Y и Z.

1. Найдите функцию плотности величины X;
2. Найдите совместную функцию плотности пары величин X и Y;

11. Геометрия

Минитеория:

1. Скалярное произведение, $\langle R, S\rangle = \mathbb{E}(XY)$.
2. Длина, $\|X\| = \sqrt{\langle X, X\rangle}$.
3. Косинус угла между, $\cos(X, Y) = \frac{\langle X,Y\rangle}{\|X\|\|Y\|}$.

11.1 Дайте геометрическую интепретацию следующим величинам:

1. $\mathbb{E}(X)$
2. $\mathrm{Var}(X)$
3. σ_X
4. $\mathrm{Corr}(X, Y)$

11.2 Дайте геометрическую интепретацию тождества $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

11.3 Известно, что $\mathbb{E}(X) = 4$, $\mathbb{E}(X^2) = 20$, $\mathbb{E}(X^3) = 10$, $\mathbb{E}(X^4) = 10000$.

1. Найдите проекцию величины X на множество констант.
2. Найдите проекцию величины X на множество случайных величин с нулевым математическим ожиданием.

3. Найдите $\cos(X, X^2)$.
4. Обозначим проекции X и X^2 на множество случайных величин с нулевым математическим ожиданием буквами R и S. Найдите $\cos(R, S)$. Как он связан с корреляцией $\mathrm{Corr}(X, X^2)$?

11.4 Случайные величины Y и X имеют совместное распределение

	$X=-1$	$X=0$	$X=1$
$Y=-1$	0.1	0.2	0.3
$Y=2$	0.1	0.1	0.2

Потомственная ясновидящая Агафья знает значение X, но не знает значение Y. Агафья хочет построить прогноз $\hat{Y}$, зависящий от X так, чтобы в среднем квадрат ошибки прогноза $\mathbb{E}((Y-\hat{Y})^2)$ был минимален.

1. Найдите наилучший прогноз вида $\hat{Y} = a \cdot X$.
2. Найдите наилучший прогноз вида $\hat{Y} = a + b \cdot X$.
3. Найдите наилучший прогноз $\hat{Y}$ произвольного вида.
4. Проинтепретируйте задачи нахождения наилучшего прогноза геометрически.
5. Представим себе, что Маланья, подруга Агафьи, наоборот, знает Y и не знает X. Маланья пытается построить наилучший прогноз $\hat{X} = c + d \cdot Y$. Найдите c и d. Как связаны b, d и корреляция X и Y?

11.5 Исследователь Василий оценивает неизвестную константу θ с помощью случайной величины $\hat{\theta}$. Проинтерпретируйте геометрически тождество

$$\mathbb{E}((\hat{\theta}-\theta)^2) = \mathrm{Var}(\hat{\theta}) + (\mathbb{E}(\hat{\theta}) - \theta)^2.$$

12. Всё нормально!

12.1 Величины X_1, ..., X_n распределены нормально $\mathcal{N}(4, 100)$ и независимы.

1. Найдите вероятности $\mathbb{P}(X_1 > 4)$, $\mathbb{P}(X_1 \in [2; 20])$, $\mathbb{P}(X_1 < -5)$
2. Найдите вероятности $\mathbb{P}(\bar{X}_{100} > 10)$, $\mathbb{P}(\bar{X}_{36} \in [0; 5])$
3. Найдите такое число a, что $\mathbb{P}(X_1 > a) = 0.3$
4. Найдите такое число b, что $\mathbb{P}(\bar{X}_{100} \in [4-b; 4+b]) = 0.5$

Решите эту задачу двумя способами: с использованием R и с помощью таблиц. В R могут оказаться полезны функции pnorm и qnorm.

12.2 Величина W имеет функцию плотности $f(w) = c \cdot \exp(5w - 2w^2)$. Найдите $\mathbb{E}(W)$, $\mathrm{Var}(W)$, c.

12.3 Величина X нормальна $\mathcal{N}(\mu, \sigma^2)$.

1. Выпишите функцию плотности случайной величины X, $f(x)$
2. Найдите точку максимума и точки перегиба функции f

12.4 Величина X распределена нормально $\mathcal{N}(0, \sigma^2)$.

1. Найдите $\mathbb{E}(|X|)$

2. Найдите функцию плотности $|X|$

12.5 Известно, что $\ln Y \sim \mathcal{N}(\mu; \sigma^2)$. Найдите $\mathbb{E}(Y)$, $\mathrm{Var}(Y)$, медиану и моду величины Y.

12.6 Величина X имеет стандартное нормальное распределение.

1. Найдите $\mathbb{E}(Y)$ для $Y = \max\{X, 0\}$
2. Найдите $\mathbb{E}(X|X < 0)$ и $\mathrm{Var}(X|X < 0)$

12.7 Величина X имеет нормальное распределение $\mathcal{N}(\mu; \sigma^2)$, а функция $g(t)$ дифференцируема и не слишком быстро растёт при $t \to \infty$.

1. Докажите, что $\mathbb{E}((X - \mu)g(X)) = \sigma^2 \mathbb{E}(g'(X))$.
2. Найдите $\mathbb{E}(X^4)$ и $\mathbb{E}(X^6)$ для $X \sim \mathcal{N}(0; 1)$.
3. Какое формальное условие на g достаточно записать вместо нестрогого «не слишком быстро растёт»?

12.8 Вася хочет посчитать сумму 100 случайных слагаемых. Слагаемые независимы и равномерны распределены с математическим ожиданием равным единице. Васе достаточно оценить сумму с точностью до 0.1% с вероятностью больше 99%.

Сколько знаков после запятой достаточно записывать Васе у каждого слагаемого?

13. Долой неравенство Чебышёва и Маркова

13.1 С помощью неравенства Чебышева оцените вероятности

1. $\mathbb{P}(-2\sigma < X - \mu < 2\sigma)$, если $\mathbb{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$
2. $\mathbb{P}(8 < Y < 12)$, если $\mathbb{E}(Y) = 10$, $\mathrm{Var}(Y) = 400/12$
3. $\mathbb{P}(-2 < Z - \mathbb{E}(Z) < 2)$, если $\mathbb{E}(Z) = 1$, $\mathrm{Var}(Z) = 1$
4. Найдите точные значения, если дополнительно известно, что $X \sim \mathcal{N}(\mu; \sigma^2)$, $Y \sim U[0; 20]$ и $Z \sim Exp(1)$.

13.2 Известно, что $\mathbb{E}(X) = 100$, какой должна быть дисперсия величины X, чтобы вне зависимости от закона распределения величины X можно было бы гарантировать, что $\mathbb{P}(X \in [90; 110]) \geqslant 0.95$? А как решить аналогичный вопрос для $\mathbb{P}(X \in [90; 100]) \geqslant 0.95$?

13.3 Известно, что X — неотрицательная случайная величина с $\mathbb{E}(X) = 10$. В каких пределах может лежать вероятность $\mathbb{P}(X < 20)$?

13.4 Сравните:

1. $\mathbb{E}(X^2)$ и $(\mathbb{E}(X))^2$;
2. $\ln(\mathbb{E}(X))$ и $\mathbb{E}(\ln X)$;
3. $\mathbb{E}(1/X)$ и $1/\mathbb{E}(X)$;

13.5 Дискретная случайная величина X обладает интересным свойством $\mathbb{E}(1/X) = 1/\mathbb{E}(X)$. В каких пределах может лежать $\mathrm{Var}(X)$?

13.6 Весна, половодье. На пеньках ждут спасения зайцы. Их очень много. Количество зайев на каждом пеньке — случайная величина равновероятно принимающая значения от 1 до 19. У Деда Мазая две стратегии. Стратегия А: перевозить зайецев на безопасное место равными партиями по 10 штук. Стратегия Б: спасать за одну перевозку зайцев с одного пенька. При какой стратегии у Деда Мазая в среднем будет меньше перевозок?

13.7 Пёс Шарик и Кот Матроскин каждый день в течение месяца покупают молоко в розлив. Цена молока в i-ый день — константа m_i. Средняя цена молока за прошедший месяц оказалась равной 40 рублям. Пёс Шарик каждый день покупал литр молока. Кот Матроскин каждый день покупал молока на 40 рублей. Кто больше потратил денег? Кто больше молока купил?

13.8 Кот Матроскин забрасывает удочку 10 раз. Вероятность поймать рыбку при одном забрасывании равна p. Пёс Шарик забрасывает удочку случайное пуассоновское количество раз, N, под настроение. Известно, что $\mathbb{E}(N) = 10$. У кого шансы поймать хотя бы одну рыбку выше?

14. Полный беспредел

Предел по вероятности, $\operatorname{plim}_{n\to\infty} X_n = X$:

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0, \ \forall \varepsilon > 0$$

Закон Больших Чисел (формулировка Хинчина): Если X_i независимы, одинаково распределены и математическое ожидание $\mathbb{E}(X_i)$ существует, то $\operatorname{plim} \bar{X}_n = \mathbb{E}(X_1)$

Центральная Предельная Теорема (формулировка Линдеберга-Леви): Если X_i независимы, одинаково распределены с математическим ожиданием μ и дисперсией σ^2, то:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \to \mathcal{N}(0;1) \text{ (строго)}$$

$$\bar{X}_n \approx \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right) \text{ (практично)}$$

Дельта-метод: Если Z_n асимптотически нормальны, то есть

$$\frac{Z_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \to \mathcal{N}(0;1), \text{ (строго)}$$

$$Z_n \approx \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right) \text{ (практично)}$$

и f — дифференциируемая функция, такая что $f'(\mu) \neq 0$,
то:

$$\frac{f(Z_n) - f(\mu)}{\sqrt{\frac{(f'(\mu))^2\sigma^2}{n}}} \to \mathcal{N}(0;1) \text{ (строго)}$$

$$f(Z_n) \approx \mathcal{N}\left(f(\mu); \frac{(f'(\mu))^2\sigma^2}{n}\right) \text{ (практично)}$$

14.1 Величины X_1, ..., X_n независимы и равномерно распределены на отрезке $[0;1]$. Найдите $\operatorname{plim} \bar{X}_n$, $\operatorname{plim} 1/(1+\bar{X}_n)$, $\operatorname{plim} \sum_{i=1}^n \ln X_i/n$, $\operatorname{plim} \sqrt[n]{X_1 \cdot \ldots \cdot X_n}$, $\operatorname{plim}(X_1 \cdot \ldots \cdot X_n)$, $\operatorname{plim} \max\{X_1, \ldots, X_n\}$, $\operatorname{plim} \min\{X_1, \ldots, X_n\}$, $\operatorname{plim} \frac{\sum (X_i - \bar{X})^2}{n}$, $\operatorname{plim} X_1/\bar{X}$.

14.2 Величины X_1, ..., X_n независимы и равномерно распределены на отрезке $[0;1]$. Найдите примерный закон распределения величин $\bar{X}_{100}$, $S_{100} = X_1 + \ldots + X_{100}$. Найдите примерно $\mathbb{P}(\bar{X}_{100} > 0.55)$, $\mathbb{P}(S_n \in [50;60])$, $\mathbb{E}(\bar{X}_{100}|\bar{X}_{100} > 0.6)$. Найдите такое число a, для которого $\mathbb{P}(S_n < a) = 0.65$.

14.3 Количество смс за сутки, посылаемое каждым из 160 абонентов, имеет пуассоновское распределение со средним значением 5 смс в сутки. Какова вероятность того, что за двое суток абоненты пошлют в сумме более 1700 сообщений?

14.4 Каждый день цена акции равновероятно поднимается или опускается на один рубль. Сейчас акция стоит 1000 рублей. Найдите вероятность того, что через сто дней акция будет стоить больше 1030 рублей.

14.5 Вероятность выпадения монетки «орлом» равна 0.63.

1. Какова вероятность, что в 100 испытаниях выборочная доля выпадения орлов будет отличаться от истинной вероятности менее, чем на 0.07?
2. Каким должно быть минимальное количество испытаний, чтобы вероятность отличия выборочной доли и истинной вероятности менее чем на 0.02 была больше 0.95?

14.6 Величины X_1, ..., X_n независимы и одинаково распределены с математическим ожиданием 10 и дисперсией 20. Найдите примерный закон распределения величин $\bar{X}^2$, $(1+\bar{X})/(\bar{X}^2+5)$ при большом n.

14.7 Величина X имеет биномиальное распределение $Bin(n,p)$ и n велико. Какое распределение примерно имеют величины $\ln(X/n)$? $X/(n-X)$?

14.8 (*) Случайные величины X и Y независимы и равновероятно принимают значения от 1 до n. Вероятность того, что сумма $X+Y$ является квадратом натурального числа, обозначим p_n. Найдите предел $\lim_{n\to\infty}\sqrt{n}p_n$.

14.9 (*) Величины X_1, ..., X_n независимы и равномерны на отрезке $[0;1]$. Вероятность того, что сумма любых двух соседних иксов меньше единицы, обозначим p_n. Найдите предел $\lim_{n\to\infty} p_n^{1/n}$.

15. Условное математическое ожидание

15.1 Пусть совместное распределение X и Y задано таблицей:

	$X=-1$	$X=1$
$Y=-1$	1/8	4/8
$Y=2$	2/8	1/8

1. Найдите $\mathbb{E}(Y|X)$, $\text{Var}(Y|X)$
2. Убедитесь, что $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$

15.2 Монетка выпадает орлом с вероятностью p. Эксперимент состоит из двух этапов. На первом этапе монетку подкидывают 100 раз и записывают число орлов, Z. На втором этапе монетку подбрасывают до тех пор пока не выпадет столько орлов, сколько выпало на первом этапе. Обозначим число подбрасываний монетки на втором этапе буквой X.
Найдите $\mathbb{E}(X|Z)$, $\text{Var}(X|Z)$, $\mathbb{E}(X)$, $\text{Var}(X)$

15.3 Автобусы приходят на остановку через случайные промежутки времени (Пуассоновский поток с параметром λ). В первый день Вася приходит на остановку и замеряет время до первого автобуса. Пусть это время X. На следующий день Вася приходит на остановку и считает, сколько автобусов придет в течении времени X. Он получает количество автобусов N.

1. Найдите $\mathbb{E}(N)$ и $\mathrm{Var}(N)$
2. Найдите $\mathbb{P}(N > 0)$

15.4 Известно, что $U \sim U[0;1]$, а величина X имеет распределение Рэлея (Rayleigh density):
$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \text{ если } x \geqslant 0 \\ 0, \text{ иначе} \end{cases}$.

1. Найдите $\mathbb{E}(X|Y)$
2. Найдите $\hat{X} = aY + b$ так, чтобы $E[(X - \hat{X})^2]$ была минимальной.

15.5 Маша собрала n грибов в лесу наугад. В лесу есть рыжики, мухоморы и лисички. Рыжики попадаются с вероятностью $p_R > 0$, лисички — с вероятностью $p_L > 0$, мухоморы — с вероятностью $p_M > 0$, $p_R + p_M + p_L = 1$. Пусть R — количество собранных рыжиков, L — лисичек, а M — мухоморов. Найдите:

1. $\mathbb{E}(R + L|M)$, $\mathbb{E}(M|R + L)$
2. $\mathbb{E}(R|L)$
3. $\mathrm{Var}(R|L)$
4. $\mathbb{E}(R + L|L + M)$
5. $\mathbb{E}(R|R - L)$ (? похоже не решается в явном виде)
6. $\mathrm{Var}(R|L)$
7. $\mathbb{P}(\mathbb{E}(R|L) = 0)$
8. $\mathbb{P}(R = 0|L)$
9. $\mathbb{E}\left(\left(\frac{p_M}{p_R+p_M}\right)^{100-L}\right)$

15.6 Пусть X и Y — независимые пуассоновские случайные величины с параметрами λ_1 и λ_2. Найдите $\mathbb{E}(X|X + Y)$

15.7 Вася, Петя и Коля играют в карточного «дурака» втроём. Вася проигрывает с вероятностью p_1, Петя — с вероятностью p_2, Коля — с вероятностью p_3. Естественно, $p_1 + p_2 + p_3 = 1$. Всего они сыграли n партий. Обозначим количества проигранных ими партий X_1, X_2 и X_3, соответственно. Найдите $\mathbb{E}(X_1|X_1 + X_2)$. Может получится и $\mathrm{Var}(X_1|X_1 + X_2)$?

15.8 Пусть X_1, ..., X_{100} независимы и равномерны на $[0;1]$. Пусть $L = \max\{X_1, X_2, \ldots, X_{80}\}$ а $R = \max\{X_{81}, X_{82}, \ldots, X_{100}\}$ и $M = \max\{X_1, \ldots, X_{100}\}$

Найдите

1. $\mathbb{P}(L > R|L)$ и $\mathbb{P}(L > R|R)$ и $\mathbb{P}(L > R|M)$, $\mathbb{P}(L > R|L, M)$
2. $\mathbb{E}(X_1|L)$, $\mathbb{E}(X_1|\min\{X_1, \ldots, X_{100}\})$
3. $\mathbb{E}(\min\{X_1, \ldots, X_{100}\}|\max\{X_1, \ldots, X_{100}\})$
4. $\mathbb{E}(\min\{X_1, \ldots, X_{100}\}|X_1)$

5. Нарисуйте условную функцию распределения $\mathbb{P}(X_1 \leqslant t|L)$

15.9 Величины X_1, X_2, ...независимы и одинаково распределены $\mathcal{N}(2;9)$. Мы складываем случайное количество N слагаемых. Величина N независима от X_i и распределена по Пуассону с параметром $\lambda = 10$. Обозначим сумму буквой $S = \sum_{i=1}^{N} X_i$. Найдите $\mathbb{E}(S)$, $\text{Var}(S)$ и $Cov(S, N)$

15.10 Неправильный кубик выпадает с вероятностью $0{,}5$ шестеркой вверх. Остальные пять граней выпадают равновероятно. Случайная величина X — остаток от деления номера грани на два, Y — остаток от деления номера грани на три. Найдите

1. Закон распределения $\mathbb{E}(X \mid Y)$, $\mathbb{E}(Y \mid X)$
2. Выразите $\mathbb{E}(Y \mid X)$ через X, а $\mathbb{E}(X \mid Y)$ через Y
3. Найдите $Cov(\mathbb{E}(Y \mid X), \mathbb{E}(X \mid Y))$, $Cov(\mathbb{E}(Y \mid X), X)$, $Cov(Y, X)$

15.11 Цена литра молока, X, распределена равномерно на отрезке $[1;2]$. Количество молока, которое дает корова Мурка, Y, распределено экспоненциально с $\lambda = 1$. Надои не зависят от цены. Величина Z — выручка кота Матроскина от продажи всего объема молока.

