

Problem set 14. Отличники лишним обозначением $[t]$ не заморачиваются!

14.1. В лесу живет N удавов. Каждый из них имеет свою длину. Обозначим σ^2 дисперсию длины наугад выбранного удава. Сегодня n удавов выползли погреться на солнышке на большой поляне. Обозначим \bar{X}_n среднюю длину выползших удавов. Чему равно $Cov(X_1, \sum_{i=1}^N X_i)$? Выразите $Cov(X_i, X_j)$ через σ^2 для $i \neq j$. Выразите $Var(\bar{X}_n)$ через σ^2 .

14.2. Маша и Саша same back! Маша наблюдает реализацию двух независимых случайных величин X и Y , распределенных равномерно на $[0; 1]$. Она выбирает, значение какой из них рассказать Саше. Саша выигрывает, если угадает, какая из величин наибольшая (та, значение которой он узнал от Маши, или другая). Маша выигрывает, если Саша ошибется. Найдите оптимальные стратегии.

Решение (Winkler)

Выбирая свой ответ наугад, Саша может гарантировать себе вероятность выигрыша 50%. Передавая Саше ту величину, значение которой легло от $\frac{1}{2}$ далее, Маша лишает Сашу релевантной для отгадывания информации.

14.3. Перед Машей колода в 52 карты. Маша открывает карты одну за одной. В любой момент времени Маша может сделать заявление "Следующая карта будет красной!". Маша выигрывает, если ее предсказание сбудется. Найдите оптимальную стратегию.

Решение (Winkler)

Маше безразлично делать ли прогноз на следующую карту или на последнюю, т.к. о них в любой момент времени имеется одинаковая информация. Исходная вероятность того, что последняя карта будет красной равна $\frac{1}{2}$. Следовательно, любая Машина стратегия приводит к такой вероятности выигрыша. Маша не может ее ни увеличить, ни снизить!

14.4. Перед Машей колода в 52 карты. Маша открывает карты одну за одной. Изначально у Маши 1 доллар. Маша может делать любую ставку в пределах имеющейся у нее суммы на цвет следующей карты. Найдите оптимальную стратегию и ожидаемый выигрыш, который приносит эта стратегия. Предположим бесконечную делимость денег.

14.5. В мешке лежат бочонки. На каждом из них написана цифра. На одном бочонке написана цифра 1, на двух бочонках - цифра 2, ..., на девяти бочонках - цифра 9. Маша вытаскивает один бочонок наугад. Саша не знает, какой бочонок достала Маша. Саша может задавать Маше вопросы, на которые можно отвечать только "да" или "нет". Как выглядит стратегия, минимизирующая ожидаемое число вопросов, необходимых чтобы угадать цифру?

14.6. Маша и Саша играют в быстрые шахматы. У них одинаковый класс игры и оба предпочитают играть белыми, т.е. вероятность выигрыша того, кто играет белыми равна $p > 0,5$. Партии играют до 10 побед. Первую партию Маша играет белыми. Она считает, что в каждой последующей партии белыми должен играть тот, кто выиграл предыдущую партию. Саша считает, что ходить белыми нужно по очереди. При каком варианте правил у Маши больше шансы выиграть?

Решение (C.L. Anderson)

При любом варианте правил, Маша будет ходить белыми не больше 10 раз, а Саша - не больше 9 раз. Победитель гарантированно определяется за 19 партий. Если победитель определен раньше, то дополним турнир недостающими фиктивными партиями (чтобы Саша ходил белыми ровно 9 раз, а Маша - ровно 10 раз). В турнире из 19 партий победителем при любом варианте правил оказывается тот, кто выиграл больше партий. Следовательно, Машины шансы не зависят от выбираемого варианта правил.

14.7. Из хорошо перетасованной колоды в 52 карты, содержащей четыре туза, извлекаются сверху карты до появления первого туза. На каком месте в среднем появляется первый туз?

Решение (Саха Серова)

Сначала сделаем колоду из четырех тузов, лежащих стопкой в случайном порядке. Затем карты будем по одной класть на случайное место в формируемой колоде. Всего 5 мест. По индукции видно, что вероятность для каждой карты попасть на место впереди первого туза равна $\frac{1}{5}$. Если X - количество карт, попавших раньше первого туза, то $X = X_1 + \dots + X_{48}$ и $E(X) = 48 \cdot \frac{1}{5} = \frac{48}{5}$.

