

# 1 Семинар 1. Вперёд, в рукопашную!

Минитеория:

1. <https://github.com/bdemeshev/pr201/wiki> или <http://pokrovka11.wordpress.com>
2. Константы. Строчные английские буквы,  $a, x, z$ .
3. События. Заглавные английские буквы начала алфавита  $A, B, C, D$ . Вероятность  $\mathbb{P}(A)$ .
4. Случайные величины. Заглавные английские буквы конца алфавита  $X, Y, W, Z$ . Математическое ожидание  $\mathbb{E}(X)$ .

Задачи:

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  $\mathbb{P}(X = 2)$ ,  $\mathbb{P}(X > 1)$ ,  $\mathbb{E}(X)$
2. Неправильную монетку с вероятностью «орла» равной  $p$  подбрасывают до первого «орла». Чему равно среднее количество подбрасываний? Орлов? Решек? Какова вероятность того, что будет чётное число бросков?
3. Саша и Маша по очереди подбрасывают кубик. Посуду будет мыть тот, кто первым выбросит шестерку. Маша бросает первой. Каковы ее шансы отдохнуть за «Cosmo»?
4. 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.
  - (a) Какова вероятность того, что игра рано или поздно закончится выпадением 6-ки?
  - (b) Какова ожидаемая продолжительность игры?
  - (c) Чему равен ожидаемый выигрыш в эту игру?
  - (d) Чему равен ожидаемый выигрыш в эту игру, если изначально на кону лежит 100 рублей?
  - (e) Изменим условие: если выпадает 5, то сумма на кону сгорает, а игра продолжается. Как изменятся ответы на предыдущие вопросы?
5. 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.
  - (a) У кого какие шансы выиграть?
  - (b)  $\mathbb{P}(X = 4)$ ,  $\mathbb{P}(Y = 1)$ ,  $\mathbb{E}(X)$ ,  $\mathbb{E}(Y)$
  - (c) Решите аналогичную задачу для ОРО и ООР.

6. «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  $\mathbb{P}(X = 2)$ ,  $\mathbb{P}(X = 3)$ ,  $\mathbb{P}(X = \infty)$
7. 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.
  - (а) Как выглядит оптимальная стратегия? Чему равен ожидаемый выигрыш при использовании оптимальной стратегии?
  - (б) Какова средняя продолжительность игры при использовании оптимальной стратегии?
  - (с) Как выглядит оптимальная стратегия и чему равен ожидаемый выигрыш, если за каждое подбрасывание Вася платит 35 копеек?
8. 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?
9. 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  $\mathbb{P}(X = 4)$ ,  $\mathbb{E}(X)$
10. В каждой вершине треугольника по ёжику. Каждую минуту с вероятностью  $0.7$  каждый ежик независимо от других движется по часовой стрелке, с вероятностью  $0.3$  — против часовой стрелки. Обозначим  $T$  — время до встречи всех ежей в одной вершине. Найдите  $\mathbb{P}(T = 1)$ ,  $\mathbb{P}(T = 2)$ ,  $\mathbb{P}(T = 3)$ ,  $\mathbb{E}(T)$ .

## 2 Семинар 2. Хочу ещё задач!

1. Две команды равной силы играют в волейбол до трех побед одной из них, не обязательно подряд. Ничья невозможна. Из-за равенства сил можно считать, что вероятность победы каждой равна  $0.5$ . Величина  $N$  — количество сыгранных партий. Составьте табличку возможных значений  $N$  с их вероятностями. Найдите  $\mathbb{P}(N — \text{четное})$ ,  $\mathbb{E}(N)$
2. Какова вероятность того, что у 10 человек не будет ни одного совпадения дней рождений?
3. Наугад из четырех тузов разных мастей выбираются два.  $\mathbb{P}$ (они будут разного цвета)?
4. События  $A$  и  $B$  несовместны, т.е. не могут произойти одновременно. Известно, вероятности  $\mathbb{P}(A) = 0,3$ ,  $\mathbb{P}(B) = 0,4$ . Найдите  $\mathbb{P}(A^c \cap B^c)$
5. Вероятность  $\mathbb{P}(A) = 0,3$ ,  $\mathbb{P}(B) = 0,8$ . В каких пределах может лежать  $\mathbb{P}(A \cap B)$ ?
6. Множество исходов  $\Omega = \{a, b, c\}$ ,  $\mathbb{P}(\{a, b\}) = 0,8$ ,  $\mathbb{P}(\{b, c\}) = 0,7$ . Найдите  $\mathbb{P}(\{a\})$ ,  $\mathbb{P}(\{b\})$ ,  $\mathbb{P}(\{c\})$
7. 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?