Найдите

1. $E(Z|X)$, $\text{Var}(Z|X)$, корреляцию Z и X
2. Закон распределения $E(Z|X)$
3. Функцию плотности величины $\text{Var}(Z|X)$

15.12 Кубик подбрасывают бесконечное количество раз. Величина X — номер подбрасывания, когда впервые выпала единица, а Y — номер подбрасывания, когда впервые выпала шестерка. Найдите $\mathbb{E}(Y|X)$.

15.13 The random variables X_1, X_2, ...are independent uniformly distributed on $[0;1]$. I am summing them until the first X_i greater than 0.5 is added. After this term I stop. Let's denote by S the total sum and by N — the number of terms added. Find $\mathbb{E}(S|N)$, $\text{Var}(S|N)$, $\mathbb{E}(S)$, $\text{Var}(S)$

15.14 Известно, что $X = (Z_1 + Z_2)^2 + Z_3$ и $Y = (Z_1 + Z_2)^3 + Z_3$, величины Z_i независимы, $Z_1 \sim U[0;2]$, $Z_2 \sim N(1,4)$, $Z_3 \sim N(-2,9)$. Найдите

1. $\mathbb{E}(X \mid Z_1)$, $\mathbb{E}(X \mid Z_2)$, $\mathbb{E}(X \mid Z_3)$
2. $\mathbb{E}(Y \mid Z_1)$, $\mathbb{E}(Y \mid Z_2)$, $\mathbb{E}(Y \mid Z_3)$

15.15 Вася случайно выбирает между 0 и 1 число X_1, затем случайно выбирает между 0 и X_1 число X_2, затем X_3 между 0 и X_2, и так до бесконечности.

1. Найдите $\mathbb{E}(X_n)$, $\text{Var}(X_n)$;
2. Найдите функцию плотности распределения X_n;
3. Найдите $\mathbb{E}(X_2|X_1, X_3)$;
4. Найдите $\text{plim}\, X_n$

15.16 Приведите пример:

1. X и Y зависимы, но $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

2. X и Y зависимы, но $\mathbb{E}(Y|X) = \mathbb{E}(Y)$ (as)

3. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, но $\mathbb{E}(Y|X) \neq \mathbb{E}(Y)$ (as)

15.17 Пусть X — равномерна на отрезке $[0; 1]$. В шляпе лежат две свернутые бумажки. На одной бумажке написано X, на другой X^2. Вы тяните одну бумажку наугад. Пусть Z — число, написанное на вытянутой Вами бумажке, а W — число на другой бумажке. Увидев число Вы решаете, оставить себе эту бумажку, или отказаться от этой и забрать оставшуюся. Ваш выигрыш — число на оставшейся у Вас бумажке.

1. Найдите $\mathbb{E}(W|Z)$

2. Максимально подробно (кубическое уравнение там будет суровое, не решайте его) опишите стратегию максимизирующую Ваш выигрыш
3. Как изменится результат, если на одной бумажке написано значение X, а на второй — значение случайной величины имеющей такое же распределение, как и X^2, но независимой от X?

16. Многомерное нормальное

16.1 Рассмотрим две предпосылки:

Rot Распределение вектора не изменяется при любом повороте вектора.

Ind Компоненты вектора независимы.

Каким предпосылкам удовлетворяют данные распределения

1. Вектор $(X, Y)'$ равномерно распределён на множестве $\{x^2 \leqslant 1,\ y^2 \leqslant 1\}$.
2. Вектор $(X, Y)'$ равномерно распределён на множестве $\{x^2 + y^2 \leqslant 1\}$.
3. Вектор $(X, Y)'$ имеет функцию плотности $f(x, y) = \exp(-(x^2 + y^2)/2)/2\pi$.

16.2 Какими свойствами должна обладать матрица V, чтобы быть матрицей поворота? То есть при умножении на V должны сохраняться длины векторов и углы между векторами.

16.3 Теорема Хершела-Максвелла

Рассмотрим замкнутую плоскую фигуру внутри которой случайно летают частицы. Обозначим $Z = (X, Y)'$ — вектор скоростей случайно выбираемой частицы.

Максвелл предположил, что:

1. Распределение вектора не должно меняться при повороте вектора на любой угол.
2. Вертикальная и горизонтальная составляющая скорости должны быть независимы.

1. Какой вектор получится, если вектор Z поверхнуть на 90° по часовой стрелке?
2. Докажите, что X распределена так же, как Y и также как $-Y$.
3. Чему равное $\mathbb{E}(X)$? Верно ли, что $\operatorname{Var}(X) = \operatorname{Var}(Y)$?
4. Докажите, что совместная функция плотности $f(x, y)$ представима в виде

$$f(x, y) = h(x^2 + y^2)$$

5. Докажите, что совместная функция плотности $f(x,y)$ представима в виде
$$f(x,y) = g(x^2)\cdot g(y^2)$$
6. Докажите, что отношение $h'(t)/h(t)$ равно константе.
7. Найдите функцию $h(t)$ с точностью до константы.
8. Пусть U — угол, а $R = \sqrt{x^2+y^2}$. Найдите совместную функцию плотности U и R с точностью до константы.
9. Найдите функцию плотности X, если единицы измерения скорости выбраны так, что $\mathrm{Var}(X) = 1$.

16.4 Пара X и Y имеет двумерное нормальное распределение
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 1 \\ -2 \end{pmatrix}; \begin{pmatrix} 9 & -1 \\ -1 & 4 \end{pmatrix}\right]$$

1. Найдите $\mathbb{E}(X)$, $\mathrm{Var}(X)$, $\mathbb{E}(X+3Y-7)$, $\mathrm{Var}(X+3Y-7)$, $\mathrm{Cov}(X-Y, 2X+3Y)$, $\mathrm{Corr}(X-9, X+3Y)$
2. Найдите $\mathbb{P}(X>5)$, $\mathbb{P}(X+Y>5)$
3. Найдите $\mathbb{E}(X|Y)$, $\mathrm{Var}(X|Y)$, $\mathbb{P}(X>1|Y=1)$
4. Найдите $\mathbb{E}(Y-3|X)$, $\mathrm{Var}(2X+7Y+2|Y)$, $\mathbb{P}(Y+X>1|X=2)$

16.5 Ермолай Лопахин решил приступить к вырубке вишневого сада. Однако выяснилось, что растут в нём не только вишни, но и яблони. Причём, по словам Любови Андреевны Раневской, среднее количество деревьев (а они периодически погибают от холода или жары, либо из семян вырастают новые) в саду распределено в соответствии с нормальным законом (X — число яблонь, Y — число вишен) со следующими параметрами:
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 25 \\ 125 \end{pmatrix}; \begin{pmatrix} 5 & 4 \\ 4 & 10 \end{pmatrix}\right) \tag{1}$$

Найдите вероятность того, что Ермолаю Лопахину придется вырубить более 150 деревьев. Каково ожидаемое число подлежащих вырубке вишен, если известно, что предприимчивый и последовательный Лопахин, не затронув ни одного вишнёвого дерева, начал очистку сада с яблонь и все 35 яблонь уже вырубил?

Автор: Кирилл Фурманов, Ира Чернухина

16.6 В самолете пассажирам предлагают на выбор «мясо» или «курицу». В самолет 250 мест. Каждый пассажир с вероятностью 0.6 выбирает курицу, и с вероятностью 0.4 — мясо. Сколько порций курицы и мяса нужно взять, чтобы с вероятностью 99% каждый пассажир получил предпочитаемое блюдо, а стоимость «мяса» и «курицы» для компании одинаковая?

Как изменится ответ, если компания берет на борт одинаковое количество «мяса» и «курицы»?

16.7 Сэр Фрэнсис Гальтон — учёный XIX-XX веков, один из основоположников как генетики, так и статистики — изучал, среди всего прочего, связь между ростом детей и родителей. Он исследовал данные о росте 928 индивидов. Обозначим X_1 — рост случайного человека, а X_2 — среднее арифметическое роста его отца и матери. По результатам исследования Гальтона:
$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 68.1 \\ 68.3 \end{pmatrix}; \begin{pmatrix} 6.3 & 2.1 \\ 2.1 & 3.2 \end{pmatrix}\right]$$

1. Обратите внимание на то, что дисперсия роста детей выше дисперсии среднего роста родителей. С чем это может быть связано? Учтите, что рост детей измерялся уже по достижении зрелости, так что разброс не должен быть связан с возрастными различиями.
2. Рассчитайте корреляцию между X_1 и X_2
3. Один дюйм примерно равен 2.54 сантиметра. Пусть X_1' и X_2' — это те же X_1 и X_2, только измеренные в сантиметрах. Найдите вектор математических ожиданий и ковариационную матрицу вектора $X' = (X_1', X_2')$.
4. Определите, каков ожидаемый рост и дисперсия роста человека, средний рост родителей которого составляет 72 дюйма?
5. Найдите вероятность того, что рост человека превысит 68 дюймов, если средний рост его родителей равен 72 дюймам. Подсказка: используйте предыдущий пункт и нормальность распределения!

Автор: Кирилл Фурманов, Ира Чернухина

16.8 «Регрессия к среднему»

Каждый день независимо от других муж дарит Машке случайное количество роз. Логарифм количества роз (в тысячах цветов), подаренных мужем в день t, ε_t, имеет нормальное распределение $\mathcal{N}(0, 1)$. Улыбчивость Машки в день t, обозначаемая Y_t, зависит от количества цветов, подаренных в этот и в предыдущей день, $Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$. Подружка Машки не наблюдает подарки Машкиного мужа, ε_t, однако видит улыбчивость Машки, Y_t. Машкина подружка хочет спрогнозировать завтрашнюю улыбчивость Машки, исходя из прошлой информации.

1. Найдите $\mathrm{Corr}(Y_t, Y_{t-1})$
2. Найдите прогноз Машкиной улыбчивости завтра при известной сегодняшней улыбчивости $\mathbb{E}(Y_t|Y_{t-1})$
3. Проинтерпретируйте величину коэффициента при Y_{t-1}, становится ли Машка в среднем улыбчивей или грустней со временем?
4. Исследователь Вениамин собрал данные по 1000 семей и изобразил диаграмму рассеяния в осях (рост мамы в 20 лет, рост дочки в 20 лет). Далее он провел наиболее похожую на эти точки прямую. Наклон этой прямой скорее всего будет около 1, меньше 1, больше 1?
5. Если Вовочка плохо пишет контрольную, то его лишают мороженого. После этого успеваемость Вовочки как правило улучшается. В чём сходство этой ситуации с данной задачей?
6. Найдите прогноз $\mathbb{E}(Y_t|Y_{t-1}, Y_{t-2})$.

16.9 Пара случайных величин X и Y имеет совместное нормальное распределение:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}; \begin{pmatrix} 5 & 6 \\ 6 & 10 \end{pmatrix}\right)$$

1. Найдите корреляцию X и Y
2. Найдите собственные числа и собственные векторые ковариационной матрицы
3. Постройте линии уровня совместной функции плотности

16.10 Случайная величина X нормально распределена, $\mathcal{N}(0;4)$. При фиксированном X случайная величина Y нормально распределена, $\mathcal{N}(2X-1,9)$.

1. Найдите безусловный закон распределения величины Y
2. Найдите $\mathrm{Cov}(X,Y)$
3. Найдите $\mathbb{P}(Y>2)$
4. Найдите $\mathbb{E}(X|Y)$, $\mathbb{E}(Y|X)$, $\mathrm{Var}(X|Y)$, $\mathrm{Var}(Y|X)$

16.11 Маша прячется от Медведей в случайной точке на числовой прямой. Место, где спряталась Маша — случайная величина X, имеющая нормальное распределение, $\mathcal{N}(0;4)$. Каждый из n Медведей, обнюхав числовую прямую, имеет своё мнение о том, где спряталась Маша. Эти мнения — случайные величины Y_i. При фиксированном X случайные величины Y_1, ..., Y_n условно независимы и нормально распределены, $\mathcal{N}(X,9)$.

1. Михаилу Потапычу, Медведю номер 1, кажется, что Машей сильнее всего пахнет в точке Y_1. Где ему следует искать Машу, т.е. чему равно $\mathbb{E}(X|Y_1)$?
2. Теперь n Медведей объединились и зная Y_1, Y_2, ..., Y_n хотят понять, где же разумнее всего искать Машу. Помогите им посчитать $\mathbb{E}(X|Y_1,Y_2,\ldots,Y_n)$:
   а) Найдите безусловный закон распределения вектора Y_1, ..., Y_n
   б) Маленькое техническое задание. Пусть I — единичная матрица, а S — матрица строевого леса, то есть матрица, в которой все элементы равны единицам. Найдите $(aI+bS)^{-1}$. Подсказка: ответ имеет вид $cI+dS$.
   в) Найдите $\mathbb{E}(X|Y_1,\ldots,Y_n)$

16.12 Величины X_1 и X_2 имеют совместное нормальное распределение, причем каждая из них имеет стандартное нормальное распределение $\mathcal{N}(0;1)$, а корреляция между ними равна ρ. Найдите $\mathbb{P}(X_1>0,X_2>0)$.

16.13 Величина X имеет стандартное нормальное распределение $\mathcal{N}(0;1)$. Храбрая исследовательница Мишель подкидывает правильную монетку. Если монетка выпадает орлом, то Мишель домножает величину X на единицу, а иначе — на минус единицу, и получает величину Y.

1. Какое распределение имеет величина Y?
2. Верно ли, что пара X и Y имеет совместное нормальное распределение?
3. Чему равна корреляция X и Y?
4. Приведите пример таких случайных величин Z и W, что каждая из них имеет нормальное распределение, корреляция между ними равна 0.5, однако распределение пары (Z,W) не является совместным нормальным.

17. Случайные вектора

17.1 Пусть $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, $\mathbb{E}(x)=\begin{pmatrix}1\\2\end{pmatrix}$, $\mathrm{Var}(x)=\begin{pmatrix}2&1\\1&2\end{pmatrix}$. Найдите $\mathbb{E}(y)$, $\mathrm{Var}(y)$ и $\mathbb{E}(z)$, если

1. $y=x-\mathbb{E}(x)$;
2. $y=\mathrm{Var}(x)x$;
3. $y=\mathrm{Var}(x)(x-\mathbb{E}(x))$;

4. $y = \text{Var}(x)^{-1}(x - \mathbb{E}(x))$;
5. $y = \text{Var}(x)^{-1/2}(x - \mathbb{E}(x))$;
6. $z = (x - \mathbb{E}(x))' \text{Var}(x)(x - \mathbb{E}(x))$;
7. $z = (x - \mathbb{E}(x))' \text{Var}(x)^{-1}(x - \mathbb{E}(x))$;
8. $z = x' \text{Var}(x) x$;
9. $z = x' \text{Var}(x)^{-1} x$.

17.2 Известно, что случайные величины x_1, x_2 и x_3 имеют следующие характеристики:

1. $\mathbb{E}(x_1) = 5$, $\mathbb{E}(x_2) = 10$, $\mathbb{E}(x_3) = 8$;
2. $\text{Var}(x_1) = 6$, $\text{Var}(x_2) = 14$, $\text{Var}(x_3) = 1$;
3. $\text{Cov}(x_1, x_2) = 3$, $\text{Cov}(x_1, x_3) = 1$, $\text{Cov}(x_2, x_3) = 0$.

Пусть случайные величины y_1, y_2 и y_3, представляют собой линейные комбинации случайных величин x_1, x_2 и x_3:

$$\begin{cases} y_1 = x_1 + 3x_2 - 2x_3 \\ y_2 = 7x_1 - 4x_2 + x_3 \\ y_3 = -2x_1 - x_2 + 4x_3 \end{cases}$$

1. Выпишите математическое ожидание и ковариационную матрицу случайного вектора $x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}'$.
2. Напишите матрицу A, которая позволяет перейти от случайного вектора $x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}'$ к случайному вектору $y = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}'$.
3. С помощью матрицы A найдите математическое ожидание и ковариационную матрицу случайного вектора $y = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}'$.

17.3 Случайные величины w_1 и w_2 независимы с нулевым ожиданием и единичной дисперсией. Из них составлено два вектора, $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ и $z = \begin{pmatrix} -w_2 \\ w_1 \end{pmatrix}$

1. Являются ли векторы w и z перпендикулярными?
2. Найдите $\mathbb{E}(w)$, $\mathbb{E}(z)$.
3. Найдите $\text{Var}(w)$, $\text{Var}(z)$, $\text{Cov}(w, z)$.

17.4 Известна ковариационная матрица вектора $\varepsilon = (\varepsilon_1, \varepsilon_2)$,

$$\text{Var}(\varepsilon) = \begin{pmatrix} 9 & -1 \\ -1 & 9 \end{pmatrix}$$

Найдите четыре различных матрицы A, таких что вектор $v = A\varepsilon$ имеет некоррелированные компоненты с единичной дисперсией, то есть $\text{Var}(A\varepsilon) = I$.

17.5 Известно, что $\text{Corr}(X, Z) = 0.7$, $\text{Corr}(X, Y) = 0.6$. В каких пределах может лежать корреляция $\text{Corr}(Y, Z)$?

17.6 Пусть r_1, r_2 и r_3 — годовые доходности трёх рисковых финансовых инструментов. Пусть α_1, α_2 и α_3 — доли, с которыми данные инструменты входят в портфель инвестора. Считаем, что $\sum_{i=1}^{3} \alpha_i = 1$ и $\alpha_i \geqslant 0$ для всех $i = 1, 2, 3$. Пусть $r = \begin{pmatrix} r_1 & r_2 & r_3 \end{pmatrix}'$, $\mathbb{E}(r) = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}'$, $\text{Var}(r) = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$. Параметры $\{a_i\}$ и $\{c_i\}$ известны.