14.8. Игра состоит из последовательности партий, в каждой из которых вы или ваш партнер выигрывает очко, вы - с вероятностью $p < 0,5$, он - с вероятностью $1 - p$. Число игр должно быть четным. Для выигрыша надо набрать больше половины очков. Предположим, что вам известно, что $p = 0,45$, и в случае выигрыша вы получаете приз. Если число партий в игре выбирается заранее, то каков будет ваш выбор?

14.9. Два кубика подбрасываются неограниченное число раз. Какова вероятность того, что сумма очков равная пяти, появится раньше суммы очков, равной семи?

Решение

Применяя метод первого шага получаем уравнение, $p = \frac{4}{36} \cdot 1 + \frac{6}{36} \cdot 0 + \frac{26}{36} \cdot p$.

Problem set 15. HSE. Probability theory and MS. 2 year. 17.01.06.

15.1. [т] Пусть наблюдаются $Y_1 = \beta + u_1$ и $Y_2 = 2 \cdot \beta + u_2$, где ненаблюдаемые u_i независимо и одинаково распределены, причем $E(u_i) = 0$ и $Var(u_i) = \gamma$. Найдите оценки методом максимального правдоподобия для β и γ , если: а) u_i - равномерно распределены; б) u_i - нормально распределены;

15.2. Имеются результаты экзамена в двух группах. Группа 1: 45, 67, 87, 71, 34, 12, 54, 57; группа 2: 46, 66, 81, 72, 11, 47, 55, 51, 9, 99. На уровне значимости 5% проверить гипотезу о том, что результаты двух групп не отличаются.

15.3. Имеются результаты нескольких студентов до и после апелляции (в скобках указан результат до апелляции): 48(47), 54(52), 67(60), 56(60), 55(58), 55(60), 90(70), 71(81), 72(87), 69(60). На уровне значимости 5% проверьте гипотезу о том, что апелляция в среднем не сказывается на результатах.

15.4. Имеются наблюдения за говорливостью 30 попугаев (слов/день): 34, 56, 32, 45, 34, 45, 67, 1, 34, 12, 123, ..., 37 (всего 13 наблюдений меньше 40). Проверить гипотезу о том, что медиана равна 40 (слов/день).

15.5. Вашему вниманию представлены результаты прыжков в длину Васи Сидорова на двух соревнованиях. На первых среди болельщиц присутствовала Аня Иванова (его первая любовь): 1,83; 1,64; 2,27; 1,78; 1,89; 2,33; 1,61; 2,31. На вторых Аня среди болельщиц не присутствовала: 1,26; 1,41; 2,05; 1,07; 1,59; 1,96; 1,29; 1,52; 1,18; 1,47. С помощью теста (Mann-Whitney) проверьте гипотезу о том, что присутствие Ани Ивановой положительно влияет на результаты Васи Сидорова. Уровень значимости $\alpha = 0.05$.

15.6. Некоторые результаты 2-х контрольных по теории вероятностей выглядят следующим образом (указан результат за вторую контрольную и в скобках результат за первую): 43(55), 113(108), 97(53), 68(42), 94(67), 90.5(97), 35(91), 126(127), 102(78), 89(83). Можно ли считать (при $\alpha = 0.05$), что вторую контрольную написали лучше?

15.7. Пусть с.в. X_i независимы и имеют функцию плотности $p(t) = \frac{a}{2}e^{-a \cdot |t|}$. Найдите оценку параметра a методом максимального правдоподобия и методом моментов а) приравняв теоретическую $Var(X_i)$ и эмпирическую $\hat{\sigma}^2$; б) приравняв $E(X^2)$ и соответствующий эмпирический момент.

15.8. [т?] Пусть с.в. X_i независимы и имеют функцию плотности $p(t) = \frac{a}{2}e^{-a \cdot |t-b|}$. Найдите ML оценку параметров a и b .