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.
- (a) Как выглядит оптимальная стратегия в этой игре?
- (b) Чему при этом будет равен средний выигрыш?
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 (т.е. может зайти в одну и ту же). Сейчас мышка в коридоре. Сколько времени ей в среднем потребуется, чтобы найти сыр?
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 Змей 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?
11. У Пети — монетка, выпадающая орлом с вероятностью  $p \in (0; 1)$ . У Васи — с вероятностью  $q \in (0; 1)$ . 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?
12. Треугольник с вершинами  $(0; 0)$ ,  $(2; 0)$  и  $(1; 1)$ . Внутри него случайным образом выбирается точка,  $X$  — абсцисса точки. Найдите  $\mathbb{P}(X > 1)$ ,  $\mathbb{P}(X \in [0.5; 1])$ ,  $\mathbb{E}(X)$
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?
14. Исследовательница Мишель подкидывает игральный кубик неограниченное количество раз и складывает выпадающие количества очков.
- (a) Чему примерно равна вероятность того, что однажды сумма в точности будет равна 123456789?
- (b) Чему точно равна указанная вероятность?

### 3 Семинар 3. К чёрту условности!

1. Имеется три монетки. Две «правильных» и одна — с «орлами» по обеим сторонам. Петя выбирает одну монетку наугад и подкидывает ее два раза. Оба раза выпадает «орел». Какова условная вероятность того, что монетка «неправильная»?
2. Два охотника одновременно выстрелили в одну утку. Первый попадает с вероятностью 0,4, второй — с вероятностью 0,7. В утку попала ровно одна пуля. Какова условная вероятность того, что утка была убита первым охотником?
3. Кубик подбрасывается два раза. Найдите вероятность получить сумму равную 8, если при первом броске выпало 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?
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?
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?
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?
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?
9. Ген карих глаз доминирует ген синих. Т.е. 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?
10. Из колоды в 52 карты извлекается одна карта наугад. Являются ли события «извлечен туз» и «извлечена пика» независимыми?
11. Из колоды в 52 карты извлекаются по очереди две карты наугад. Являются ли события «первая карта — туз» и «вторая карта — туз» независимыми?
12. Известно, что  $P(A) = 0,3$ ,  $P(B) = 0,4$ ,  $P(C) = 0,5$ . События  $A$  и  $B$  несовместны, события  $A$  и  $C$  независимы и  $P(B|C) = 0,1$ . Найдите  $P(A \cup B \cup C)$ .
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 тети Маши есть дети обоих полов.
  - (а) Известно, что хотя бы один ребенок — мальчик.

- (b) Тетя Маша наугад выбирает одного своего ребенка и посылает к тете Оле, вернуть учебник по теории вероятностей. Это оказывается мальчик.
- (c) Известно, что старший ребенок — мальчик.
- (d) На вопрос: «А правда ли тетя Маша, что у вас есть сын, родившийся в пятницу?» тетя Маша ответила: «Да».