1. Найдите годовую доходность портфеля U инвестора.
2. Докажите, что дисперсия доходности портфеля равна $\sum_{i=1}^{3}\sum_{j=1}^{3}\alpha_i c_{ij}\alpha_j$.
3. Для случая $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, $\alpha 3 = 0.4$, $\mathbb{E}(r) = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}' = \begin{pmatrix} 0.10 & 0.06 & 0.05 \end{pmatrix}'$,

$$\mathrm{Var}(r) = \begin{pmatrix} 0.04 & 0 & -0.005 \\ 0 & 0.01 & 0 \\ -0.005 & 0 & 0.0025 \end{pmatrix}$$

найдите $\mathbb{E}(U)$ и $\mathrm{Var}(U)$.

17.7 Пусть ξ_1, ξ_2, ξ_3 — случайные величины, такие что $\mathrm{Var}(\xi_1) = 2$, $\mathrm{Var}(\xi_2) = 3$, $\mathrm{Var}(\xi_3) = 4$, $\mathrm{Cov}(\xi_1, \xi_2) = 1$, $\mathrm{Cov}(\xi_1, \xi_3) = -1$, $\mathrm{Cov}(\xi_2, \xi_3) = 0$. Пусть $\xi = \begin{pmatrix} \xi_1 & \xi_2 & \xi_3 \end{pmatrix}'$. Найдите $\mathrm{Var}(\xi)$ и $\mathrm{Var}(\xi_1 + \xi_2 + \xi_3)$.

18. Большая сила о-малых

18.1 Вероятность того, что непросветлённый Ученик достигнет Просветления за малый интервал времени, прямо пропорциональна длине этого интервала, то есть

$$\mathbb{P}(\text{достигнуть Просветления за отрезок времени } [0; 0 + \Delta]) = \lambda\Delta + o(\Delta)$$

Какова точная вероятность того, что Ученик, начавший искать Просветление, так и не достигнет его к моменту времени t?

18.2 Дворняги благородных кровей, Шарик и Тузик, очень любят тусоваться вместе. Изначально блоха Изабелла сидит на Шарике. Вероятность перескока Изабеллы с одной собаки на другую за малый интервал времени прямо пропорциональна длине этого интервала, то есть:

$$\mathbb{P}(\text{перескок за отрезок времени } [t; t + \Delta]) = \lambda\Delta + o(\Delta)$$

Какова точная вероятность того, что блоха Изабелла будет сидеть на Шарике в момент времени t?

18.3 В десятиэтажном доме есть невероятный лифт. Лифт мгновенно способен подняться или опуститься на один этаж. Вероятность отправки лифта вверх или вниз за малый промежуток времени $[t; t + \Delta]$ прямо пропорциональна длине этого промежутка. А именно,

$$\mathbb{P}(\text{лифт отправится вверх за интервал } [t; t + \Delta]) = \lambda_u\Delta + o(\Delta)$$

$$\mathbb{P}(\text{лифт отправится вниз за интервал } [t; t + \Delta]) = \lambda_d\Delta + o(\Delta)$$

Если лифт стоит на первом или последнем этаже и нет возможности двигаться вверх или вниз, то соответствующая вероятность зануляется.

Обозначим N_t — этаж, на котором находится лифт в момент времени t. Система находится в стохастическом равновесии, то есть вероятности $\mathbb{P}(N_t = k)$ постоянны во времени.

1. Найдите вероятность встретить лифт на каждом этаже;
2. При каком соотношении λ_u и λ_d вероятность встретить лифт на втором этаже максимальна?

18.4 На одного покупателя кассирша Марфа Петровна тратит экспоненциальное время с параметром λ_s. Покупатели приходят пуассоновским потоком с параметром λ_{in}.

Длина очереди к Марфе Петровне в момент времени t — случайная величина L_t. Предположим, что система находится в стохастическом равновесии, то есть вероятности $\mathbb{P}(L_t = k)$ не зависят от t.

1. Найдите вероятности $\mathbb{P}(L_t = k)$;
2. Найдите среднюю длину очереди $\mathbb{E}(L_t)$;

18.5 Исследователь Василий выбирает равномерно и независимо друг от друга 10 точек на отрезке $[0;1]$. Затем Василий записывает их координаты в порядке возрастания, $Y_1 \leqslant Y_2 \leqslant \ldots \leqslant Y_{10}$.

Не производя вычислений, *по определению*, выпишите функции плотности:

1. величины Y_1;
2. величины Y_{10};
3. величины Y_4;
4. пары величин Y_2 и Y_7;
5. пары величин Y_3 и Y_5;
6. тройки величин Y_1, Y_4 и Y_9;
7. всех величин Y_1, ..., Y_{10};

18.6 про нормальное распределение

19. Броуновское движение

19.1

19.2 Пусть Z — стандартная нормальная случайная величина. Определим случайный процесс $X_t = \sqrt{t}Z$.

1. Найдите $\mathbb{E}(X_t)$
2. Найдите $\text{Var}(X_t)$
3. Верно ли, что у процесса X_t независимые приращения?
4. Нарисуйте три типичные траектории процесса X_t?
5. Будет ли X_t броуновским движением?

19.3 Пусть W_t — стандартное броуновское движение и $a > 0$. Являются ли следующие процессы броуновским движением?

1. $X_t = -W_t$;
2. $X_t = W_{a+t} - W_a$;
3. $X_t = \frac{1}{a} W_{a^2 t}$
4. $X_t = W_t^3$

19.4

19.5

20. Статистика ноль

20.1 Имеется пять действительных чисел: $x, 9, 5, 4, 7$. При каком значении x медиана будет равна среднему?

20.2 Измерен рост 25 человек. Средний рост оказался равным 160 см. Медиана оказалась равной 155 см. Машин рост в 163 см был ошибочно внесен как 173 см. Как изменятся медиана и среднее после исправления ошибки?

А как могут измениться медиана и среднее, если истинный рост Маши равен 153?

20.3 Возможно ли чисто теоретически, что риск катастрофы в расчете на 1 час пути больше для самолета, чем для автомобиля, а в расчете на 1 километр пути — наоборот?

20.4 Деканат утверждает, что если студента N перевести из группы A в группу B, то средний рейтинг каждой группы возрастет. Возможно ли это?

20.5 Есть три группы по 10 человек, две группы по 20 человек и одна группа по 40 человек. У каждой из групп свой преподаватель.

1. Каков средний размер группы, для которой читает лекции наугад выбранный профессор?
2. Каков средний размер группы, в которой учится наугад выбранный студент?
3. Творческий вопрос. Мы ловим студентов наугад и спрашиваем каждого размер группы, в которой он учится. Можно ли как-то восстановить средний размер группы с точки зрения преподавателя?

20.6 Приведите примеры случайных величин, для которых:

1. $\mathrm{Med}(X+Y) = \mathrm{Med}(X) + \mathrm{Med}(Y)$
2. $\mathrm{Med}(X+Y) \neq \mathrm{Med}(X) + \mathrm{Med}(Y)$
3. $\mathrm{Med}(X^k) = \mathrm{Med}(X)^k$ для всех k
4. $\mathrm{Med}(X^2) \neq \mathrm{Med}(X)^2$

20.7 Исследователь Вениамин измерил рост пяти случайно выбранных человек. Какова вероятность того, что истинная медиана роста лежит между минимумом и максимумом из этих пяти наблюдений? Предположим, что рост имеет непрерывное распределение.

20.8 Во время Второй Мировой войны американские военные собрали статистику попаданий пуль в фюзеляж самолёта. По самолётам, вернувшимся из полёта на базу, была составлена карта повреждений среднестатистического самолёта. С этими данными военные обратились к статистику Абрахаму Вальду с вопросом, в каких местах следует увеличить броню самолёта.

Что посоветовал Абрахам Вальд и почему?

20.9 Два лекарства испытывали на мужчинах и женщинах. Каждый человек принимал только одно лекарство. Общий процент людей, почувствовавших улучшение, больше среди принимавших лекарство A. Процент мужчин, почувствовавших улучшение, больше среди мужчин, принимавших лекарство B. Процент женщин, почувствовавших улучшение, больше среди женщин, принимавших лекарство B.

1. Возможно ли это?

2. Какое лекарству нужно порекомендовать больному, не зная его пола?

20.10 Из набора чисел $\{2, 4, 10, 14\}$ случайным образом равновероятно по очереди выбираются три числа с возможностью повторения.

1. Найдите закон распеределения (табличку с вероятностями) величины X_1. Найдите закон распределения величины X_2.
2. Найдите совместный закон распределения пары X_1, X_2. Найдите совместный закон распределения пары X_1, X_3.
3. Являются величины X_1, X_2, X_3 независимыми? Одинаково распределеннными?
4. Верно ли, что $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3)$? Верно ли, что $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \mathrm{Var}(X_3)$?
5. Верно ли, что $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_1, X_3)$?
6. Как изменятся ответы на предыдущие вопросы, если числа выбираются без возможности повторения?

20.11 Из фиксированного множества N чисел случайным образом выбирают n чисел. Известно, что если бы выбирать наугад всего одно число, то тогда математическое ожидание и дисперсие этого одного случайного числа были бы равны $\mathbb{E}(X_1) = \mu$ и $\mathrm{Var}(X_1) = \sigma^2$. Обозначим среднее арифметическое выбранных n чисел с помощью $\bar{X}_n$. Чему равны $\mathbb{E}(\bar{X}_n)$ и $\mathrm{Var}(\bar{X}_n)$, если:

1. Мы выбираем n чисел из N с возвращениями
2. Мы выбираем n чисел из N без возвращений
3. Во что превращаются полученные формулы при $n = 1$? при $n = N$? при $N \to \infty$?

20.12 Исследовательница Мишель подбрасывает кубик 100 раз. Пусть X_1 — количество выпадений единицы, а X_6 — количество выпадений шестёрки.

1. Как распределена величина X_1? Величина X_6? Найдите $\mathbb{E}(X_1)$, $\mathrm{Var}(X_1)$.
2. Верно ли, что величины X_1 и X_6 независимы? Одинаково распереданы?
3. Найдите $\mathrm{Cov}(X_1, X_1 + X_2 + X_3 + X_4 + X_5 + X_6)$, $\mathrm{Cov}(X_1, X_6)$
4. Найдите $\mathrm{Corr}(X_1, X_6)$, проинтерпретируйте эту величину

20.13 В множестве A всего два числ, $A = \{24, 42\}$. Случайным образом из множества A выбираются 3 числа с возможностью повторений. Явно найдите закон распределения выборочного среднего, выборочной медианы, выборочной моды, выборочного минимима и выборочного максимума.

20.14 Величины X и Y независимы и одинаково распределены на отрезке $[0; 1]$ с функцией плотности $f(x) = 2x$.

1. Найдите теоретическую медиану $\mathrm{Med}(X)$.
2. Найдите теоретическую медиану $\mathrm{Med}(X + Y)$.

21. Случайная выборка

21.1 Создайте случайную выборку объемом $n = 1000$ из равномерного на отрезке $[0; 1]$ распределения.

1. Найдите выборочные характеристики: среднее, медиану, минимум и максимум, стандартную ошибку, 10%-ый и 95%-ый квантили.
2. Постройте гистограмму распределения, выборочную функцию распределения для первых 20 чисел из случайной выборки
3. Повторите данный опыт для нормального $\mathcal{N}(5, 1)$ распределения и для экспоненциального распределения с параметром $\lambda = 1$
4. Насколько сильно выборочные характеристики отличаются от истинных?

21.2 Придумайте способ, как сгенерировать 100 одинаково распределенных случайных величин, таких что $\sum_{i=1}^{100} X_i = 50$. Будут ли эти величины X_i зависимы?Модифицируйте способ, так чтобы он давал одинаково распределенные величины, такие что $\sum_{i=1}^{100} Y_i^2 = 50$. Будут ли эти новые величины Y_i зависимы?

21.3 Придумайте детерминистическую функцию, такую, которая бы превращала одну равномерную на $[0; 1]$ случайную величину X в

1. случайную величину Y, принимающую значения 1 и 0 с вероятностями 0.7 и 0.3 соответственно
2. случайную величину Z с функцией плотности $f(z) = 2z$ на отрезке $z \in [0; 1]$
3. пару независимых одинаково распределенных случайных величин (Y_1, Y_2), принимающих значения 1 и 0 с вероятностями 0.7 и 0.3 соответственно
4. пару независимых равномерных на $[0; 1]$ случайных величин

21.4 Постройте случайную выборку в $n = 200$ наблюдений из двумерного нормального распределения с параметрами:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 25 \\ 125 \end{pmatrix}; \begin{pmatrix} 5 & 4 \\ 4 & 10 \end{pmatrix} \right) \tag{2}$$

1. Посчитайте выборочную ковариацию, выборочную корреляцию
2. Постройте диаграмму рассеяния, нанесите на диаграмму рассеяния линию $y(x) = E(Y|X = x)$
3. Насколько сильно выборочные характеристики отличаются от истинных?

21.5 Создайте 500 выборок объемом $n = 20$ каждая из равномерного на отрезке $[0; 1]$ распределения и вычислите выборочное среднее для каждой из выборок.

1. Каково теоретическое математическое ожидание и дисперсия каждого из выборочных средних?
2. Постройте гистограмму выборочных средних
3. На фоне функции плотности стандартного нормального распределения изобразите в подходящем масштабе гистограмму стандартизированных выборочных средних

22. Проецируй!

Если:

1. вектор Z имеет многомерное нормальное стандартное распределение, $Z \sim \mathcal{N}(0; I)$;
2. $\hat{Z}$ — это проекция вектора Z на некоторое d-мерное подпространство V;
3. Q — это квадрат длины проекции, $Q = ||\hat{Z}||^2$;

то закон распределения величины Q называется хи-квадрат распределением с d степенями свободы и обозначается $Q \sim \chi^2_d$.

22.1 Вектор Z имеет многомерное нормальное распределение, $Z_i \sim \mathcal{N}(0; 1)$, и все Z_i независимы. Для каждого случая найдите проекцию $\hat{Z}$ вектора Z на подпространство V; найдите квадрат длины проекции, Q; укажите закон распределения величины Q:

1. $V = \mathcal{L}in(e)$, где $e = (1, 1, 1, 1, \ldots, 1)$;
2. $V = \mathcal{L}in(e)$, где $e = (1, 2, 3, 4, \ldots, n)$;
3. $V = \mathcal{L}in(e_1, e_2)$, где $e_1 = (1, 0, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 1, 1, \ldots, 1)$;
4. $V = \mathcal{L}in^{\perp}(e)$, где $e = (1, 1, 1, 1, \ldots, 1)$;
5. $V = \mathcal{L}in(e_5, e_7, e_9)$, вектор e_i содержит 1 на i-ом месте и 0 на остальных;

22.2 Найдите минимум функции $f(a, b, c) = (6-a)^2 + (3-b)^2 + (7-b)^2 + (8-c)^2 + (9-c)^2 + (10-c)^2 + 11^2$;

Проекцией какого вектора на какое пространство является вектор $\hat{Z} = (a^*, b^*, b^*, c^*, c^*, c^*, 0)$?

22.3 Вектор Z из n случайных величин имеет многомерное нормальное распределение, $Z_i \sim \mathcal{N}(0; 1)$, и все Z_i независимы. Определите, какое распределение имеет величина Q, на какое подпространство проецировали вектор Z, и найдите вероятность:

1. $Q = Z_1^2$, $\mathbb{P}(Q > 6.6)$;
2. $Q = n\bar{Z}^2$, $\mathbb{P}(Q < 3.8)$;
3. $Q = \sum (Z_i - \bar{Z})^2$;
4. $Q = Z_5^2 + Z_6^2 + Z_{32}^2$, $\mathbb{P}(Q < 0.58)$;
5. $Q = Z_2^2 + (Z_7 + Z_{11})^2/2$, $\mathbb{P}(Q > 6)$;
6. $Q = (Z_7 + Z_{11})^2/2 + (Z_3 + Z_9 + Z_{12})^2/3$, $\mathbb{P}(Q < 0.21)$;

Для каждого случая укажите подпространство, для которого величина Q будет квадратом длины проекции исходного вектора Z;

22.4 Вектор Z из n случайных величин имеет многомерное нормальное распределение, $Z_i \sim \mathcal{N}(0; 1)$, и все Z_i независимы.

1. Какое распределение имеет величина $Q = Z_1^2 + Z_2^2 + \ldots + Z_d^2$?
2. Чему равно $\mathbb{E}(Q)$?
3. Чему равна дисперсия $\mathrm{Var}(Q)$?
4. Величина Q_a имеет хи-квадрат распределение с a степенями свободы, а величина Q_b — с b степенями свободы. Величины Q_a и Q_b независимы. Какое распределение имеет величина $S = Q_a + Q_b$?

22.5 Найдите функцию плотности χ-квадрат распределения с одной степенью свободы;

22.6 Вектор Z из n случайных величин имеет многомерное нормальное распределение, $Z_i \sim \mathcal{N}(0;1)$, и все Z_i независимы. Вектор v имеет единичную длину.

1. Найдите вектор $\hat{Z}$, проекцию вектора Z на подпространство $\mathcal{L}in(v)$; Чему равно $\hat{Z}_i$?
2. Найдите дисперсию $\text{Var}(\langle Z, v\rangle)$;
3. Найдите ковариацию $\text{Cov}(\hat{Z}_i, \hat{Z}_j)$;
4. Как выглядит ковариационная матрица вектора $\hat{Z}$?

22.7 Пусть величины X и Y независимы и имеют хи-квадрат распределение с одной и двумя степенями свободы. Введём величины $R = X/(X+Y)$ и $S = X+Y$.

1. Выпишите совместную функцию плотности $f(x,y)$;
2. Найдите совместную функцию плотности $f(r,s)$;
3. Верно ли, что R и S независимы?
4. С точностью до сомножителя найдите функцию плотности S;
5. Какой закон распределения имеет величина S?
6. Предположите вид функции плотности хи-квадрат распределения с d степенями свободы и докажите догадку по индукции;

23. Максимально правдоподобно — 1!