15.9. [к] Проведя 1000 экспериментов на компьютере найдите 5%-ое пороговое значение статистик U_1 (Mann-Whitney, $n_1 = 5$ и $n_2 = 4$) и T^+ (Wilcoxon Signed Rank Test, $n = 9$) для двусторонней альтернативной гипотезы. Сравните его с асимптотическим.

15.10. В темно-синем лесу, где трепещут осины, отловлено 100 зайцев. Каждому из них на левое ухо завязали бант из красной ленточки и отпустили. Через неделю будет снова отловлено 100 зайцев. Из них N (с.в.) окажутся с бантами. Найдите $E(N)$, если в лесу a зайцев. Придумайте ММ оценку общего числа зайцев.

15.11. [ш] Вася утверждает, что оценки метода максимального правдоподобия являются состоятельными. Является ли Васино утверждение а) правдоподобным; б) состоятельным?

15.12. Найдите ММ и ML оценки параметра a , если X_i - независимы и одинаково распределены с функцией плотности $p(t) = a^2 \cdot t \cdot e^{-a \cdot t}$ при $t > 0$.

Problem set 2⁴. HSE. Probability theory and MS. 2 year. 24.01.06.

Theory: If X_i - iid, $N(\mu, \sigma^2)$, then $\frac{(n-1) \cdot \hat{\sigma}^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n-1)}$.

16.1. Найдите $P(Y > 2)$, если $Y = \sum_{i=1}^9 X_i^2$, а X_i - iid $N(0; 1)$.

16.2. Пусть Y_1 имеет χ^2 распределение с 5-ю степенями свободы, а Y_2 - χ^2 распределение с 15-ю степенями свободы, причем Y_1 и Y_2 независимы. Как распределена их сумма?

16.3. Пусть $p_X(t)$ - функция плотности с.в. X . Найдите функцию плотности с.в. $Y = aX + b$, если $a > 0$.

16.4. По 820 посетителям супермаркета средние расходы на одного человека составили 340 рублей. Из достоверных источников известно, что дисперсия равна 90000 руб. Постройте 95% доверительные интервалы для средних расходов одного посетителя (двусторонний и два односторонних).

16.5. Контрольные камеры ДПС, установленные на некотором участке МКАД, фиксируют скорость движения автомобилей. Случайная выборка из 6 автомобилей: 89, 83, 78, 96, 81, 79 (км/ч). Вычислите выборочное среднее и выборочную дисперсию. Предполагая нормальность скорости постройте 90%-ый доверительный интервал для дисперсии скорости. Дополнительно предположив, что настоящая дисперсия равна 50 (км/ч)² постройте 90%-ый доверительный интервал для средней скорости. На 10%-ом уровне значимости проверьте гипотезу о том, что средняя скорость равна 90 км/ч.

16.6. Средний балл по диплому студента - с.в. $N(\mu; 0.04)$. Средний балл, рассчитанный по выборке из 25 абитуриентов этого года, составил 4.30. По данной выборке был построен доверительный интервал для μ : (4.2424; 4.3576). Какой уровень доверия соответствует этому интервалу?

16.7. В прошлом году средняя длина ушей зайцев в темно-синем лесу была 20 см, $\sigma = 4$. В этом году у случайно попавшихся 15 зайцев средняя длина оказалась 24 см. Предполагая нормальность распределения, проверьте гипотезу о том, что средняя длина ушей не изменилась (против альтернативной гипотезы о росте длины).

16.8. Вася Сидоров хвастался перед Аней Ивановой, что в среднем прыгает не меньше, чем на 2 метра. Напомним его результаты: 1,83; 1,64; 2,27; 1,78; 1,89; 2,33; 1,61; 2,31. В предположении нормальности постройте 80%-ый доверительный интервал для дисперсии длины прыжка. Дополнительно предположив, что $\sigma = 0,3$ проверьте гипотезу о том, что Вася действительно прыгает на 2 метра (Аня, естественно, ему не верит, и утверждает, что он прыгает меньше, чем на 2 метра). Постройте двусторонний 90%-ый доверительный интервал для μ , если $\sigma = 0,3$.

16.9. On 384 out of 600 randomly selected farms, the operator was also the owner. Find a 95% confidence interval for the true proportion of owner operated farms.