## 4 Семинар 4. Use R

## 5 Семинар 5. Use R or die!

1. Самая простая. Случайная величина  $N$  имеет пуассоновское распределение с  $\lambda = 2$ . С помощью симуляций оцените  $\mathbb{E}(N^3)$ ,  $\mathbb{P}(N \geq 4)$ ,  $\mathbb{P}(N \geq 10 \mid N \geq 5)$ ,  $\mathbb{E}(N \mid N \geq 5)$ . Функция `rpois` может помочь :)
2. Случайные величины  $X_1, \dots, X_5$  имеют равномерное распределение на отрезке  $[0; 1]$  и независимы. С помощью симуляций оцените  $\mathbb{P}(\min\{X_1, \dots, X_5\} > 0.2)$ ,  $\mathbb{P}(\min\{X_1, \dots, X_5\} > 0.2 \mid X_1 + X_2 < 0.5)$ ,  $\mathbb{E}(\min\{X_1, \dots, X_5\})$ ,  $\mathbb{E}(\min\{X_1, \dots, X_5\} \mid X_1 + X_2 < 0.5)$
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. В колоде 52 карты. Мы вытаскиваем карты из колоды до первого туза, пусть  $X$  — количество вытянутых карт. С помощью симуляций оцените  $\mathbb{E}(X^2)$ ,  $\mathbb{P}(X > 10)$ ,  $\mathbb{P}(X > 5 \mid X < 15)$ ,  $\mathbb{E}(X^2 \mid X < 15)$
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  $\mathbb{P}(Y > 3X)$ ,  $\mathbb{E}(XY)$ ,  $\mathbb{P}(Y > 3X \mid Y < X^2)$ ,  $\mathbb{E}(XY \mid Y < X^2)$

## 6 Семинар 6. Разлагай и властвуй!

1. Из грота ведут 10 штретов, с длинами 100м, 200м, ... 1000м. Самый длинный штрет оканчивается выходом на поверхность. Остальные — тупиком. Вася выбирает штреты наугад, в тупиковый штрет два раза не ходит. Какова вероятность того, что Вася посетит самый короткий штрет? Какой в среднем путь он нагуляет прежде чем выберется на поверхность?
2. У Маши 30 разных пар туфель. И она говорит, что мало! Пес 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  $\mathbb{E}(X)$ ,  $\mathbb{E}(Y)$ ,  $\text{Var}(X)$ ,  $\text{Var}(Y)$ .
3. У меня в кармане 3 рубля мелочью. Среди монет всего одна монета достоинством 50 копеек. Я извлекаю монеты по одной наугад до извлечения 50 копеечной монеты. Какую сумму в среднем я извлеку?
4. «Модница». В шкатулке у Маши 100 пар сережек. Каждый день утром она выбирает одну пару наугад, носит ее, а вечером возвращает в шкатулку. Проходит год.
  - (a) Сколько в среднем пар окажутся ни разу не надетыми?
  - (b) Сколько в среднем пар окажутся одетыми не менее двух раз?
  - (c\*) Как изменятся ответы, если каждый день Маша покупает себе новую пару сережек и вечером добавляет ее в шкатулку?
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?
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  $\mathbb{E}(X)$ ,  $\text{Var}(X)$ .
7. Над озером взлетело 20 уток. Каждый из 10 охотников стреляет в случайно выбираемую им утку. Величина  $Y$  — количество убитых уток,  $X$  — количество попавших в цель охотников. Найдите  $\mathbb{E}(X)$ ,  $\text{Var}(X)$ ,  $\mathbb{E}(Y)$ ,  $\text{Var}(Y)$ , если охотники стреляют без промаха. Как изменится ответ, если вероятность попадания равна 0.7?
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  $\mathbb{E}(X)$ ,  $\text{Var}(X)$ . Вероятность совпадения роста будем считать равной нулю.
9. По 10 коробкам наугад раскладывают 7 карандашей. Каково среднее количество пустых коробок? Дисперсия?
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  $\mathbb{E}(X)$ ,  $\text{Var}(X)$ ? Как это объяснить ребёнку?

11. Из колоды в 52 карты извлекается 5 карт. Сколько в среднем извлекается мастей? Достоинств? Тузов? Дисперсии этих величин?
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  $\text{Var}(X)$
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  $\text{Var}(X)$
14. Кубик подбрасывается  $n$  раз. Величина  $X_1$  — число выпадений 1, а  $X_6$  — число выпадений 6. Найдите  $\text{Corr}(X_1, X_6)$

Пуассоновский поток событий.