23.1 Кот Матроскин каждый вечер ходит на рыбалку. Поймав одну «рыбку» кот Матроскин возвращается домой. В пруду встречаются караси, щуки и бегемоты. Кот Матроскин хочет оценить вероятность p поймать карася. От своей бабушки Кот Матроскин достоверно знает, что щуки встречаются в два раза чаще карасей. За ночь экосистема пруда успевает восстановиться от воздействия кота Матроскина.

1. Оцените $\hat{p}_{KM}$ методом максимального правдоподобия, если Кот Матроскин ловил «рыбку» четыре дня и имеются наблюдения: $X_1 =$ щука, $X_2 =$ карась, $X_3 =$ карась, $X_4 =$ бегемот.
2. Постройте оценку $\hat{p}_{KM}$ методом максимального правдоподобия в общем виде. То есть, Кот Матроскин ходил на пруд n дней, поймал $Y_{\text{к}}$ карасей, $Y_{\text{щ}}$ щук и $Y_{\text{б}}$ бегемотов.
3. Зависимы ли величины $Y_{\text{к}}$, $Y_{\text{щ}}$ и $Y_{\text{б}}$? Как распределена величина $Y_{\text{б}}$? Найдите $\mathbb{E}(Y_{\text{б}})$, $\text{Var}(Y_{\text{б}})$.
4. Найдите $\mathbb{E}(\hat{p}_{\text{КМ}})$, $\text{Var}(\hat{p}_{\text{КМ}})$
5. Является ли оценка $\hat{p}_{\text{КМ}}$ несмещенной, состоятельной?
6. Постройте аналогичную оценку $\hat{p}_{\text{ПШ}}$ для Пса Шарика. В отличие от Кота Матроскина Пёс Шарик не знает, что щуки встречаются в два раза чаще карасей. Является ли оценка Пса Шарика несмещенной и состоятельной?
7. Какая из двух оценок является более эффективной? Почему?

23.2 «Про зайцев». В темно-синем лесу, где трепещут осины, живут $n >> 0$ зайцев. Мы случайным образом отловили 100 зайцев. Каждому из них на левое ухо мы завязали бант из красной ленточки и потом всех отпустили. Через неделю будет снова отловлено 100 зайцев. Из них случайное количество S зайцев окажутся с бантами.

1. Постройте ML и MM оценку для неизвестного параметра n, если оказалось, что $s = 80$
2. Постройте ML и MM оценку для неизвестного параметра n в общем случае

23.3 Вася и Петя независимо друг от друга прочитали всю Википедию. Вася всего нашёл 100 опечаток, Петя — 200 опечаток. При этом 80 опечаток оказались найдены и Петей, и Васей. С помощью ML и MM:

1. Оцените количество опечаток в Википедии
2. Оцените внимательность Васи, то есть вероятность, с которой Вася находит опечатки

23.4 У Васи есть два одинаковых золотых слитка неизвестной массы m каждый и весы, которые взвешивают с некоторой погрешностью. Сначала Вася положил на весы один слиток и получил результат $Y_1 = m + u_1$, где u_1 — случайная величина, ошибка первого взвешивания. Затем Вася положил на весы сразу оба слитка и получил результат $Y_2 = 2m + u_2$, где u_2 — случайная величина, ошибка второго взвешивания. Оказалось, что $y_1 = 0.9$, а $y_2 = 2.3$.

Используя ML оцените вес слитка m и параметр погрешности весов b, если

1. u_i — независимы и $\mathcal{N}(0; b)$
2. u_i — независимы и $U[-b; b]$

23.5 Задача о немецких танках [4]

Предположим, что все выпущенные танки имеют порядковый номер. От самого первого выпущенного танка, имеющего номер 1, до самого последнего танка, имеющего номер n. В бою удалось подбить танки с номерами 15, 29 и 23.

1. Постройте оценку количества танков методом моментов
2. Постройте оценку количества танков методом максимального правдоподобия
3. Постройте несмещённую оценку количества танков с наименьшей дисперсией

24. Максимально правдоподобно — 2!

Минитеория
Метод моментов (MM, method of moments): найти θ из уравнения $\bar{X}_n = \mathbb{E}(X_i)|_{\theta=\hat{\theta}}$
Метод максимального правдоподобия (ML, maximum likelihood):
найти θ при котором вероятность получить имеющиеся наблюдения будет максимальной
Наблюдаемая информация Фишера: $\hat{I} = -\frac{\partial^2 l}{\partial^2 \theta}(\hat{\theta})$, $\widehat{\mathrm{Var}}(\hat{\theta}_{ML}) = \hat{I}^{-1}$.

Пусть $l(\theta)$ — логарифмическая функция правдоподобия ($l(\theta) = \ln(f(X_1, \ldots, X_n, \theta))$).
Ожидаемая информация Фишера $I(\theta) = \mathbb{E}\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] = -E\left(\frac{\partial^2 l}{\partial \theta^2}\right)$
Сколько информации о неизвестном θ содержится в выборке X_1, ..., X_n
Неравенство Крамера-Рао (Cramer-Rao) («слишком хорошей оценки не бывает»):
Если $\hat{\theta}$ - несмещенная оценка и ..., то $\mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{I(\theta)}$
Оценки ML — самые лучшие (асимптотически несмещенные и с минимальной дисперсий):
Если X_i — iid, ..., и $n \to \infty$ то $\hat{\theta}_{ML} \sim N(\theta, \frac{1}{I(\theta)})$.

[4] Незадолго до высадки союзников в Нормандии, 6 июня 1944 года, в распоряжении союзников было всего два (!) немецких танка «Пантера V». По серийным номерам на шасси танков союзники оценили выпуск в феврале 1944 в 270 танков. Фактический выпуск «Пантер V» согласно немецким документам в феврале 1944 составил 276 танков, [RB47].

24.1 Допустим, что X_i — независимы и имеют закон распределения, заданный табличкой:

X	-1	0	2
$\mathbb{P}()$	θ	$2\theta-0.2$	$1.2-3\theta$

Имеется выборка: $X_1=0$, $X_2=2$.

1. Найдите оценки $\hat{\theta}_{ML}$ и $\hat{\theta}_{MM}$

2. Первоначально ничего о θ не было известно и поэтому предполагалось, что θ распределена равномерно на $[0.1;0.4]$. Как выглядит условное распределение θ, если известно что $X_1=0$, $X_2=2$?

3. Постройте ML и MM оценки для произвольной выборки X_1, X_2, ...X_n

24.2 Пусть Y_1 и Y_2 независимы и распределены по Пуассону. Известно также, что $\mathbb{E}(Y_1)=e^a$ и $\mathbb{E}(Y_2)=e^{a+b}$.

1. Найдите ML оценки $\hat{a}$ и $\hat{b}$ для случая $y_1=7$ и $y_2=3$
2. Найдите ML оценки $\hat{a}$ и $\hat{b}$ в общем виде

24.3 Пусть X_1, ..., X_n распределены одинаково и независимо. Оцените значение θ с помощью ML (везде) и MM (в «а» и «б»), оцените дисперсию ML оценки, если функция плотности X_i, $p(t)$ имеет вид:

1. $\theta t^{\theta-1}$ при $t\in[0;1]$;
2. $\frac{2t}{\theta^2}$ при $t\in[0;\theta]$

3. $\frac{\theta e^{-\frac{\theta^2}{2t}}}{\sqrt{2\pi t^3}}$ при $t\in[0;+\infty)$;
4. $\frac{\theta\left(\ln^{\theta-1}t\right)}{t}$ при $t\in[1;e]$;
5. $\frac{e^{-|t|}}{2\left(1-e^{-\theta}\right)}$ при $t\in[-\theta;\theta]$

24.4 Пусть X_1, ..., X_n - независимы и экспоненциальны с параметром λ. Постройте MM и ML оценки параметра λ. Оцените дисперсию ML оценки.

24.5 Пусть X_1, ..., X_n - независимы и $\mathcal{N}(\mu;\sigma^2)$. Значение σ^2 известно. Постройте MM и ML оценки параметра μ.

24.6 Пусть X_i независимы и одинаково распределены $\mathcal{N}(\alpha,2\alpha)$
По выборке X_1, ..., X_n постройте оценку для α с помощью ML и MM. Оцените дисперсию ML оценки.

24.7 Пусть $Y_1\sim N(0;\frac{1}{1-\theta^2})$. Найдите ML оценку для θ. Оцените дисперсию ML оценки.

24.8 Пусть X_1, X_2,..., X_n независимы и их функции плотности имеет вид:
$f(x)=\begin{cases}(k+1)x^k, & x\in[0;1];\\ 0, & x\notin[0;1].\end{cases}$
Найдите оценки параметра k с помощью ML и MM. Оцените дисперсию ML оценки.

24.9 Пусть X_1, X_2,..., X_n независимы и равномерно распределены на отрезке $[0;\theta]$, $\theta > 1$

1. Постройте ММ и ML оценки для неизвестного θ.
2. Как изменятся ответы на «а», если исследователь не знает значений самих X_i, а знает только количество X_i оказавшихся больше единицы?

24.10 В озере водятся караси, окуни, щуки и налимы. Вероятности их поймать занесены в табличку

Рыба:	Карась	Окунь	Щука	Налим
$\mathbb{P}()$	0.1	p	p	$0.9-2p$

Рыбак поймал 100 рыб и среди пойманных 100 рыб он посчитал количества карасей, окуней, щук и налимов.

1. Постройте $\hat{p}_{ML}$
2. Найдите ожидаемую и наблюдаемую информацию Фишера
3. Несмещенная оценка $\hat{\theta}$ получена по 100 наблюдениям: X_1, ..., X_{100}. В каких пределах может лежать $\mathrm{Var}(\hat{\theta})$?

24.11 Известно, что X_i — независимы и имеют закон распределения, заданный таблицей:

X_i	0	1
$\mathbb{P}()$	p	$1-p$

1. Постройте $\hat{p}_{ML}$
2. Найдите ожидаемую и наблюдаемую информацию Фишера. Постройте возможные графики $I(p)$.
3. Пусть $\hat{\theta}$ — несмещенная оценка, полученная по 100 наблюдениям: X_1, ..., X_{100}. В каких пределах может лежать $\mathrm{Var}(\hat{\theta})$?

24.12 Пусть X_i независимы и имеют экспоненциальное распределение с параметром λ, т.е. $p(t) = \lambda e^{-\lambda t}$.

1. Найдите $I(\lambda)$, если наблюдаются X_1, ..., X_n
2. Пусть $\lambda = 1/\theta$, т.е. $p(t) = \frac{1}{\theta}e^{-\frac{1}{\theta}t}$. Найдите $I(\theta)$, если наблюдается X_1, ..., X_n

24.13 Пусть X_i — независимы и одинаково распределены. Пусть $I_{X_i}(\theta)$ - информация Фишера о θ, получаемая при наблюдении X_i.

1. Верно ли, что $I_{X_1}(\theta) = I_{X_2}(\theta)$?
2. Как найти $I_{X_1,\ldots,X_n}(\theta)$ зная $I_{X_i}(\theta)$?

24.14 Величины X_1, ..., X_n — независимы и одинаково распределены с функцией плотности $f(t) = \frac{\theta(\ln t)^{\theta-1}}{t}$ при $t \in [1; e]$. По выборке из 100 наблюдений оказалось, что $\sum \ln(\ln(X_i)) = -30$

1. Найдите ML оценку параметра θ и ожидаемую и наблюдаемую информацию Фишера
2. Постройте 95% доверительный интервал для θ

24.15 Величины X_1, ..., X_n — независимы и одинаково распределены с функцией плотности $\frac{\theta e^{-\frac{\theta^2}{2t}}}{\sqrt{2\pi t^3}}$ при $t \in [0; +\infty)$. По выборке из 100 наблюдений оказалось, что $\sum 1/X_i = 12$

1. Найдите ML оценку параметра θ и информацию Фишера $I(\theta)$
2. Пользуясь данными по выборке постройте оценку $\hat{I}$ для информации Фишера
3. Постройте 90% доверительный интервал для θ Hint: $\mathbb{E}(1/X_i) = 1/\theta^2$ (интеграл берется заменой $x = \theta^2 a^{-2}$)

24.16 Известно, что X_i независимы, $\mathbb{E}(X_i) = 5$, $\text{Var}(X_i) = 4$ и n велико. Как примерно распределены следующие величины:

1. $\bar{X}_n$,
2. $Y_n = (\bar{X}_n + 3)/(\bar{X}_n + 6)$,
3. $Z_n = \bar{X}_n^2$,
4. $W_n = 1/\bar{X}_n$

24.17 Известно, что X_i независимы и равномерны на $[0; 1]$.

1. Найдите $\mathbb{E}(\ln(X_i))$, $\text{Var}(\ln(X_i))$, $\mathbb{E}(X_i^2)$, $\text{Var}(X_i^2)$
2. Как примерно распределены величины $X_n = \frac{\sum \ln(X_i)}{n}$, $Y_n = (X_1 \cdot X_2 \cdots X_n)^{1/n}$, $Z_n = \left(\frac{\sum X_i^2}{n}\right)^3$ при больших n?

24.18 Величины X_i независимы и имеют функцию плотности $f(x) = a \cdot x^{a-1}$ на отрезке $[0; 1]$.

1. Постройте оценку $\hat{a}$ методом моментов, укажите её примерный закон распределения
2. По 100 наблюдениям оказалось, что $\sum X_i = 25$. Посчитайте численное значение $\hat{a}$ и оцените дисперсию случайной величины $\hat{a}$.

24.19 Начинающий каратист Вася тренируется бить кирпичи ударом ладони. Каждый день он бьёт ладонью по кирпичу до пор, пока тот не расколется от одного удара. Предположим, что вероятность разбить кирпич с одного удара равна p и неизменна во времени. Величины X_1, X_2, ..., X_n — количества ударов которые потребовались Васе в соответствующий день.

1. Найдите оценку p методом максимального правдоподобия
2. Найдите достаточную статистику T
3. Выразите $\widehat{\text{Var}}(\hat{p})$ через достаточную статистику T
4. Найдите $\mathbb{P}(X_1 = 1 \mid T = t)$.

24.20 Продавщица Глафира отдаёт псу Шарику в конце каждого дня нерасфасованные остатки мясного фарша. Фарш фасуется упаковками по a грамм, поэтому нерасфасованный остаток в i-ый день, X_i, случаен и равномерно распределен на отрезке $[0; a]$. Пёс Шарик хорошо помнит все X_1, ..., X_n. Помогите псу Шарику:

1. Найдите оценку a методом максимального правдоподобия
2. Найдите достаточную статистику T
3. Выразите $\widehat{\text{Var}}(\hat{a})$ через достаточную статистику T
4. Найдите $\mathbb{P}(X_1 < 10 \mid T = t)$.

24.21 Величины X_i равномерны на отрезке $[-a; 3a]$ и независимы. Есть несколько наблюдений, $X_1 = 0.5$, $X_2 = 0.7$, $X_3 = -0.1$.

1. Найдите $\mathbb{E}(X_i)$ и $\mathbb{E}(|X_i|)$
2. Постройте оценку метода моментов, используя $\mathbb{E}(X_i)$
3. Постройте оценку метода моментов, используя $\mathbb{E}(|X_i|)$
4. Постройте оценку обобщёного метода моментов используя моменты $\mathbb{E}(X_i)$, $\mathbb{E}(|X_i|)$ и взвешивающую матрицу
$$W = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$
5. Найдите оптимальную теоретическую взвешивающую матрицу для обобщённого метода моментов
6. Постройте двухшаговую оценку обобщённого метода моментов, начав со взвешивающей матрицы W

24.22 Величины X_i имеют Пуассоновское распределение с параметром λ и независимы. Есть несколько наблюдений, $X_1 = 5$, $X_2 = 7$, $X_3 = 1$.

1. Найдите $\mathbb{E}(X_i)$ и $\mathbb{E}((X_i - \bar{X})^2)$
2. Постройте оценку метода моментов, используя $\mathbb{E}(X_i)$
3. Постройте оценку метода моментов, используя $\mathbb{E}((X_i - \bar{X})^2)$
4. Постройте оценку обобщёного метода моментов используя моменты $\mathbb{E}(X_i)$, $\mathbb{E}((X_i - \bar{X})^2)$ и взвешивающую матрицу
$$W = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$
5. Найдите оптимальную теоретическую взвешивающую матрицу для обобщённого метода моментов
6. Постройте двухшаговую оценку обобщённого метода моментов, начав со взвешивающей матрицы W

24.23 Величины X_1, ..., X_n независимы и равномерны на отрезке $[-1; 1]$. Величины Y_1, ..., Y_n независимы и имеют на отрезке $[-1; 1]$ функцию плотности

$$h(y) = \frac{3\exp(\alpha^{-4})}{4}(1 - \exp(2\alpha^{-4})(y + \alpha - 1)^2)$$

Исследовательница Зинаида создаёт величины Z_1, ..., Z_n по простому принципу. Каждая Z_i равна Y_i с вероятностью α и X_i с вероятностью $(1 - \alpha)$.

Пусть $n = 500$.

1. Постройте выборку из X_i и изобразите результат с помощью гистограммы.
2. Постройте выборку из Y_i при $\alpha = 0.5$ и изобразите результат на гистограмме.
3. Постройте выборку из Z_i при $\alpha = 0.5$ и изобразите результат на гистограмме.

У рассеянной Зинаиды сохранилась только выборка из Z_i. Она не помнит ни α, ни X_i, ни Y_i. И поэтому Зинаида решила восстановить α с помощью максимального правдоподобия.

4. Для полученной выборки Z_i постройте график функции правдоподобия. Будьте осторожны! Если по-быстрому строить график командой встроенной в R/python/julia/..., то получится неверный график! Обратите внимание на значения функции в точках Z_i.
5. Найдите оценку максимального правдоподобия для параметра α по полученной выборке.
6. Найдите $\operatorname{plim}_{n\to\infty} \hat{\alpha}_{ML}(n)$ для произвольного α. Является ли оценка $\hat{\alpha}_{ML}$ состоятельной?