16.10. Стандартное отклонение количества иголок у ежа равно 130. По выборке из 12 ежей было получено среднее количество иголок 5120. Допустим, что количество иголок на одном еже можно считать нормально распределенным. Постройте 90%-ый доверительный интервал для среднего количества иголок. На 5%-ом уровне значимости проверьте гипотезу о том, что среднее количество иголок равно 5000.

Problems 9 is borrowed from www.math.cornell.edu/~durrett/ep4a/ep4a.html

Problem set 17. HSE. Probability theory and MS. 2 year. 31.01.06.

Theory: $\sum \frac{(X_i - np_i)^2}{np_i} \sim \chi_{r-1}^2$; $\sum \frac{(X_{i,j} - n\hat{p}_{i,j})^2}{n\hat{p}_{i,j}} \sim \chi_{(r-1)(c-1)}^2$. If $X \sim N(0; 1)$ and $K \sim \chi_n^2$ then $Y = \frac{X}{\sqrt{\frac{K}{n}}}$ is called t_n . If X_i - iid $N(\mu, \sigma^2)$, then $\frac{X_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim t_{n-1}$.

17.1. Пусть X_i - iid $N(0; 1)$. а) Как распределена случайная величина $\frac{X_1}{|X_2|}$? б) Найдите $P(X_5 > 2.3\sqrt{X_1^2 + X_2^2 + X_3^2 + X_4^2})$.

17.2. Пусть $Y \sim \chi_n^2$ и $W \sim t_n$. Найдите $E(Y)$, $E(W)$ и $[t?] Var(Y)$.

17.3. In 1882 Michelson performed experiments to measure the speed of light. 23 trials gave an average of 299756.2 km/sec with a standard deviation of 107.12. Find a 95% confidence interval for the speed of light. The correct answer is 299710.5 so there must have be some bias in his experiments.

17.4. Контрольные камеры ДПС на МКАД, зафиксировали скорость движения 6-и автомобилей: 89, 83, 78, 96, 81, 79. В предположении нормальности скоростей: а) Постройте 90%-ый доверительный интервал для дисперсии скорости. б) Постройте 90% доверительный интервал для средней скорости автомобилей. в) На 10%-ом уровне значимости проверьте гипотезу о том, что средняя скорость равна 90 км/ч.

17.5. An English biologist named Weldon was interested in the 'pip effect' in dice – the idea that the spots, or 'pips', which on some dice are produced by cutting small holes in the surface, make the sides with more spots lighter and more likely to turn up. Weldon threw 12 dice 26306 times for a total of 315672 throws and observed that a 5 or 6 came up on 106602 throws. Find a 95% confidence interval for the true probability of getting 5 or 6 on a dice.

17.6. During a two week period (10 weekdays) a parking garage collected an average of \$126 with a standard deviation of \$15. Find a 95% confidence interval for the mean revenue.

17.8. При подбрасывании кубика грани выпали 234, 229, 240, 219, 236 и 231 раз соответственно. Проверьте гипотезу о том, что кубик "правильный".

17.9. Проверьте независимость дохода и пола по таблице:

	< \$500	\$500 – \$1000	> \$1000
М	112	266	34
Ж	140	152	11

17.10. Вася Сидоров утверждает, что ходит в кино в два раза чаще, чем в спортзал, а в спортзал в два раза чаще, чем в театр. За последние полгода он 10 раз был в театре, 17 раз - в спортзале и 39 раз - в кино. Правдоподобно ли Васино утверждение?

17.11. Проверьте независимость пола респондента и предпочитаемого им сока:

	Апельсиновый	Томатный	Вишневый
М	69	40	23
Ж	74	62	34

Problems 3, 5, 6 are borrowed from www.math.cornell.edu/~durrett/ep4a/ep4a.html

Problem set 18. HSE. Probability theory and MS. 2 year. 07.02.06.

18.1. Пусть $X \sim N(0; 1)$; Z равновероятно принимает значения 1 и -1 ; X и Z независимы. Рассмотрим $Y = X \cdot Z$. Найдите:

- а) закон распределения Y ;
- б) $Cov(X, Y)$;
- в) условное распределение X , если известно, что $Y = y$.
- г) верно ли, что $X + Y$ нормально?