Если:

1.  $X_t$  — количество происшествий за период  $[0; t]$
2. Независимость приращений. Количество происшествий на непересекающихся интервалах времени независимы.
3. Стационарность приращений. Распределение количества событий на интервале  $t \in [t_0; t_0 + \Delta t]$  зависит только от  $\Delta t$ , но не от  $t_0$ .
4. На малом интервале времени вероятность одного происшествия примерно пропорциональна длине интервала,  $\mathbb{P}(X_t = 1) = \lambda t + o(t)$ .
5. На малом интервале времени несколько происшествий происходят существенно реже одного происшествия,  $\mathbb{P}(X_t \geq 2) = o(t)$ .

То:

1. Количество событий за единицу времени,  $X$ , имеет пуассоновское распределение  $X \sim Pois(\lambda)$

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\mathbb{E}(X) = \lambda$$

$$\text{Var}(X) = \lambda$$

2. Отсюда смысл  $\lambda$  — 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  $X_t \sim Pois(\lambda t)$

$$\mathbb{P}(X_t = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$

$$\mathbb{E}(X_t) = \lambda t$$

$$\text{Var}(X_t) = \lambda t$$

4. Время между  $(i - 1)$ -ым и  $i$ -ым происшествием,  $Y_i$ , имеет экспоненциальное распределение  $Y_i \sim exp(\lambda)$ .

$$f(y) = \begin{cases} \lambda e^{-\lambda y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

$$\mathbb{E}(Y_i) = 1/\lambda$$

$$\text{Var}(Y_i) = 1/\lambda^2$$

5. Величины  $Y_i$  независимы
6. Сумма двух независимых пуассоновских процессов с интенсивностями  $\lambda_1$  и  $\lambda_2$  — пуассоновский процесс с интенсивностью  $\lambda_1 + \lambda_2$

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



## 7 Семинар ( $n+1$ ). За время моего дежурства происшествий не было!

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?
  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?
  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?
  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?
  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.
    - (а) Какова вероятность того, что тетя Зина обслужит клиента быстрее?
    - (б) Как распределено время обслуживания того клиента, который освободится быстрее?
    - (с) Каково условное среднее время обслуживания клиента тетей Зиной, если известно, что она обслужила клиента быстрее тети Маши?
  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).
    - (а) Какова вероятность того, что в очереди ровно  $n$  студентов?
    - (б) Какова средняя длина очереди?
- Подсказка: если сейчас в очереди  $n$  человек, то через малый промежуток времени  $dt \dots$
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. 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?

9. Величина  $X$  распределена экспоненциально с параметром  $\lambda$ , а константа  $a > 0$ . Как распределена величина  $Y = aX$ ?
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.
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.
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.
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.
  - (a) Найдите функцию плотности  $Y_1$
  - (b) Верно ли, что процесс количества звонков, которые услышит Василиса, имеет независимые приращения?
14. Владелец таверны «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  $E(X)$ ,  $Var(X)$ .

## 8 Многомерное нормальное

1. Ермолай Лопехин решил приступить к вырубке вишневого сада. Однако выяснилось, что растут в нём не только вишни, но и яблони. Причём, по словам Любви 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:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 25 \\ 125 \end{pmatrix}; \begin{pmatrix} 5 & 4 \\ 4 & 10 \end{pmatrix} \right) \quad (1)$$

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

Автор: Кирилл Фурманов, Ира Чернухина

2. 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»?

3. 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:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[ \begin{pmatrix} 68.1 \\ 68.3 \end{pmatrix}; \begin{pmatrix} 6.3 & 2.1 \\ 2.1 & 3.2 \end{pmatrix} \right]$$

- (a) 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.
- (b) Рассчитайте корреляцию между  $X_1$  и  $X_2$
- (c) 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  $X' = (X'_1, X'_2)$ .
- (d) Определите, каков ожидаемый рост и дисперсия роста человека, средний рост родителей которого составляет 72 дюйма?
- (e) Найдите вероятность того, что рост человека превысит 68 дюймов, если средний рост его родителей равен 72 дюймам. Подсказка: используйте предыдущий пункт и нормальность распределения!