24.24 Величины X_1, ..., X_n независимы и одинаково распределены с функцией распределения

$$F(x) = 1/(1 + \exp(a - x))$$

Рассмотрим три оценки: оценку максимального правдоподобия, $\hat{a}_{ML}$, простую медиану $\hat{a}_{Med} = \operatorname{Med} X_1, \ldots, X_n$ и скорректированную медиану $\hat{a}_{1step} = \hat{a}_{Med} - \frac{\ell'(\hat{a}_{Med})}{\ell''(\hat{a}_{Med})}$.

1. Существует ли в явном виде формула для $\hat{a}_{ML}$ в этой задаче?
2. Сгенерируйте 1000 выборок по $n = 20$ наблюдений в каждой для $a = 2$. Для каждой выборки найдите оценки трёх видов. Таким образом получится 1000 оценок каждого вида. Рассчитайте выборочную дисперсию и выборочное среднее каждой оценки.
3. Прокомментируйте полученные результаты

25. Настоящий ценитель

25.1 Случайные величины X_1, X_2, ...независимы и одинаково распределены с неизвестными $\mathbb{E}(X_i) = \mu$ и $\operatorname{Var}(X_i) = \sigma^2$. Исследовательница Борислава хочет использовать оценку вида $\hat{\mu} = c(X_1 + X_2 + \ldots + X_n)$ для неизвестного параметра μ.

1. При каком c оценка Бориславы будет несмещённой? Возможно ли использовать такое c в практической задаче?
2. При каком c будет минимальной величина $MSE = \mathbb{E}((\hat{\mu} - \mu)^2)$? Возможно ли использовать такое c в практической задаче?
3. Святозар минимизирует по $\hat{\mu}$ штрафную функцию

   $$Q(\hat{\mu}) = \sum (y_i - \hat{\mu})^2 + \lambda \hat{\mu}^2.$$

   При каком λ оценка Святозара совпадёт с несмещённой оценкой Бориславы? С оценкой минимизирующей MSE?

25.2 Исследовательница Радомира размышляет о том, как оценить неизвестное математическое ожидание по выборке из независимых одинаково распределённых случайных величин. Она мучительно выбирает из нескольких оценок. Для каждой оценки определите, является ли она несмещённой, состоятельной и линейной по наблюдениям. Для линейных несмещённых оценок определите, являются ли они эффективными среди линейных несмещённых оценок.

1. Удалить из выборки наблюдение номер 13 и посчитать среднее арифметическое.
2. Удалить из выборки все нечётные наблюдения и посчитать среднее арифметическое.
3. Удалить из выборки все наблюдения после 13-го и посчитать среднее арифметическое.

4. Домножить наблюдение номер 13 на 13 и посчитать среднее арифметическое.
5. Прибавить число 13 к наблюдению номер 13 и посчитать среднее арифметическое.
6. Продублировать 13-ое наблюдение 13 раз и посчитать среднее арифметическое.
7. Продублировать каждое наблюдение 13 раз и посчитать среднее арифметическое.
8. Домножить первое наблюдения на 1, второе — на 2, третье — на 3, и так далее и посчитать среднее арифметическое.
9. Прибавить к первому наблюдению 1, ко второму — 2, к третьему — 3, и так далее и посчитать среднее арифметическое.
10. Продублировать первое наблюдения 1 раз, второе — 2 раза, третье — 3 раза, и так далее и посчитать среднее арифметическое.

25.3 Величины Y_i независимы и имеют функцию плотности

$$f(y) = \begin{cases} 5y^4/\theta^5, \text{ если } y \in [0; \theta]; \\ 0, \text{ иначе.} \end{cases}$$

1. Найдите ML оценку неизвестного параметра θ.
2. Устно, не производя вычислений, определите, является ли оценка $\hat{\theta} = \max\{Y_1, Y_2, \ldots, Y_n\}$ несмещённой.
3. Найдите функцию распределения Y_1, функцию распределения $\hat{\theta}$, функцию плотности $\hat{\theta}$.
4. Найдите $\mathbb{E}(\hat{\theta})$.
5. Если $\hat{\theta}$ смещённая, то скорректируйте оценку так, чтобы она стала несмещённой.

25.4 Величина Y имеет биномиальное распределение $Bin(n, p)$.

1. Является ли оценка $\hat{p} = Y/n$ для p несмещённой? Если является смещённой, то скорректируйте оценку так, чтобы она стала несмещённой.
2. Чему равна теоретическая дисперсия σ^2 величины Y?
3. Является ли оценка $\hat{\sigma}^2 = n\hat{p}(1-\hat{p})$ для σ^2 несмещённой? Если является смещённой, то скорректируйте оценку так, чтобы она стала несмещённой.

25.5 Величины X_i независимы и одинаково распределениы. Какая из приведенных оценок для $\mu = \mathbb{E}(X_i)$ является несмещенной? Обладает наименьшей дисперсией среди несмещённых оценок? Обладает наименьшей среднеквадратичной ошибкой MSE?

1. $X_1 + 3X_2 - 2X_3$;
2. $(X_1 + X_2)/2$;
3. $(X_1 + X_2 + X_3)/3$;
4. $(X_1 + \ldots + X_{20})/21$;
5. $X_1 - 2X_2$.

25.6 Величина X равномерна на $[0; a]$. Придумайте несмещённую оценку параметра a вида $\hat{a} = \alpha + \beta X$.

25.7 Величины X_i — независимы и одинаково распределены. При каком значении параметра β

1. оценка $2X_1 - 5X_2 + \beta X_3$ будет несмещённой для $\mathbb{E}(X_i)$?
2. оценка $\beta (X_1 + X_2 - 2X_3)^2$ будет несмещённойдля $\text{Var}(X_i)$?

25.8 Величины X_1 и X_2 независимы и равномерны на $[0; a]$. При каком β оценка $Y = \beta \cdot \min\{X_1, X_2\}$ для параметра a будет несмещённой?

25.9 Величины X_i независимы и одинаково распределены. Какая из приведенных оценок для $\sigma^2 = \text{Var}(X_i)$ является несмещённой?

1. $X_1^2 - X_1X_2$;
2. $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$;
3. $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$;
4. $\frac{1}{2}(X_1 - X_2)^2$;
5. $X_1 - 2X_2$;
6. X_1X_2.

25.10 Величина X равномерна на $[3a - 2; 3a + 7]$. При каких α и β оценка $Y = \alpha + \beta X$ неизвестного параметра a будет несмещённой?

25.11 Закон распределения величины X имеет вид

1.

x_i	0	1	a
$P(X = x_i)$	1/4	1/4	2/4

;

2.

x_i	0	1	2
$P(X = x_i)$	1/4	a	$(3/4 - a)$

Постройте несмещённую оценку вида $Y = \alpha + \beta X$ для неизвестного параметра a.

25.12 Время горения лампочки распределено экспоненциально с ожиданием равным θ . Вася включил одновременно 20 лампочек. Величина X обозначает время самого первого перегорания. Как с помощью X построить несмещенную оценку для θ ?

25.13 Величины X_i независимы и одинаково распределены, причем $\text{Var}(X_i) = \sigma^2$, а $\mathbb{E}(X_i) = \frac{\theta}{\theta+1}$, где $\theta > 0$ — неизвестный параметр. С помощью $\bar{X}$ постройте состоятельную оценку для θ.

25.14 Величины $Y_i = \beta x_i + \varepsilon_i$, константа β и случайные величины ε_i являются ненаблюдаемыми. Известно, что $\mathbb{E}(\epsilon_i) = 0$, $\text{Var}(\epsilon_i) = 1$, ϵ_i являются независимыми. Константы x_i наблюдаемы, и известно, что $20 < x_i < 100$. У исследователя есть две оценки для β: $\hat{\beta}_1 = \frac{\bar{Y}}{\bar{x}}$ и $\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2}$.

1. Проверьте несмещенность, состоятельность.
2. Определите, какая оценка из двух является наиболее эффективной.

25.15 Исследователи Иван да Марья интересуются, какая доля населения берёт взятки. Они независимо друг от друга задают разным людям вопрос: «Берёте ли Вы взятки?»

Иван использует следующий механизм: предлагает респонденту тайно подбросить правильную монетку, если монетка выпадает орлом, то предлагается ответить «да», если решкой — то предлагается ответить правду. Предположим, что все респонденты действую точно так, как предлагает Иван.

У Марьи кристально чистые голубые глаза, она юна и невинна, и солгать ей просто невозможно.

Пусть p — доля берущих взятки, а $\hat{q}_I$ и $\hat{q}_M$ — доля ответивших «да» на вопросы Ивана да Марьи. Марья использует оценку $\hat{p}_M = \hat{q}_M$, а Иван — $\hat{p}_I = 2\hat{q}_I - 1$.

1. Являются ли оценки $\hat{p}_M$, $\hat{p}_I$ несмещёнными? состоятельными?
2. Иван планирует опросить 100 человек. Сколько человек в зависимости от p нужно опросить Марье, чтобы $\mathrm{Var}(\hat{p}_I) = \mathrm{Var}(\hat{p}_M)$?

Идея: Кирилл Фурманов

25.16 Величины X_i независимы от друг друга равны единице с вероятностью p и нулю с вероятностью $1-p$.

Рассмотрим две оценки $\hat{p}_1 = \frac{\sum_{i=1}^{10} X_i}{10}$ и $\hat{p}_2 = \frac{\sum_{i=1}^{10} X_i}{11}$.

При каких значениях параметра p вторая оценка будет лучше по критерию среднеквадратичной ошибки MSE?

26. Классические интервальные оценки

Здесь интервальные оценки без гипотез :)

26.1 Пусть X — равномерна на участке $[0; 2a]$. С какой вероятностью интервал $[0.9X; 1.1X]$ накрывает неизвестное a? Постройте 95%-ый доверительный интервал для a вида $[0; kX]$.

26.2 Пусть X — экспоненциальна с параметром λ и $\mu = \mathbb{E}(X)$. С какой вероятностью интервал $[0.9X; 1.1X]$ накрывает μ? Постройте 90%-ый доверительный интервал для μ вида $[0; kX]$.

26.3 Пусть X_i — независимы и нормальны $\mathcal{N}(\mu, 1)$. Какова вероятность того, что интервал $[\bar{X}_{10} - 1; \bar{X}_{10} + 1]$ накроет неизвестное μ? Постройте 90%-ый доверительный интервал для μ вида $[\bar{X}_{10} - k; \bar{X}_{10} + k]$.

26.4 Вася наугад поймал 400 покемонов-девочек и 100 покемонов-мальчиков. Среди девочек 250 оказались ядовитыми, среди мальчиков — 60.

1. Найдите точечную оценку $\hat{p}$ для доли ядовитых покемонов среди мальчиков.
2. Найдите точечную оценку $\hat{\sigma}^2$ для дисперсии X_i, где X_i — индикатор ядовитости i-го покемона-мальчика.
3. Постройте 95%-ый доверительный интервал для доли ядовитых покемонов среди мальчиков.
4. Постройте 95%-ый доверительный интервал для доли ядовитых покемонов среди девочек.
5. Постройте 95%-ый доверительный интервал для разницы долей ядовитых покемонов среди девочек и мальчиков.
6. Нужно ли предположение о нормальности X_i и Y_i для решения предыдущих пунктов?

26.5 Вася наугад поймал 400 покемонов-девочек и 100 покемонов-мальчиков. Средний рост покемонов-девочек равен 0.9 метра, и сумма квадратов ростов равна $\sum Y_i^2 = 1000$. Для покемонов мальчиков средний рост равен 1 метру, а сумма квадратов ростов равна $\sum X_i^2 = 2000$.

1. Постройте точечную оценку для ожидания роста покемона-мальчика μ_X.
2. Постройте точечную оценку для дисперсии роста покемона-мальчика σ_X^2.
3. Постройте 95%-ый доверительный интервал для ожидаемого роста покемона-мальчика.
4. Постройте 95%-ый доверительный интервал для ожидаемого роста покемона-девочки.
5. Постройте 95%-ый доверительный интервал для разницы ожидаемых ростов покемона-мальчика и покемона-девочки.
6. Нужно ли предположение о нормальности X_i и Y_i для решения предыдущих пунктов?

26.6 В одном тропическом лесу длина удавов имеет нормальное распределение $\mathcal{N}(\mu, \sigma^2)$. По выборке из 10 удавов оказалось, что $\sum Y_i = 20$ метрам, а $\sum Y_i^2 = 1000$.

1. Постройте 95%-ый доверительный интервал для μ.
2. Постройте 95%-ый доверительный интервал для σ^2.
3. Важна ли предпосылка о нормальности при решении предыдущих пунктов?

26.7 В одном тропическом лесу водятся удавы и питоны. Длина удавов имеет нормальное распределение $\mathcal{N}(\mu_X, \sigma_X^2)$. По выборке из 10 удавов оказалось, что $\sum X_i = 20$ метрам, а $\sum X_i^2 = 1000$. Длина питонов имеет нормальное распределение $\mathcal{N}(\mu_Y, \sigma_Y^2)$. По выборке из 20 питонов оказалось, что $\sum Y_i = 60$ метрам, а $\sum Y_i^2 = 4000$.

1. Постройте точечные оценки для μ_X, σ_X^2, μ_Y, σ_Y^2.
2. Постройте 95%-ый доверительный интервал для σ_X^2/σ_Y^2.
3. Постройте 95%-ый доверительный интервал для разницы $\mu_X - \mu_Y$ предполагая равенство дисперсий $\sigma_X^2 = \sigma_Y^2$.
4. Важна ли предпосылка о нормальности при решении предыдущих пунктов?

27. Проверка гипотез: общая теория

27.1 Кальямпуди Радхакришна Рао и Карл Харальд Крамер строят доверительный интервал для μ по случайной выборке из n наблюдений. Наблюдения, величины X_i, распределены нормально и независимы друг от друга. Рао знает величину σ^2; а Крамер не знает, и поэтому вынужден использовать $\hat{\sigma}^2$ при проверке гипотезы.

1. Какова вероятность того, что Рао получит более короткий доверительный интервал при $n = 10$?
2. К чему стремится эта вероятность при $n \to \infty$?

27.2 Как распределено P-значение при верной H_0? Вовочка использует следующий статистический критерий: «Если P-значение больше 0.95, то H_0 отвергается». Чему де-факто равна вероятность ошибки первого рода в этом случае? Разумно ли использовать данный критерий?

27.3 Величины X_1 и X_2 независимы и равномерны на отрезке $[0; a]$. Есть две гипотезы, H_0: $a = 1$ и H_a: $a = 2$. Мальвина отвергает H_0 в том случае, если $X_1 + X_2 > 1.5$. Найдите вероятность ошибок первого и второго рода.

27.4 Величины X_1 и X_2 независимы и нормальны $\mathcal{N}(a; 1)$. Есть две гипотезы, H_0: $a = 1$ и H_a: $a = 2$. Мальвина отвергает H_0 в том случае, если $X_1 + X_2 > 1.5$. Найдите вероятность ошибок первого и второго рода.

27.5 Величины X_1 и X_2 независимы и распределены экспоненциально с интенсивностью a. Есть две гипотезы, H_0: $a = 1$ и H_a: $a = 2$. Мальвина отвергает H_0 в том случае, если $\min\{X_1, X_2\} < 1$. Найдите вероятность ошибок первого и второго рода.

27.6 Величины X_1 и X_2 независимы и распределены по Пуассону с интенсивностью a. Есть две гипотезы, H_0: $a = 1$ и H_a: $a = 2$. Мальвина отвергает H_0 в том случае, если $X_1 + X_2 \geqslant 2$. Найдите вероятность ошибок первого и второго рода.

27.7 Бабушка Аксинья утверждает, что обладает сверхспособностями и умеет слышать Внутренний Голос. Стоит лишь Аксинье глянуть на стакан с водой, как Внутренний Голос подсказывает, что налито в стакан, Бон-Аква или Аква Минерале.

Исследователь Кирилл проводит с Аксиньей следующий опыт. Правила опыта Аксинье известны. Кирилл в тайне от Аксиньи наливает в 3 стакана Акву Минерале и 2 стакана Бон-Аквы. Затем предлагает их Аксинье в случайной порядке. Задача Аксиньи после осмотра всех стаканов определить, в каком порядке они предлагались.

Кирилл проверяет две гипотезы. Нулевую H_0: Аксинья не обладает сверхспособностями и её Внутренний Голос верно определяет содержимое каждого стакана с вероятностью $p = 0.5$ и альтернативную H_a: Аксинья обладает сверхспособностями и $p = 0.9$.

Критерий: если Аксинья ошиблась хотя бы один раз, то H_0 не отвергается; если не ошиблась ни разу, то H_0 отвергается.

1. Предположим, Аксинья, не задумываясь, говорит ровно то, что подсказывает ей Внутренний Голос. Найдите вероятности ошибок первого и второго рода.
2. (*) Предположим, Аксинья осознаёт что, Внутренний Голос может ошибаться с вероятностью $1 - p$. И поэтому при очевидной ошибке Внутренного Голоса старается её исправить. Например, если Внутренний Голос говорит ААААБ, то, следуя ему, угадать все стаканы невозможно. Найдите вероятности ошибок первого и второго рода.

27.8 Пусть $X_1, \ldots, X_n$ — случайная выборка из нормального распределения с неизвестным математическим ожиданием μ и известной дисперсией $\sigma^2 = 4$.

Объем выборки $n = 16$. Тестируются основная гипотеза $H_0 : \mu = 0$ против альтернативной гипотезы $H_a : \mu = 2$. С помощью леммы Неймана–Пирсона найдите наиболее мощный критерий, имеющий уровень значимости $\alpha = 0.05$.

27.9 Величины X_1 и X_2 независимы и одинаково распределены на отрезке $[0; 1]$. Есть две гипотезы, H_0: $X_i \sim U[0; 1]$ и H_a: $f(x_i) = \begin{cases} 2x_i, & \text{если } x_i \in [0;1]; \\ 0, & \text{иначе} \end{cases}$. С помощью леммы Неймана–Пирсона найдите наиболее мощный критерий, имеющий уровень значимости $\alpha = 0.05$.