18.2. Пусть $X \sim t_n$. Как распределена величина $Y = X^2$?

18.3. Пусть X_i - iid $N(0; 1)$. Найдите $P(X_1^2 + X_2^2 > 6.37 \cdot (X_3^2 + X_4^2 + X_5^2))$.

18.4. Изучалось воздействие модератора на количество идей, сочиняемых группой людей. По выборке из 4-х групп с модератором, среднее количество идей оказалось равным 78, при стандартном отклонении 24.4, по выборке из 4-х групп без модератора, среднее количество идей оказалось равным 63.5, при стандартном отклонении 20.2. Предположим нормальность распределения.

- а) Проверьте гипотезу о равенстве дисперсий.
 - б) Предполагая равенство дисперсий проверьте гипотезу о равенстве средних.
- 18.5. Маркетинговый отдел банка опросил 300 женщин и 400 мужчин. Оказалось, что реклама банка вызывает положительные эмоции у 74% опрошенных женщин и 69% опрошенных мужчин.
- а) Можно ли считать, что реклама банка одинаково нравится мужчинам и женщинам?
 - б) Постройте 90% доверительный интервал для разницы долей мужчин и женщин, одобряющих рекламу банка.
 - в) Проверьте гипотезу о том, что рекламу одобряет 70% женщин (против гипотезы о том, что рекламу одобряет более 70% женщин).
 - г) Предположим, что среди потребителей рекламы мужчин и женщин поровну. Постройте 95% доверительный интервал для доли людей, которым нравится реклама банка.

18.6. Исследователь сравнивал суровость климата (дисперсию температуры) в двух странах. Для этого случайным образом были выбраны 37 наблюдений за среднедневной температурой в первой стране и 46 наблюдений за среднедневной температурой во второй стране. Известно, что $\bar{X}_I = 14$, $\bar{X}_{II} = 11$, $\hat{\sigma}_I^2 = 2341$ и $\hat{\sigma}_{II}^2 = 3079$.

- а) Постройте 95% доверительный интервал для разности математических ожиданий среднедневных температур в двух странах.
- б) Предполагая, что среднедневная температура распределена нормально, проверьте гипотезу об одинаковой суровости климата. С помощью компьютера найдите точное Р-значение.
- в) В предположениях о нормальности постройте 90%-ые доверительные интервалы для дисперсий среднедневной температуры в двух странах. Почему один из интервалов оказался шире?

18.7. Известно, что X_i iid $N(\mu; 900)$. Исследователь проверяет гипотезу $H_0: \mu = 10$ против $H_A: \mu = 30$ по выборке из 20 наблюдений. Критерий выглядит следующим образом: если $\bar{X} > c$, то выбрать H_A , иначе выбрать H_0 .

- а) Рассчитайте вероятности ошибок первого и второго рода, мощность критерия для $c = 25$.
- б) Что произойдет с указанными вероятностями при росте количества наблюдений ($c \in [10; 30]$)?
- в) Каким должно быть c , чтобы вероятность ошибки второго рода равнялась 0,15?
- г) Как зависят от c вероятности ошибок первого и второго рода?
- д) Объясните, почему при сравнении данных гипотез действительно разумно использовать предложенный тест

18.8. Дама утверждает, что обладает особыми способностями и безошибочно отличает "бонакву" без газа от "святого источника" без газа. H_0 : дама не обладает особыми способностями, H_A : дама обладает особыми способностями. При даме в 3 стаканчика из 8-ми налили "бонакву" а в 5 оставшихся - "святой источник". При отгадывании стаканчики предлагаются даме в неизвестном ей порядке. Критерий: принимается основная гипотеза, если дама ошиблась хотя бы один раз и альтернативная иначе.