## 9 Статистика ноль

1. Имеется пять чисел:  $x$ , 4, 5, 7, 9. При каком значении  $x$  медиана будет равна среднему?
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?
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?
4. Деканат утверждает, что если студента N перевести из группы А в группу В, то средний рейтинг каждой группы возрастет. Возможно ли это?
5. Есть три группы по 10 человек, две группы по 20 человек и одна группа по 40 человек. У каждой из групп свой преподаватель.
  - (a) Каков средний размер группы, для которой читает лекции наугад выбранный профессор?
  - (b) Каков средний размер группы, в которой учится наугад выбранный студент?

(с) 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?

6. Приведите примеры, когда  $\text{Med}(X + Y) = \text{Med}(X) + \text{Med}(Y)$  и  $\text{Med}(X + Y) \neq \text{Med}(X) + \text{Med}(Y)$ , где  $\text{Med}$  — медиана

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.

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 статистiku Абрахаму Вальду с вопросом, в каких местах следует увеличить броню самолёта.

Что посоветовал Абрахам Вальд и почему?

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

Возможно ли это?

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

(а) Мы выбираем  $n$  чисел из  $N$  с возвращениями

(б) Мы выбираем  $n$  чисел из  $N$  без возвращений

(с) Во что превращаются полученные формулы при  $n = 1$ ? при  $n = N$ ? при  $N \rightarrow \infty$ ?

## 10 Максимально правдоподобно!

### Минитеория

Метод моментов (ММ, method of moments): найти  $\theta$  из уравнения  $\bar{X}_n = \mathbb{E}(X_i)$

Метод максимального правдоподобия (МЛ, maximum likelihood):

найти  $\theta$  при котором вероятность получить имеющиеся наблюдения будет максимальной

Наблюдаемая информация Фишера:  $\hat{I} = -\frac{\partial^2 l}{\partial \theta^2}(\hat{\theta})$ ,  $\widehat{\text{Var}}(\hat{\theta}_{ML}) = \hat{I}^{-1}$ .

Байесовский подход (bayesian approach):

1. Сделать изначальное предположение о распределении  $\hat{\theta}$

2. Обновлять закон распределения  $\hat{\theta}$  по формуле условной вероятности

Пусть  $l(\theta)$  - логарифмическая функция правдоподобия ( $l(\theta) = \ln(f(X_1, \dots, X_n, \theta))$ ).

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  $\theta$  содержится в выборке  $X_1, \dots, X_n$

Неравенство Крамера-Рао (Cramer-Rao) («слишком хорошей оценки не бывает»):

Если  $\hat{\theta}$  - несмещенная оценка и  $\dots$ , то  $\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$

Оценки ML - самые лучшие (асимптотически несмещенные и с минимальной дисперсией):  
Если  $X_i$  - iid, ..., и  $n \rightarrow \infty$  то  $\hat{\theta}_{ML} \sim N(\theta, \frac{1}{I(\theta)})$ .

1. Допустим, что  $X_i$  - независимы и имеют закон распределения, заданный табличкой:

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

Имеется выборка:  $X_1 = 0, X_2 = 2$ .

(a) Найдите оценки  $\hat{\theta}_{ML}$  и  $\hat{\theta}_{MM}$

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

(c) Постройте ML и MM оценки для произвольной выборки  $X_1, X_2, \dots, X_n$

2. У Васи есть два одинаковых золотых слитка неизвестной массы  $m$  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

(a)  $u_i$  — независимы и  $N(0; b)$

(b)  $u_i$  — независимы и  $U[-b; b]$

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

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

(a)  $\theta t^{\theta-1}$  при  $t \in [0; 1]$ ;

(b)  $\frac{2t}{\theta^2}$  при  $t \in [0; \theta]$

(c)  $\frac{\theta e^{-\frac{\theta^2}{2t}}}{\sqrt{2\pi t^3}}$  при  $t \in [0; +\infty)$ ;

(d)  $\frac{\theta(\ln^{\theta-1} t)}{t}$  при  $t \in [1; e]$ ;

(e)  $\frac{e^{-|t|}}{2(1-e^{-\theta})}$  при  $t \in [-\theta; \theta]$

5. «Про зайцев». В темно-синем лесу, где трепещут осины, живут  $n$  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  $Z$  зайцев окажутся с бантами.