27.10 Величины X_1 и X_2 одинаково распределены на отрезке $[0; 1]$. Есть две гипотезы, H_0: $X_i \sim U[0; 1]$ и H_a: $f(x_1, x_2) = \begin{cases} x_1 + x_2, & \text{если } x_1, x_2 \in [0;1]; \\ 0, & \text{иначе} \end{cases}$. С помощью леммы Неймана–Пирсона найдите наиболее мощный критерий, имеющий уровень значимости $\alpha = 0.05$.

27.11 Пусть $X_1, \ldots, X_n$ — случайная выборка из нормального распределения с известным математическим ожиданием $\mu = 1$ и неизвестной дисперсией σ^2.

Объем выборки $n = 16$. Тестируются основная гипотеза $H_0 : \sigma^2 = 4$ против альтернативной гипотезы $H_a : \sigma^2 = 9$. С помощью леммы Неймана–Пирсона найдите наиболее мощный критерий, имеющий уровень значимости $\alpha = 0.05$.

28. Таблицы сопряжённости

28.1 Помимо широко известной системы групп крови AB0, существуют также и другие, например система MN. В этой системе существует три группы крови: M, N и MN. В гене существует всего два аллеля, M и N, кодоминантные по отношению друг к другу. Обладатели группы крови M имеют гомозиготный генотип MM, обладатели группы N — гомозиготный NN, обладатели группы MN — гетерозиготный MN. Наследование систем групп крови AB0 и MN происходит независимо друг от друга.

В идеальных условиях частоты генотипов должны находится в равновесии Харди-Вайнберга. Частоты генотипов должны быть равны: $p_{NN} = p_N^2$, $p_{MN} = 2p_N(1-p_N)$, $p_{MM} = (1-p_N)^2$, где p_N — частота аллеля N в популяции.

У 50 людей определили их генотипы: оказалось 10 человек с генотипом NN, 20 — с генотипом MM, и 20 — с генотипом MN.

1. С помощью критерия хи-квадрат Пирсона проверьте гипотезу, что условия Харди-Вайнберга выполнены и $p_N = 0.3$.
2. В рамках условий Харди-Вайнберга оцените неизвестную вероятность p_N с помощью максимального правдоподобия.
3. С помощью критерия хи-квадрат Пирсона проверьте гипотезу, что условия Хадри-Вайнберга выполнены.

29. Многомерщина

29.1 Известна ковариационная матрица и математическое ожидание вектора y:

$$\mathbb{E}(y) = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}, \mathrm{Var}(y) = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 4 & 2 \\ -1 & 2 & 9 \end{pmatrix}$$

Найдите $\mathbb{E}(Ay)$ и $\mathrm{Var}(Ay)$, где $A = \begin{pmatrix} 2 & 0 & 2 \\ -1 & 2 & 3 \end{pmatrix}$

29.2 Пусть $A = X(X^TX)^{-1}X^T$ и матрица X^TX обратима.

1. Найдите A^2, A^{2017}, A^T.
2. В каком случае матрица A обратима?

29.3 Найдите матрицу A для каждой из ситуаций:

1. Матрица A проецирует n-мерные вектора на вектор $\nVdash = (1, 1, 1, \ldots, 1)^T$.
2. Матрица A проецирует 3-мерные вектора на вектор $(1, 2, 9)^T$.
3. Матрица A проецирует 3-мерные вектора на плоскость, порождённую векторами $(1, 1, 1)^T$ и $(1, 2, 3)^T$.

29.4 Про каждую из матриц проверьте, является ли она проектором. ... Для матриц-проекторов определите, на какие вектора они проецируют.

29.5 Известно, что ... Найдите закон распределения $y^TB^{-1}y$, y^TP^y....

29.6 Определим информацию Фишера как ковариационную матрицу вектора $\frac{\partial \ell}{\partial \theta}$, $I = \mathrm{Var}\left(\frac{\partial \ell}{\partial \theta}\right)$.

1. Найдите $\mathbb{E}\left(\frac{\partial \ell}{\partial \theta}\right)$.
2. Докажите, что $I = -\mathbb{E}(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T})$
3. Докажите, что $I = \mathbb{E}(\frac{\partial \ell}{\partial \theta}\frac{\partial \ell}{\partial \theta}^T)$
4. Докажите, что $\operatorname{Var}(\hat{\theta}) \operatorname{Var}(\frac{\partial \ell}{\partial \theta}) - I$ является положительно определённой

30. Самая главная компонента

30.1 Найдите максимум функции $2x^2 + 5y^2 + 3z^2 + 7w^2$ при ограничении $x^2 + y^2 + z^2 + w^2 = 1$.

30.2 Пусть $A = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$.

1. Найдите собственные векторы и собственные числа матрицы A
2. Найдите максимум $v^T A v$ при ограничении $v^T v = 1$.
3. Найдите максимум $v^T A v$ при ограничениях $v^T v = 1$ и $v \perp a$, где $a^T = (1, 1, 0)$.

30.3 Вектора z_1 и z_2 имеют выборочную ковариационную матрицу

$$M = \begin{pmatrix} 6 & 2 \\ 2 & 10 \end{pmatrix}.$$

1. Как изменится коварицонная матрица, если центрировать вектора z_1 и z_2?
2. Как изменится коварицонная матрица, если центрировать и нормировать вектора z_1 и z_2?

30.4 Вектора-столбцы z_1 и z_2 содержат по пять наблюдений. Матрица X состоит из столбцов x_1 и x_2. Выборочная ковариационная матрица векторов z_1 и z_2 равна $\begin{pmatrix} 6 & 2 \\ 2 & 10 \end{pmatrix}$.

1. Найдите матрицу $X^T X$, если вектора x_i получены центрированием векторов z_i.
2. Найдите матрицу $X^T X$, если вектора x_i получены центрированием и нормирование векторов z_i.

30.5 Вениамин находит главные компоненты набора данных из трёх переменных. По каждой из переменных есть 100 наблюдений. Вениамин центрирует и не нормирует переменные, так как они изначально измеряются в одних и тех же единицах.

Собственные числа выборочной ковариационной матрицы исходных переменных равны 5, 4 и 1.

1. Найдите сумму выборочных дисперсий исходных переменных.
2. Найдите длины и выборочные дисперсии всех трёх главных компонент.
3. Какую долю от суммы выборочных дисперсий объясняют первые две главные компоненты?

30.6 Рассмотрим результаты пяти студентов за две контрольные работы:

ФИО	Контрольная 1	Контрольная 2
Маша	4	5
Вася	5	5
Лена	3	4
Коля	5	4
Рита	4	3

1. Найдите выборочную ковариационную матрицу.
2. Найдите собственные числа и собственные векторы единичной длины для ковариационной матрицы.
3. Выпишите первую и вторую главные компоненты.
4. Найдите сумму выборочных дисперсий исходных переменных.
5. Найдите длины и выборочные дисперсии всех главных компонент.
6. Какую долю от суммы выборочных дисперсий объясняет первая главная компоненты?

31. Я не врач!

31.1 Какие 10 лекарств являются лидерами по объёму продаж в денежном выражении в России за прошлый год? Сколько фаз клинических испытаний каждое из них прошло согласно https://www.drugbank.ca?

31.2 что-то про шансы

31.3 здесь LR тест в качестве аналога MH теста для агрегирования отдельных таблиц сопряженности

32. И потребленье возрастает, а производство отстаёт!

Испить мудрость жадными глотками можно в источнике [Wil13].
Обобщенный бином Ньютона для произвольной степени $a \in \mathbb{R}$:

$$(1+x)^a = \sum_{k=0}^{\infty} C_a^k x^k, \quad \text{где } C_a^k = \frac{a(a-1)(a-2)\cdot \ldots \cdot (a-k+1)}{k!}$$

32.1 Найдите коэффициент при x^{17} в многочлене $(1+x^5+x^7)^{20}$.

32.2 Найдите и запишите максимально просто производящие функции для последовательностей

1. 1, 1, 1, 1, 0, 0, 0, …
2. 1, 1, 1, 1, 1, 1, 1, …
3. 1, 4, 6, 4, 1, 0, 0, …

32.3 Найдите первые четыре коэффициента в многочленах $(1+2x)^{-3}$ и $(1+x+x^2)^{-4}$.

32.4 Кубик подбрасывают три раза. Обозначим результаты подбрасываний X_1, X_2 и X_3. Найдите производящие функции для величин X_1, $X_1 + X_2$ и $X_1 + X_2 + X_3$.

32.5 Производящая функция случайной величины X линейна и проходит через точки $(1, 1)$ и $(5, 4)$. Какие значения принимает величина X и с какими вероятностями?

32.6 На месте убийства преступники оставили улику — маленький-маленький кусочек графика производящей функции. Сможет ли Шерлок Холмс по этому кусочку восстановить все вероятности?

32.7 Леонардо Пизанский (Фибоначчи) приобрёл новорождённую пару кроликов. Каждая взрослая пара кроликов даёт в качестве приплода ежемесячно новую пару кроликов. Только что рождённая пара кроликов начинает приносить приплод со второго месяца жизни. Пусть F_n — количество пар кроликов в n-ый месяц.

1. Каким соотношением связаны F_n, F_{n-1} и F_{n-2}?
2. Пусть для удобства $F_0 = 0$ и $F_1 = 1$. Найдите производящую функцию для количества пар кроликов.
3. Используя хороший софт, поделите 1 на 99999999999999999999999899999999999999999999999999. Здесь 23 девятки, восьмёрка и потом 24 девятки. Почему мы видим последовательность чисел Фибонначи? Например, можно поделить в julia установив высокую точность:

```julia
setprecistion(BigFloat, 10000)
1 / 999999999999999999999998999999999999999999999999
```

4. Разложите найденную функцию в сумму двух дробей и, вспомнив формулу для геометрической прогрессии, найдите явную формулу для чисел Фибоначчи.

32.8 Сева стартует в точке 0 на числовой прямой. За каждый час он перемещается вправо на единицу с вероятностью 0.7 или влево на единицу с вероятностью 0.3. Пусть X_1 — момент времени, когда Сева впервые достигнет точки 1, а X_2 — точки 2.

1. Как связаны производящие функции $h_1(t)$ и $h_2(t)$ величин X_1 и X_2?
2. Рассмотрев возможные положения Севы после первого шага, составьте квадратное уравнение на $h_1(t)$.
3. Найдите $h_1(t)$.
4. Найдите $\mathbb{E}(X_1)$.
5. Выпишите формулу для $\mathbb{P}(X_1 = 2017)$.

32.9 Ангела Меркель подбрасывает правильную монетку до выпадения трёх орлов подряд. Найдите производящую функцию для количества подбрасываний.

33. Сэр Томас Байес

Байесовский подход (bayesian approach):

1. Сделать изначальное априорное предположение о распределении θ, $p(\theta)$;
2. Сформулировать модель для данных, $p(data|\theta)$;

3. Получить апостериорный закон распределения θ по формуле условной вероятности, $p(\theta|data) \propto p(\theta) \cdot p(data|\theta)$

33.1 В своём труде 1763 года «An Essay towards solving a Problem in the Doctrine of Chances» Томас Байес решает следующую задачу: «Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named». А вам слабо?

Имеется монетка, возможно неправильная. Мы не знаем вероятность выпадения орла α, поэтому считаем, что α равномерно распределена на отрезке $[0; 1]$.

1. Какова безусловная вероятность того, что α лежит в диапазоне $[0; 0.5]$?
2. Какова условная вероятность того, что α лежит в диапазоне $[0; 0.5]$, если монетка выпала 5 раз орлом и 7 раз решкой?
3. В байесовском подходе α — это константа или случайная величина?
4. В каком году умер Томас Байес?

http://rstl.royalsocietypublishing.org/content/53/370.full.pdf

33.2 Время, которое Вася тратит на задачу — равномерно распределенная случайная величина: на простую — от 1 до 15 минут, на сложную — от 10 до 20 минут. Известно, что на некую задачу Вася потратил 13 минут.

1. С помощью метода максимального правдоподобия определите, простая она или трудная.
2. С помощью байесовского подхода посчитайте вероятности того, что задача была простая, если на экзамене было 7 легких и 3 трудных задачи.

34. А энтропия никогда не убывала

34.1 Найдите энтропию X, спутанность (perplexity) X, индекс Джини X, если

1. величина X равновероятно принимает значения 1, 7 и 9;
2. величина X равновероятно принимает $k \geqslant 2$ значений;
3. величина X равномерно распределена на отрезке $[0; a]$;
4. величина X нормальна $\mathcal{N}(\mu; \sigma^2)$;

34.2 У Васи была дискретная случайная величина X, принимавшая натуральные значения. Вася решил изменить закон распределения величины X. Он увеличил количество возможных значений величины X в два раза, разделив каждое событие $X = k$ на два равновероятных подсобытия: $X = k - 0.1$ и $X = k + 0.1$. Как при этом изменились энтропия, спутанность (perplexity) и индекс Джини?

34.3 Найдите дивергенцию Кульбака-Лейблера

1. из биномиального $Bin(n = 2, p = 1/3)$ в равновероятное на $0, 1, 2$;
2. из равновероятного на $0, 1, 2$ в биномиальное $Bin(n = 2, p = 1/3)$;
3. из $\mathcal{N}(0; l)$ в $\mathcal{N}(0; \sigma^2)$;
4. из $\mathcal{N}(0; 1)$ в экспоненциальное с $\lambda = 1$;

5. из $\mathcal{N}(0; \sigma^2)$ в $\mathcal{N}(0; l)$;
6. из экспоненциального с $\lambda = 1$ в $\mathcal{N}(0; 1)$;

34.4

34.5

35. Шпаргалка

35.1. Определения

1. Оценка $\hat{a}$ неизвестного параметра a называется *несмещённой*, если $\mathbb{E}(\hat{a}) = a$.

2. Последовательность случайных величин R_1, R_2, ...сходится к величине R по вероятности, если для любого положительного числа $\varepsilon > 0$ вероятность отклонения R_i от R больше, чем на ε, стремится к нулю:
$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0, \ \forall \varepsilon > 0$$
Сходимость по вероятности обозначается с помощью оператора plim, $\operatorname{plim}_{n\to\infty} R_n = R$:

3. Последовательность оценок $\hat{a}_1$, $\hat{a}_2$, ...неизвестного параметра a называется *состоятельной*, если $\operatorname{plim} \hat{a}_n = a$.

4. Последовательность случайных величин Z_n сходится к случайной величине Z по вероятности, если для любого положительного числа ε оказывается, что $\lim \mathbb{P}(|Z_n - Z| > \varepsilon) = 0$. Сходимость по вероятности записывается как $\operatorname{plim} Z_n = Z$.

5. Среднеквадратическое отклонение оценки $\hat{a}$ от истинного значение параметра a, $MSE(\hat{a}) = \mathbb{E}((\hat{a} - a)^2)$. По теореме Пифагора величина MSE представима в виде $MSE(\hat{a}) = \operatorname{Var}(\hat{a}) + (\mathbb{E}(\hat{a}) - a)^2$. Если оценка несмещённая, то $MSE = \operatorname{Var}(\hat{a})$.

6. Оценка $\hat{a}$ неизвестного параметра a называется эффективной среди некоторого набора оценок, если оценка $\hat{a}$ обладает наименьшей среднеквадратичной ошибкой MSE среди рассматриваемого набора оценок. Если рассматриваемые оценки являются несмещёнными, то эффективная оценка обладает наименьшей дисперсией.

35.2. Гипотезы

35.2.1. Про единственную выборку

1. Гипотеза о математическом ожидании при большом количестве наблюдений

   а) Наблюдаем: X_1, X_2, ..., X_n;

   б) Предполагаем: X_i независимы и одинаково распределены (не обязательно нормально), количество наблюдений n велико.

   в) Проверяемая гипотеза: H_0: $\mu = \mu_0$ против H_a: $\mu \neq \mu_0$;

   г) Статистика:
$$Z = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

   д) При верной H_0 оказывается, что $Z \to \mathcal{N}(0; 1)$;

2. Гипотеза о математическом ожидании при нормальных наблюдениях

   а) Наблюдаем: X_1, X_2, ..., X_n;

   б) Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu;\sigma^2)$, количество наблюдений n может быть мало.

   в) Проверяемая гипотеза: H_0: $\mu = \mu_0$ против H_a: $\mu \neq \mu_0$;

   г) Статистика:
$$t = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

   д) При верной H_0 оказывается, что $t \sim t_{n-1}$;

3. Гипотеза о математическом ожидании при нормальных наблюдениях и известной дисперсии

   а) Наблюдаем: X_1, X_2, ..., X_n, знаем величину σ^2;

   б) Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu;\sigma^2)$, количество наблюдений n может быть мало.

   в) Проверяемая гипотеза: H_0: $\mu = \mu_0$ против H_a: $\mu \neq \mu_0$;

   г) Статистика:
$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

   д) При верной H_0 оказывается, что $Z \sim \mathcal{N}(0;1)$;

4. Гипотеза о вероятности при наблюдениях с распределением Бернулли (0 или 1)

   а) Наблюдаем: X_1, X_2, ..., X_n

   б) Предполагаем: X_i независимы и имеют распределение Бернулли: равны 1 с вероятностью p и 0 с вероятностью $1-p$. Количество наблюдений n велико.

   в) Проверяемая гипотеза: H_0: $p = p_0$ против H_a: $p \neq p_0$;

   г) Статистика:
$$Z = \frac{\hat{p} - p_0}{se(\hat{p})} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

   Возможен вариант этой статистики:
$$Z = \frac{\hat{p} - p_0}{se(\hat{p})} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

   д) При верной H_0 оказывается, что $Z \to \mathcal{N}(0;1)$;

   е) Гипотеза о вероятностях является частным случаем гипотезы о математическом ожидании при большом количестве наблюдений. Можно заметить, что $\hat{p} = \bar{X}$ и $\hat{\sigma}^2 = \hat{p}(1-\hat{p}) \cdot \frac{n}{n-1}$. И потому также корректен вариант статистики
$$Z = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

5. Гипотеза о дисперсии при нормальных наблюдениях

а) Наблюдаем: X_1, X_2, ..., X_n

б) Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu;\sigma^2)$, количество наблюдений n может быть мало.