- а) Рассчитайте вероятности ошибок первого и второго рода, мощность критерия.
- б) Сколько из 8 стаканчиков надо наполнить "бонаквой" и сколько "святым источником" чтобы вероятность ошибки первого рода была минимальной?
- 18.9. Имеются две нормальные выборки: $\{X_i\} = \{34, 28, 29, 41, 32\}$, $\{Y_i\} = \{32, 30, 31, 25, 24, 29\}$.
- а) Проверьте гипотезу о равенстве дисперсий. С помощью компьютера укажите точное Р-значение
- б) В предположении о равенстве дисперсий проверьте гипотезу о равенстве математических ожиданий. Укажите точное Р-значение.

19.1. The psychologist Tversky and his colleagues say that about four out of five people will answer (a) to the following question:

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital 15 babies are born each day. Although the overall proportion of boys is about 50 percent, the actual proportion at either hospital may be more or less than 50 percent on any day. At the end of a year, which hospital will have the greater number of days on which more than 60 percent of the babies born were boys?

(a) the large hospital (b) the small hospital (c) neither (about the same).

Дайте верный ответ и попытайтесь объяснить, почему большинство людей ошибается при ответе на этот вопрос.

19.2. In the early 1600s, Galileo was asked to explain the fact that, although the number of triples of integers from 1 to 6 with sum 9 is the same as the number of such triples with sum 10, when three dice are rolled, a 9 seemed to come up less often than a 10-supposedly in the experience of gamblers. Верный ответ?

19.3. Четыре свидетеля, A, B, C и D, говорят правду независимо друг от друга с вероятностью $\frac{1}{3}$. A утверждает, что B отрицает, что C заявил, что D солгал. Какова (условная) вероятность того, что D сказал правду?

19.4. Саша едет на день рождения к Маше и ищет ее дом. Ее дом находится южнее. Одна треть встречаемых прохожих - местные. Местные всегда лгут, неместные говорят правду с вероятностью $\frac{3}{4}$. Изначально Саша оценивает вероятность того, что дом находится южнее, как a . Саша спросил первого встречного прохожего и получил ответ "севернее". Как Саша изменит свою субъективную вероятность?

19.5. [Problem of the points] Монетка выпадает орлом с вероятностью p . Игрок A выигрывает, если k орлов появятся раньше чем l решек (не обязательно подряд). Игрок B выигрывает в противоположном случае. Какова вероятность того, что A выиграет?

19.6. Возможно ли, что $U = X + Y$, X и Y - iid, а U - равномерна на $[0; 1]$?

19.7. [экзамен 1858 года St John's College, Cambridge] A large quantity of pebbles lies scattered uniformly over a circular field; compare the labour of collecting them on by one: (i) at the center O of the field, (ii) at a point A on the circumference.

А вам слабо? Ответ: $\frac{E(L_A)}{E(L_O)} = \frac{16}{3\pi}$

19.8. Имеется пять чисел: x , 4, 5, 7, 9. Когда медиана будет равна среднему?

19.9. [Сумасшедшая старушка] В автобус дальнего следования, имеющий n мест, проданы все билеты. На каждом билете написан номер места пассажира. Для посадки в автобус пассажиры выстроились в очередь (не обязательно по номерам билетов). Среди пассажиров есть сумасшедшая старушка. Она растолкала всех локтями, первой ворвалась в салон и села на первое понравившееся ей место. Нормальный пассажир садится на свое место, если оно не занято; если оно занято, то пассажир садится произвольным образом на любое свободное. Какова вероятность того, что последний в очереди пассажир сядет на свое место?

19.10. [Парадокс голосования] Пусть X , Y , Z - дискретные случайные величины, их значения попарно различны с вероятностью 1. Докажите, что

$\min \{P(X > Y), P(Y > Z), P(Z > X)\} \leq \frac{2}{3}$. Приведите пример, при котором эта граница точно достигается.

Introduction to Probability Charles, M. Grinstead, J. Laurie Snell, открыто доступна в интернете Tversky et. al. in Judgement Under Uncertainty: Heuristics and Biases (Cambridge: Cambridge University Press, 1982).

Probability and Random Processes, Grimmett, Stirzaker, доступна в библиотеке ГУ-ВШЭ

Существует аргумент в пользу того, что неправильное интуитивное представление об условных вероятностях способствует выживанию человека как вида. Если в игре равновесие по Нэшу не Парето-оптимально, то игрок, который считается иррациональным, может способствовать достижению Парето-оптимальности.