С помощью величины  $Z$  постройте MM и ML оценку для неизвестного параметра  $n$ .

6. Пусть  $X_1, \dots, X_n$  - независимы и экспоненциальны с параметром  $\lambda$ . Постройте MM и ML оценки параметра  $\lambda$ . Оцените дисперсию ML оценки.

7. Пусть  $X_1, \dots, X_n$  - независимы и  $N(\mu; \sigma^2)$ . Значение  $\sigma^2$  известно. Постройте ММ и МЛ оценки параметра  $\mu$ .
8. Пусть  $X_i$  независимы и одинаково распределены  $N(\alpha, 2\alpha)$   
По выборке  $X_1, \dots, X_n$  постройте оценку для  $\alpha$  с помощью МЛ и ММ. Оцените дисперсию МЛ оценки.
9. Пусть  $Y_1 \sim N(0; \frac{1}{1-\theta^2})$ . Найдите МЛ оценку для  $\theta$ . Оцените дисперсию МЛ оценки.
10. Пусть  $X_1, X_2, \dots, X_n$  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 МЛ и ММ. Оцените дисперсию МЛ оценки.
11. Пусть  $X_1, X_2, \dots, X_n$  независимы и равномерно распределены на отрезке  $[0; \theta]$ ,  $\theta > 1$
- Постройте ММ и МЛ оценки для неизвестного  $\theta$ .
  - Как изменятся ответы на «а», если исследователь не знает значений самих  $X_i$ , а знает только количество  $X_i$  оказавшихся больше единицы?
12. В озере водятся караси, окуни, щуки и налимы. Вероятности их поймать занесены в табличку
- | Fish           | 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.
- Постройте  $\hat{p}_{ML}$
  - Найдите ожидаемую и наблюдаемую информацию Фишера
  - Несмещенная оценка  $\hat{\theta}$  получена по 100 наблюдениям:  $X_1, \dots, X_{100}$ . В каких пределах может лежать  $\text{Var}(\hat{\theta})$ ?
13. Известно, что  $X_i$  — независимы и имеют закон распределения, заданный таблицей:
- | $X_i$          | 0   | 1       |
|----------------|-----|---------|
| $\mathbb{P}()$ | $p$ | $1 - p$ |
- Постройте  $\hat{p}_{ML}$
  - Найдите ожидаемую и наблюдаемую информацию Фишера. Постройте возможные графики  $I(p)$ .
  - Пусть  $\hat{\theta}$  — несмещенная оценка, полученная по 100 наблюдениям:  $X_1, \dots, X_{100}$ . В каких пределах может лежать  $\text{Var}(\hat{\theta})$ ?
14. Пусть  $X_i$  независимы и имеют экспоненциальное распределение с параметром  $\lambda$ , т.е.  $p(t) = \lambda e^{-\lambda t}$ .
- Найдите  $I(\lambda)$ , если наблюдаются  $X_1, \dots, X_n$
  - Пусть  $\lambda = 1/\theta$ , т.е.  $p(t) = \frac{1}{\theta} e^{-\frac{1}{\theta} t}$ . Найдите  $I(\theta)$ , если наблюдается  $X_1, \dots, X_n$