в) Проверяемая гипотеза: H_0: $\sigma=\sigma_0$ против H_a: $\sigma\neq\sigma_0$;

г) Статистика:
$$S=\frac{\sum(X_i-\bar{X})^2}{\sigma_0^2}=\frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}$$

д) При верной H_0 оказывается, что $S\sim\chi^2_{n-1}$;

35.2.2. Про пару выборок

6. Гипотеза о разнице ожиданий при большом количестве наблюдений

а) Наблюдаем: X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}. Возможно, что $n_x\neq n_y$. Дисперсии σ^2_x и σ^2_y не знаем и не уверены, что они равны.

б) Предполагаем: X_i одинаково распределены между собой (не обязательно нормально), Y_i одинаково распределены между собой, но возможно совсем не так, как X_i (не обязательно нормально). Все величины независимы. Количества n_x и n_y велики.

в) Проверяемая гипотеза: H_0: $\mu_x-\mu_y=\delta_0$ против H_a: $\mu_x-\mu_y\neq\delta_0$;

г) Статистика:
$$Z=\frac{\bar{X}-\bar{Y}-\delta_0}{se(\bar{X}-\bar{Y})}=\frac{\bar{X}-\bar{Y}-\delta_0}{\sqrt{\frac{\hat{\sigma}^2_x}{n_x}+\frac{\hat{\sigma}^2_y}{n_y}}}$$

д) При верной H_0 оказывается, что $Z\to\mathcal{N}(0;1)$;

7. Гипотеза о разнице ожиданий при нормальности распределения обеих выборок и известных дисперсиях

а) Наблюдаем: X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}. Возможно, что $n_x\neq n_y$. Дисперсии σ^2_x и σ^2_y знаем. Возможно, что дисперсии не равны.

б) Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x,\sigma^2_x)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y,\sigma^2_y)$. Все величины независимы. Количества n_x и n_y любые.

в) Проверяемая гипотеза: H_0: $\mu_x-\mu_y=\delta_0$ против H_a: $\mu_x-\mu_y\neq\delta_0$;

г) Статистика:
$$Z=\frac{\bar{X}-\bar{Y}-\delta_0}{\sigma_{\bar{X}-\bar{Y}}}=\frac{\bar{X}-\bar{Y}-\delta_0}{\sqrt{\frac{\sigma^2_x}{n_x}+\frac{\sigma^2_y}{n_y}}}$$

д) При верной H_0 оказывается, что $Z\sim\mathcal{N}(0;1)$;

8. Гипотеза о разнице ожиданий при нормальности распределения обеих выборок и неизвестных но равных дисперсиях

а) Наблюдаем: X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}. Возможно, что $n_x\neq n_y$. Дисперсии σ^2_x и σ^2_y равны, но неизвестны.

б) Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x,\sigma^2)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y,\sigma^2)$. Все величины независимы. Количества n_x и n_y любые.

в) Проверяемая гипотеза: H_0: $\mu_x - \mu_y = \delta_0$ против H_a: $\mu_x - \mu_y \neq \delta_0$;

г) Статистика:

$$t = \frac{\bar{X} - \bar{Y} - \delta_0}{se(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\hat{\sigma}^2}{n_x} + \frac{\hat{\sigma}^2}{n_y}}},$$

где

$$\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{n_x + n_y - 2}$$

д) При верной H_0 оказывается, что $t \sim t_{n_x + n_y - 2}$;

9. Гипотеза о разнице ожиданий в связанных парах

а) Наблюдаем: X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n. Количество X_i и Y_i одинаковое.

б) Предполагаем: внутри пары X_i и Y_i зависимы, а наблюдения с разными номерами независимы. Рассматриваем разницу $D_i = X_i - Y_i$ и получаем одномерную выборку. Величины D_i независимы и одинаково распределены. Возможно три описанных ранее случая :) Здесь для примера рассмотрим случай, когда $D_i \sim \mathcal{N}(\mu_d, \sigma_d^2)$ с неизвестной дисперсией.

в) Проверяемая гипотеза: H_0: $\mu_d = \mu_0$ против H_a: $\mu_d \neq \mu_0$;

г) Статистика:

$$t = \frac{\bar{D} - \mu_d}{se(\bar{D})} = \frac{\bar{X} - \bar{Y} - \mu_d}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}},$$

где

$$\hat{\sigma}_d^2 = \frac{\sum (D_i - \bar{D})^2}{n-1} = \frac{\sum (X_i - Y_i - (\bar{X} - \bar{Y}))^2}{n-1}$$

д) При верной H_0 оказывается, что $t \sim t_{n-1}$;

10. Гипотеза о равенстве дисперсий при нормальности распределения обеих выборок

а) Наблюдаем: X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}. Возможно, что $n_x \neq n_y$. Дисперсии σ_x^2 и σ_y^2 не знаем. Возможно, что дисперсии не равны.

б) Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x, \sigma^2)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y, \sigma^2)$. Все величины независимы. Количества n_x и n_y любые.

в) Проверяемая гипотеза: H_0: $\sigma_x = \sigma_y$ против H_a: $\sigma_x \neq \sigma_y$;

г) Статистика:

$$F = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$$

д) При верной H_0 оказывается, что $F \sim F_{n_x - 1, n_y - 1}$;

36. Решения

1.1. $\mathbb{P}(X = 1) = 3/5, \mathbb{P}(X = 2) = 3/10, \mathbb{P}(X = 3) = 1/10, \mathbb{E}(X) = 1.5$

1.2.

1.3.

1.4. N 3 4 5
2/8 3/8 3/8

1.5.

1.6. $6/11, 6$

1.7. N — количество подбрасываний, G — количество орлов, R — решек
$\mathbb{E}(N) = 10/7, \mathbb{E}(G) = 1, \mathbb{E}(R) = 10/7 - 1 = 3/7$

Для вероятности чётного числа бросков можно сложить ряд или составить уравнение. Если выпала решка, то в продолжении игры мы хотим получить нечётное количество бросков.

$$v = 0.7 \cdot 0 + 0.3 \cdot (1 - v)$$

1.8.

1.9. у Саши 3/4

1.10. стоп на 4-5-6 или стоп на 5-6

1.11. $\mathbb{P}(X = 4) = 1/8, \mathbb{E}(X) = 3$

1.12.

2.1.

2.2.

2.3.

2.4.

2.5.

2.6. 1/2

2.7. $\mathbb{E}(X) = 0.25 \cdot 2 + 0.5 \cdot (\mathbb{E}(X) + 1) + 0.25 \cdot (\mathbb{E}(X) + 2)$
$\mathbb{E}(Y) = 0.25 \cdot 0 + 0.5 \cdot (\mathbb{E}(Y) + 1) + 0.25 \cdot (\mathbb{E}(Y) + 1)$

2.8. стратегия 1: говорить стоп, если на кону 200 рублей вне зависимости от числа набранных красных карточек

стратегия 2: если нет красных карточек или одна, то останавливаться при 200 рублях, а при двух карточках останавливаться на 100 рублях.

2.9.

2.10.

2.11.

2.12.

2.13.

2.14.

2.15.

2.16.

2.17.

1. Выбрать три, чтобы взять, или выбрать семь, чтобы не взять.
2. Число дорог с 11 ходами вниз и 4 вправо равно сумму числа дорог.
3. Число всех подмножеств в множестве из 4-х элементов.
4. Из 10 человек хотим выбрать одного лидера и 3-х помощников.
5. Из 10 человек хотим выбрать 3-х орлов и 2-х слонов.
6. Число всех подмножеств с одним лидером в множестве из 5-ти элементов.

2.18.

3.1. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1/3}{1/6+1/3} = 2/3$

3.2. A — первый охотник попал в утку, B — в утку попала ровно одна пуля
$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{0.4\cdot 0.3}{0.4\cdot 0.3+0.6\cdot 0.7} = 2/9$

3.3. $\mathbb{P}(A \mid B) = 1/6$

3.4. Обозначим: X — количество попавшихся тузов, S — количество попавшихся тузов пик.

$$\mathbb{P}(X = 2) = \frac{C_4^2 C_{48}^{11}}{C_{52}^{13}}$$

$\mathbb{P}(X \geqslant 2 \mid X \geqslant 1) = \frac{\mathbb{P}(X\geqslant 2)}{\mathbb{P}(X\geqslant 1)} = \approx 0.37$

Можно представить, что мы берём первые 13 карт из случайно перемешанной колоды. Поэтому вероятность того, что туз пик попадет на одно из этих 13-и мест равна:

$$\mathbb{P}(S = 1) = \frac{13}{52}$$

Далее,

$$\mathbb{P}(X \geqslant 2 \mid S = 1) = \frac{\mathbb{P}(S = 1, X \geqslant 2)}{\mathbb{P}(S = 1)}$$

Ищем вероятность в числителе. Есть много способов, один из:

$$\mathbb{P}(S = 1, X \geqslant 2) = \mathbb{P}(S = 1) - \mathbb{P}(S = 1, X = 1) = \frac{13}{52} - \frac{1\cdot C_{48}^{12}}{C_{52}^{13}}$$

Итого, $\mathbb{P}(X \geqslant 2 \mid S = 1) \approx 0.56$.

3.5. $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{7 \cdot 11 / C^2_{7+5+11}}{(7 \cdot 11 + 5 \cdot 11 + 7 \cdot 5)/C^2_{7+5+11}}$

3.6. B — событие, второй — черный, $\mathbb{P}(B) = 11/16$
$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \frac{\frac{5}{16}\frac{11}{15}}{11/16}$

3.7. A — партия мяса заражена, B — партия мяса по тесту заражена
$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \frac{0.04 \cdot 0.9}{0.04 \cdot 0.9 + 0.96 \cdot 0.1} \approx 0.27$

3.8. A — Петя из Б класса, B — Петя обожает географию
$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \frac{0.4/3}{0.3/3 + 0.4/3 + 0.7/3} = 2/7$

3.9. $1/3$

3.10. да

3.11. нет, $(4/52) \cdot (3/51) = \mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) = (4/52)^2$

3.12.

3.13. $2/3$, $1/2$, $1/2$, $14/27$

4.1.

4.2.

4.3.

4.4.

4.5.

5.1.

5.2.

5.3.

5.4.

5.5.

5.6.

5.7.

5.8.

5.9.

5.10.

5.11.

5.12. при $c = 1/\mathbb{E}(X)$

5.13.

5.14.

5.15.

5.16.

6.1.

6.2.

6.3.

6.4.

7.1.

7.2.

7.3. 1.75

7.4. $\mathbb{E}(X_0) = 100 \cdot \frac{99}{464} + \frac{100}{464} + \ldots + \frac{464}{464}$

7.5.

7.6. $\mathbb{E}(X) = 1$, $\mathrm{Var}(X) = 1$

7.7. $\mathbb{E}(Y) = 20 - 20\left(\frac{19}{20}\right)^{10} \approx 8.025$

7.8. $\mathbb{E}(X) = 9$, $\mathrm{Var}(X) = 1.2$

7.9. $10 \cdot 0.9^7$

7.10. $\mathbb{E}(X) = \frac{30}{30} + \frac{30}{29} + \ldots + \frac{30}{1} \approx 119.85$
аппроксимация через логарифм $30 \ln 30 \approx 102$

7.11.

7.12.

7.13.

7.14.

8.1. $\mathbb{P}(X_8 = 13) = e^{-8.8} 8.8^{13}/13! \approx 0.046$
$\mathbb{P}(Y_{n+1} > 1) = e^{-1.1} \approx 0.33$

8.2.

8.3.

8.4.

8.5. $\frac{a}{a+b}$, экспоненциально с параметром $a+b$, $\frac{1}{a+b}$

8.6. геометрическое распределение
$\mathbb{E}(N) = \frac{\lambda_{in}}{\lambda_{capacity} - \lambda_{in}}$

8.7. The probability of observing a taxi before a bus is given by $3/(3+2) = 3/5$ since the waiting times are independent and exponentially distributed. By the memoryless property both processes then restart and hence the probability of observing (at least) two taxis before the first bus is $(3/5)^2 = 9/25$. The probability of observing exactly two taxis before the first bus is $(3/5)^2 * (2/5) = 18/125$.

8.8.

8.9.

8.10.

8.11.

8.12.

8.13.

8.14.

8.15.

8.16.

$$a(t+\Delta) = a(t)(1-\Delta) + (1-a(t))\Delta + o(\Delta)$$

сейчас четное = за секунду было четное и никто не зашел + за секунду было нечетное и один зашел $a'(t) = 1 - 2a(t)$, $a(0) = 1$ $a(t) = (1 + \exp(-2t))/2$

8.17.

8.18.

$$(X)_k = X \cdot (X-1) \cdot \ldots \cdot (X-k+1)$$

Для Пуассоновского распределения с $\lambda = 1$ получаем, что $\mathbb{E}((X)_k) = 1$.

Пусть π — произвольное разбиение множества A на помножества. Замечаем, что

$$X^n = \sum_{\pi} (X)_{|\pi|}$$

Берём ожидание и получаем:
$$\mathbb{E}(X^n) = \sum_{\pi} 1$$

То есть $\mathbb{E}(X^n)$ — это и есть количество различных разбиений множества из n элементов на подмножества!

И получаем, что количество разбиений на подмножества, целое число, представимо в виде:
$$B_n = \frac{1}{e} \sum \frac{k^n}{k!}$$

9.1.

9.2.

9.3.

9.4.

9.5. https://people.xiph.org/~greg/attack_success.html, ≈ 0.017

9.6.

10.1.

10.2.

10.3. $\text{Cov}(X, Y) = 0$, зависимы

10.4. $\text{Cov}(X, Y) = 0$, зависимы

10.5. $-1/5$

10.6.

10.7.

10.8. $\text{Corr}(X, Y) = 1$, $\text{Corr}(X, Z) = -1$

10.9.

10.10.

10.11. нет, например, берем независимые X и Z и возьмём $Y = X + Z$.

10.12. нет

10.13.

10.14.

10.15.

10.16.

10.17.

$$\mathbb{P}(X > 0.5) = 1 - \mathbb{P}(X \leqslant 0.5) = 1 - P(X \leqslant 0.5, Y \leqslant +\infty) = 1 - F(0.5, +\infty)$$

$F_X(x) = F(x, +\infty)$; $\mathbb{P}(X = Y + 0.2) = 0$ в силу непрерывности величин; X и Y независимы, так как $F(x, y) = F_X(x) \cdot F_Y(y)$;

10.18.

10.19. Например, $X \sim U[0; 1]$ и $Y = X$

10.20. $\mathrm{Cov}(X, Y) = 0$, зависимы

10.21. корреляция равна 0, зависимы, так как знание X несёт информацию об Y, например, при $X = 0$ можно утверждать, что $|Y| \in [1; 2]$.

10.22.

10.23.

10.24.

10.25. X равномерна на $[-1; 1]$; пара (X, Y) равномерна в круге $x^2 + y^2 \leqslant 1$.

11.1.

11.2. теорема Пифагора

11.3.

11.4. теорема Пифагора

11.5. теорема Пифагора

12.1.

12.2.

12.3. $x_{max} = \mu$, $x = \mu \pm \sigma$

12.4.

$$f_{|X|}(t) = \begin{cases} 2f_X(t), \ t > 0 \\ 0, \quad \text{иначе} \end{cases}$$

12.5. медиана — $\exp(\mu)$ мода — $\exp(\mu - \sigma^2)$ $\mathbb{E}(Y) = \exp(\mu + \sigma^2/2)$ $\text{Var}(Y) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$

12.6.

12.7. Применяем формулу интегрирования по частям.

12.8.

13.1. b $4/20 \geqslant 1 - 100/12$
c $1 - e^{-3} \geqslant 0.75$

13.2. $\text{Var}(X) \leqslant 5$

13.3. $\mathbb{P}(X < 20) \geqslant 0.5$

13.4.

13.5. $\text{Var}(X) = 0$, так как X почти наверное константа;

13.6.

13.7. Потратили одинаково, молока купил больше Кот Матроскин.

13.8. У Кота Матроскина.

14.1.

14.2.

14.3.

14.4. $S_{100} \sim \mathcal{N}(1000, 100)$, $\mathbb{P}(S_{100} > 1030) = \mathbb{P}(Z > 3) = 0.0013$

14.5.

14.6.

14.7.

14.8. $4(\sqrt{2} - 1)/3$, 20 problems

14.9. $2/\pi$, 20 problems

15.1.

15.2.

15.3.

15.4.

15.5.

15.6.

15.7.

15.8.

15.9.

15.10.

15.11.

15.12.

15.13.

15.14.

15.15.

15.16.

15.17.

16.1. Ind, Rot, Rot + Ind

16.2. Чтобы сохранялись длины векторов и углы между векторами, должно сохраняться скалярное произведение. Значит $\langle Vx, Vy\rangle = \langle x, y\rangle$.

Или $x'V'Vy = x'y$. Необходимо и достаточно, чтобы $V'V = I$.

16.3. [Muk17]

16.4.

16.5.

16.6. $K = 170$, $M = 120$ (симметричный интервал) или $K = M = 168$ (площадь с одного края можно принять за 0)
Вариант: театр, два входа, два гардероба а) только пары, б) по одному

16.7.

16.8. $\mathrm{Corr}(Y_1, Y_2) = 2/5$, $E(Y_t|Y_{t-1}) = 0.4Y_{t-1}$ В среднем Машка не становиться ни грустней, ни улыбчивей Представить $\mathbb{E}(Y_t|Y_{t-1}, Y_{t-2})$ в виде $\mathbb{E}(Y_t|Y_{t-1}, Y_{t-2}) = a_1Y_{t-1} + a_2Y_{t-2} + Z$

16.9.

16.10.

16.11.

16.12.

16.13. $Y \sim \mathcal{N}(0;1)$, нет, они нормальны только по отдельности, но не в совокупности, $\mathrm{Corr}(X,Y) = 0$. Взять $Y = XZ$, где Z принимает значение 1 с вероятностью $p = 3/4$ и -1 с вероятностью $1-p = 1/4$

17.1.

17.2.

17.3.

17.4.

17.5. Корреляционная матрица должна быть положительно определена. Получаем квадратное неравенство.

17.6.

17.7. По определению ковариационной матрицы:

$$\mathrm{Var}(\xi) = \begin{pmatrix} \mathrm{Var}(\xi_1) & \mathrm{Cov}(\xi_1,\xi_2) & \mathrm{Cov}(\xi_1,\xi_3) \\ \mathrm{Cov}(\xi_2,\xi_1) & \mathrm{Var}(\xi_2) & \mathrm{Cov}(\xi_2,\xi_3) \\ \mathrm{Cov}(\xi_3,\xi_1) & \mathrm{Cov}(\xi_3,\xi_2) & \mathrm{Var}(\xi_3) \end{pmatrix} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 4 \end{pmatrix}$$

$$\mathrm{Var}(\xi_1+\xi_2+\xi_3) = \mathrm{Var}\left(\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}\right) =$$

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \mathrm{Var}\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 9$$

18.1. Обозначим искомую вероятность быть в Неведении в момент t значком p_t.

$$p_{t+\Delta} = p_t(1 - \lambda\Delta - o(\Delta))$$

Отсюда получаем, что

$$\frac{p_{t+\Delta} - p_t}{\Delta} = -\lambda p_t + \frac{o(\Delta)}{\Delta}$$

Устремляем Δ к нулю и решаем получающееся дифференциальное уравнение с начальным условием $p_0 = 1$, так как изначально Ученик находится в Неведении.

Итого:

$$p_t = \exp(-\lambda t)$$

18.2. Уже для двух перескоков вероятность не превосходит $(\lambda\Delta + o(\Delta))^2 = o(\Delta)$. Даже если сложить вероятности от двух перескоков до бесконечности, мы получаем сумму равную $o(\Delta)$. Следоветельно, вероятность нуля перескоков равна $1 - \lambda\Delta - o(\Delta)$.

Обозначим вероятность нахождения на Шарике в момент времени t значком p_t. Отсюда:

$$p_{t+\Delta} = p_t(1 - \lambda\Delta - o(\Delta)) + (1 - p_t)(\lambda\Delta + o(\Delta)) + o(\Delta)$$

Получаем дифференциальное уравнение:

$$\dot{p}_t = \lambda - 2\lambda p_t$$

Для начального условия $p_0 = 1$ получаем решение $p_t = 0.5 + 0.5\exp(-2\lambda t)$.

18.3. В равновесии выполнено соотношение

$$p_{k+1} = \frac{\lambda_u}{\lambda_d} p_k$$

Следовательно, $p_k = \left(\frac{\lambda_u}{\lambda_d}\right)^k p_0$.

Исходя из условия $\sum_k p_k = 1$ находим p_0.

18.4. При $k \geqslant 1$ получаем уравнение

$$\mathbb{P}(L_{t+\Delta} = k) = \lambda_{in}\Delta\mathbb{P}(L_t = k-1) + \lambda_s\Delta\mathbb{P}(L_t = k+1) + (1 - \lambda_s\Delta - \lambda_{in}\Delta)\mathbb{P}(L_t = k) = o(\Delta)$$

При $k = 0$ уравнение особое:

$$\mathbb{P}(L_{t+\Delta} = 0) = (1 - \lambda_{in}\Delta)\mathbb{P}(L_t = 0) + \lambda_s\Delta\mathbb{P}(L_t = 1) + o(\Delta)$$

Приравниваем $\mathbb{P}(L_{t+\Delta} = k)$ и $\mathbb{P}(L_t = k)$, для краткости обозначим p_k. Устремляем Δ к нулю.

Получаем разностные уравнения:

$$\begin{cases} p_1 = \frac{\lambda_{in}}{\lambda_s} p_0 \\ p_{k+1} = \frac{\lambda_s + \lambda_{in}}{\lambda_s} p_k - \frac{\lambda_{in}}{\lambda_s} p_{k-1} \end{cases}$$

С условием нормировки $\sum p_k = 1$ решение системы единственно,

$$p_k = (1-a)a^k, \; a = \lambda_{in}/\lambda_s$$

То есть длина очереди имеет геометрическое распределение.

Поэтому средняя длина очереди равна $\mathbb{E}(L_t) = a/(1-a) = \frac{\lambda_{in}}{\lambda_s - \lambda_{in}}$.

Заметим, что если $\lambda_s \approx \lambda_{in}$, длина очереди будет просто огромной!

18.5. Например, для Y_4:

$$\mathbb{P}(Y_4 \in [t; t+\Delta]) = C_{10}^1 \cdot \Delta \cdot C_9^3 t^3 (1-t)^6 + o(\Delta)$$

Читаем вслух:

1. Одна из десяти величин должна попасть в отрезок $[t; t+\Delta]$;

2. Три из девяти оставшихся должны оказаться меньше t;

3. Шесть из девяти оставших должны оказаться больше t;

Вероятностью попадания двух и более величин в отрезок длины Δ пренебрегаем!

18.6.

19.1.

19.2.

19.3. Являются: $X_t = -W_t$, $X_t = W_{a+t} - W_a$, $X_t = \frac{1}{a} W_{a^2 t}$

19.4.

19.5.

20.1. Среднее равно $(x + 25)/5$. Если $x < 5$, получаем $x = 0$. Если $x \in (5; 7)$, получаем $x = 25/4$. Если $x > 7$, получаем $x = 10$.

20.2. Медиана не изменится, среднее упадёт на $10/25 = 0.4$. Для случая роста 153: среднее упадёт на 0.8, медиана упадёт произвольно на некое число из отрезка $[0; 2]$.

20.3. да

20.4. да

20.5.

20.6. два независимых симметричных распределения; практически любая сумма несимметричных распределений, например, два независимых с $p(x) = 2 - 2x$ на $[0; 1]$; неотрицательные случайные величины; симметричные около нуля случайные величины

20.7. Исключим те варианты, когда все пять наблюдений оказались или синхронно выше, или синхронно ниже медианы, получаем, $p = 1 - 2 \cdot 0.5^5 = 1 - 0.5^4$.

20.8. укреплять те места, где не было следов пуль

20.9.

20.10.

20.11.

1. $\mathbb{E}(\bar{X}_n) = \mu$, $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$
2. $\mathbb{E}(\bar{X}_n) = \mu$, $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$
3. При $N \to \infty$ получится формула для выборки с возвращениями.

20.12.

20.13.

20.14. $m^2 = 1/2$, $\mathrm{Med}(X) = 1/\sqrt{2}$.

21.1.

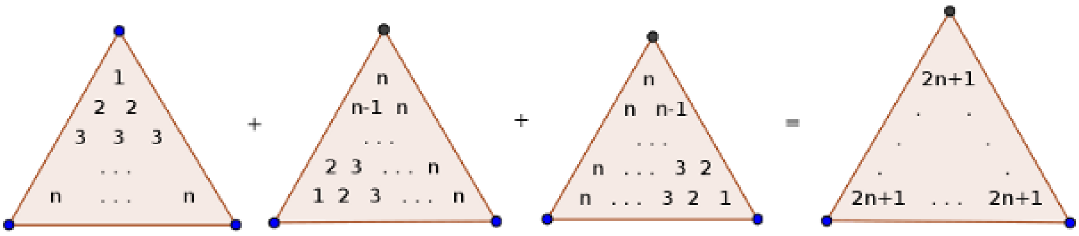
21.2.

21.3.

21.4.

21.5.

22.1. Здесь потребуется формула для $S = 1^2 + 2^2 + \ldots + n^2$.
На рисунке сложены три суммы:



На языке формул:

$$3S = (2n+1) \cdot (1 + 2 + \ldots + n) = (2n+1)n\frac{n+1}{2}.$$

22.2. Минимум, $a = 6, b = 5, c = 9$. Например, проекцией вектора $(6, 3, 7, 8, 9, 10, 11)$ на вектора вида $(a, b, b, c, c, c, 0)$.

22.3.

22.4.

1. χ^2_d
2. d
3. $2d$
4. χ^2_{a+b}

22.5. $Q = Z_1^2$, зная функцию плотности Z_1, $f(z_1) = \frac{1}{\sqrt{2\pi}} \exp(-z_1^2/2)$, находим функцию плотности Q;

22.6. $\hat{Z} = \langle Z, z\rangle \cdot v$; $\hat{Z}_i = \langle Z, z\rangle \cdot v_i$; $\operatorname{Var}(\langle Z, z\rangle) = 1$; $\operatorname{Cov}(\hat{Z}_i, \hat{Z}_j) = v_i v_j \operatorname{Var}(\langle Z, z\rangle)_{ij} = v_i v_j$;

22.7.

23.1. Метод максимального правдоподобия:

$$C_n^{Y_к} C_{n-Y_к}^{Y_щ} a^{Y_к} (2a)^{Y_щ} (1-3a)^{Y_б} \to \max_a$$

Решая задачу максимизации Кота Матроскина получаем

$$\hat{a} = \frac{Y_к + Y_щ}{3n}$$

Замечаем, что $Y_к + Y_щ \sim Bin(n, 3a)$. Отсюда $\mathbb{E}(\hat{a}_{КМ}) = a$, $\operatorname{Var}(\hat{a}_{КМ}) = \frac{a(1-3a)}{n}$. Оценка несмещённая и состоятельная.

С точки зрения Пса Шарика, неизвестными являются две вероятности, a и b. Он решает задачу максимизации по двум переменным. В результате получается вполне себе интуитивная оценка $\hat{a}_{ПШ} = Y_к/n$.

23.2. Метод правдоподобия: $\max_n \mathbb{P}(S = 80)$. Замечаем, что $S \sim Bin\left(100, p = \frac{100}{n}\right)$. Отсюда, $\hat{n}_{ML} = 125$.

Метод моментов. Рассмотрим Y_1, Y_2, ..., Y_n. Величина Y_i равна 1 если при втором отлове i-ый заяц оказался с бантом и 0 иначе.

Метод моментов: $\mathbb{E}(Y_i)|_{n=\hat{n}} = \bar{Y}$:

$$\frac{100}{\hat{n}_{MM}} = \bar{Y}$$

Отсюда

$$\hat{n}_{MM} = \frac{100}{\bar{Y}} = \frac{100^2}{S}$$

23.3.

23.4.

23.5.

24.1. $\hat{\theta}_{ML} = 0.25$, $\hat{\theta}_{MM} = 0.2$ $\hat{\theta}_{MM} = \frac{2{,}4-\bar{X}}{7}$

24.2. $\hat{a} = \ln(Y_1)$, $\hat{b} = \ln(Y_2) - \ln(Y_1)$

24.3.

24.4.

24.5.

24.6. $\hat{a}_{ml} = \sum X_i^2/2n$, $\hat{a}_{mm} = \bar{X}$.

24.7.

24.8.

24.9.

24.10.

24.11.

24.12.

24.13.

24.14.

24.15.

24.16.

24.17.

24.18.

24.19.

24.20.

24.21.

1. $\mathbb{E}(X_i) = a, \mathbb{E}(|X_i|) = 5a/4$
2. $\hat{a} = 11/30$
3. $\hat{a} = 26/75$
4. $\hat{a}_{GMM} = 108/325$
5. $\begin{pmatrix} 37 & -44 \\ -44 & 64 \end{pmatrix}$

24.22.

24.23. $\operatorname{plim}_{n\to\infty} \hat{\alpha}_{ML}(n) = 0$, не является состоятельной

24.24.

25.1.

1. $c = 1/n$, да
2. $c = 1/(n + \sigma^2/\mu^2)$, нет, так как μ и σ неизвестны
3. $\lambda = 0$ и $\lambda = \sigma^2/\mu^2$

25.2.

1. несмещённая, состоятельная, линейная, неэффективная
2. несмещённая, состоятельная, линейная, неэффективная
3. несмещённая, несостоятельная, линейная, неэффективная
4. смещённая, состоятельная, линейная
5. смещённая, состоятельная, нелинейная
6. несмещённая, состоятельная, линейная, неэффективная
7. несмещённая, состоятельная, линейная, эффективная
8. смещённая, несостоятельная, линейная
9. смещённая, несостоятельная, нелинейная
10. несмещённая, состоятельная, линейная, неэффективная

25.3.

1. $\hat{\theta}_{ML} = \max\{Y_1, Y_2, \ldots, Y_n\}$.
2. Все Y_i меньше θ, значит и $\hat{\theta}$ всегда меньше θ, значит смещённая.
3. $F_{\hat{\theta}}(t) = \mathbb{P}(\hat{\theta} \leqslant t) = \mathbb{P}(Y_1 \leqslant t, Y_2 \leqslant t, \ldots) = (\mathbb{P}(Y_1 \leqslant t))^n$, $\mathbb{P}(Y_1 \leqslant t) = t^5/\theta^5$, $f_{\hat{\theta}}(t) = dF_{\hat{\theta}}(t)/dt = \frac{5nt^{5n-1}}{\theta^{5n}}$.
4. $\mathbb{E}(\hat{\theta}) = \frac{5n}{5n+1}\theta$.
5. $\hat{\theta}_{unbiased} = \frac{5n+1}{5n}\hat{\theta}$.

25.4.

1. $\hat{p}$ несмещённая
2. $\sigma^2 = np(1-p)$.
3. $\mathbb{E}(\hat{\sigma}^2) = (n-1)p(1-p)$, смещённая, $\hat{\sigma}^2_{unbiased} = \frac{n}{n-1}\hat{\sigma}^2$.

25.5.

25.6.

25.7.

25.8.

25.9.

25.10.

25.11.

25.12. Закон распределения X также экспоненциальный, но с другим λ. Честно находим $\mathbb{E}(X) = \theta/20$, отсюда $\hat{\theta}_{unbiased} = 20X$.

25.13.

25.14. Обе оценки несмещённые, состоятельные. Более эффективна $\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2}$.

25.15.

25.16.

26.1. $\frac{1}{1.8} - \frac{1}{2.2}$, $[0; 10X]$.

26.2.

26.3.

26.4.

26.5.

26.6. Важна при обоих доверительных интервалах. Без предпосылки о нормальности интервал для дисперсии по данным формулам нельзя построить даже при больших n. При больших n можно отказаться от предпосылки о нормальности при построении интервала для μ.

26.7.

27.1. к 1/2

27.2. равномерно, $\alpha = 0.05$; нет, он резко увеличивает ошибку второго рода

27.3. $\alpha = 1/8$, $\beta = 9/32$

27.4. $\alpha = \mathbb{P}(\mathcal{N}(0;1) > -0.35) \approx 0.64$, $\beta = \mathbb{P}(\mathcal{N}(0;1) \leqslant -1.76) \approx 0.04$.

27.5.

27.6.

27.7.

27.8. Критерий Неймана-Пирсона сводится к сравнению $\bar{X}$ с порогом. При верной H_0 величина $\bar{X}$ распределена $\mathcal{N}(0; \frac{4}{n})$. Отсюда искомый критерий имеет вид: Если $\bar{X} \geqslant 0.825$, то гипотеза H_0 отвергается в пользу гипотезы H_a.

27.9. Упрощая неравенство из леммы Неймана-Пирсона, получаем критерий: если $X_1 \cdot X_2 > t$, то H_0 отвергается. Величину t находим из уравнения

$$\int_0^t (1 - t/x)\, dx = 0.05$$

27.10.

27.11.

28.1. При p_N заданном в H_0 критерий Пирсона имеет хи-квадрат распределение с двумя степенями свободы. При оцениваемом p_N критерий Пирсона имеет хи-квадрат распределение с одной степенью свободы.

https://ru.wikipedia.org/wiki/Закон_Харди_—_Вайнберга

29.1. $\mathbb{E}(Ay) = A\mathbb{E}(y)$, $\mathrm{Var}(Ay) = A\,\mathrm{Var}(y)A^T$

29.2.

29.3.

29.4.

29.5.

29.6.

30.1. Вычтем $2x^2 + 2y^2 + 2z^2 + 2w^2$. Получим, что оптимальное $x = 0$. Далее, $z = 0$, $y = 0$. В итоге $w = 1$ или $w = -1$.

30.2.

30.3.

30.4.

30.5. $(5 + 4 + 1) = 10$; длины — $\sqrt{5} \cdot \sqrt{99}$, $\sqrt{4} \cdot \sqrt{99}$, $\sqrt{1} \cdot \sqrt{99}$; выборочные дисперсии — 5, 4, 1; $(5 + 4)/10 = 0.9$.

30.6.

31.1.

31.2.

31.3.

32.1. $C_{20}^2 \cdot 18$.

32.2. $1 + x + x^2 + x^3 + x^4 = (1 - x^5)/(1 - x)$ $1 + x + x^2 + x^3 + \ldots = 1/(1 - x)$ $(1 + x)^4$

32.3.

32.4.

32.5. 0 с вероятностью 1/4 и 1 с вероятностью 3/4

32.6. да, сможет!

32.7. $F_n = F_{n-1} + F_{n-2}$. Запишем разложения для $g(x)$, $xg(x)$ и $x^2g(x)$ друг под другом. Вычитаем. Получаем, что $g(x) = x/(1 - x - x^2)$.

Указанная дробь — это и есть производящая функция при маленьком x.

Производящая функция представима в виде суммы:

$$g(x) = \frac{1}{\sqrt{5}}\left(\frac{a}{x+a} - \frac{b}{x+b}\right),$$

где $a = (1 - \sqrt{5})/2$, $b = (1 + \sqrt{5})/2$.

32.8. Замечаем, что $h_2(t) = h_1(t) \cdot h_1(t)$. После первого шага:

$$h_1(t) = 0.7t + 0.3th_1^2(t)$$

32.9.

33.1.

33.2.

34.1. Если величина X равновероятно принимает k значений, то спутанность равна k. У равномерной на $[0; a]$ спутанность равна a.

34.2.

34.3.

34.4.

34.5.

37. Источники мудрости

[RB47] Richard Ruggles и Henry Brodie. «An empirical approach to economic intelligence in World War II». В: *Journal of the American Statistical Association* 42.237 (1947), с. 72—91. Подробности про то, как захватив всего два танка, можно оценить ежемесячный выпуск с точностью в пару процентов.

[Wil13] Herbert S Wilf. *generatingfunctionology*. Elsevier, 2013. URL: https://www.math.upenn.edu/~wilf/DownldGF.html. Шикарная книжка про производящие функции.

[Muk17] Somabha Mukherjee. «A Proof of the Herschel-Maxwell Theorem Using the Strong Law of Large Numbers». В: *arXiv preprint arXiv:1701.02228* (2017).