15. Пусть  $X_i$  — независимы и одинаково распределены. Пусть  $I_{X_i}(\theta)$  — информация Фишера о  $\theta$ , получаемая при наблюдении  $X_i$ .
- Верно ли, что  $I_{X_1}(\theta) = I_{X_2}(\theta)$ ?
  - Как найти  $I_{X_1, \dots, X_n}(\theta)$  зная  $I_{X_i}(\theta)$ ?
16. 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  $[0; kX]$ .
17. Пусть  $X$  — экспоненциальна с параметром  $\lambda$  и  $\mu = \mathbb{E}(X)$ . С какой вероятностью интервал  $[0.9X; 1.1X]$  накрывает  $\mu$ ? Постройте 90%-ый доверительный интервал для  $\mu$  вида  $[0; kX]$ .
18. Пусть  $X_i$  — независимы и нормальны  $N(\mu, 1)$ . Какова вероятность того, что интервал  $[\bar{X}_{10} - 1; \bar{X}_{10} + 1]$  накроет неизвестное  $\mu$ ? Постройте 90%-ый доверительный интервал для  $\mu$  вида  $[\bar{X}_{10} - k; \bar{X}_{10} + k]$ .
19. Величины  $X_1, \dots, X_n$  — 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  $\sum \ln(\ln(X_i)) = -30$
- Найдите ML оценку параметра  $\theta$  и ожидаемую и наблюдаемую информацию Фишера
  - Постройте 95% доверительный интервал для  $\theta$
20. Величины  $X_1, \dots, X_n$  — 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  $\sum 1/X_i = 12$
- Найдите ML оценку параметра  $\theta$  и информацию Фишера  $I(\theta)$
  - Пользуясь данными по выборке постройте оценку  $\hat{I}$  для информации Фишера
  - Постройте 90% доверительный интервал для  $\theta$  Hint:  $\mathbb{E}(1/X_i) = 1/\theta^2$  (интеграл берется заменой  $x = \theta^2 a^{-2}$ )
21. 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.
- С помощью метода максимального правдоподобия определите, простая она или трудная.
  - С помощью байесовского подхода посчитайте вероятности того, что задача была простая, если на экзамене было 7 легких и 3 трудных задачи.
22. Известно, что  $X_i$  независимы,  $\mathbb{E}(X_i) = 5$ ,  $\text{Var}(X_i) = 4$ . Как примерно распределены следующие величины:
- $\bar{X}_n$ ,
  - $Y_n = (\bar{X}_n + 3)/(\bar{X}_n + 6)$ ,
  - $Z_n = \bar{X}_n^2$ ,
  - $W_n = 1/\bar{X}_n$
23. Известно, что  $X_i$  независимы и равномерны на  $[0; 1]$ .

- (a) Найдите  $\mathbb{E}(\ln(X_i))$ ,  $\text{Var}(\ln(X_i))$ ,  $\mathbb{E}(X_i^2)$ ,  $\text{Var}(X_i^2)$
- (b) Как примерно распределены величины  $X_n = \frac{\sum \ln(X_i)}{n}$ ,  $Y_n = (X_1 \cdot X_2 \cdots X_n)^{1/n}$ ,  $Z_n = \left(\frac{\sum X_i^2}{n}\right)^3$
24. Величины  $X_i$  независимы и имеют функцию плотности  $f(x) = a \cdot x^{a-1}$  на отрезке  $[0; 1]$ .
- (a) Постройте оценку  $\hat{a}$  методом моментов, укажите ее примерный закон распределения
- (b) По 100 наблюдениям оказалось, что  $\sum X_i = 25$ . Посчитайте численное значение  $\hat{a}$  и оцените дисперсию случайной величины  $\hat{a}$ .
25. Начинаящий 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  $X_1, X_2, \dots, X_n$  — количества ударов которые потребовались Васе в соответствующий день.
- (a) Найдите оценку  $p$  методом максимального правдоподобия
- (b) Найдите достаточную статистику  $T$
- (c) Выразите  $\widehat{\text{Var}}(\hat{p})$  через достаточную статистику  $T$
- (d) Найдите  $\mathbb{P}(X_1 = 1 \mid T = t)$ .
26. 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  $X_1, \dots, X_n$ . Помогите псу Шарику:
- (a) Найдите оценку  $a$  методом максимального правдоподобия
- (b) Найдите достаточную статистику  $T$
- (c) Выразите  $\widehat{\text{Var}}(\hat{a})$  через достаточную статистику  $T$
- (d) Найдите  $\mathbb{P}(X_1 < 10 \mid T = t)$ .