1 Предварительные замечания

1. Обратить внимание на формулу оценки, см. смарт лмс, в частности, нюанс с блокированием.

2. ВНИМАНИЕ! На теоретический вопрос 2.5-3 балла. ПОСМОТРИТЕ структуру этого вопроса в примере экз. билета №1. Сравните его с тем же вопросом из списка теор. вопросов. Сделайте вывод.

3. Возможна небольшая коррекция вопросов как в большую, так и в меньшую сторону в зависимости от темпа прохождения материала. Следите за смарт лмс, разделом в котором находится данный материал: здесь будет опубликована дата последнего обновления и обращено внимание на изменения, если таковые будут.

2 Теоретические вопросы

1. Дискретная случайная величина, определение, способы задания. Примеры. Могут ли две разные дискретные с.в. иметь одинаковые таблицы распределения?

2. Определение и свойства функции распределения дискретной с.в. (перечислить). Доказать, что $F(b) - F(a) = P(a < X \leq b)$. Нарисовать эскиз типичной функции распределения дискретной с.в.

3. Биномиальное распределение и его основные параметры. Докажите, что если $X \sim \mathrm{Bin}(n, p)$, то $E(X) = np$.

4. Распределение Пуассона и его основные параметры. Докажите, что для $X \sim \mathrm{Pois}(\lambda)$ дисперсия $\mathrm{Var}(X) = \lambda$. Типичные задачи, предполагающие использование пуассоновской с.в.

5. Приближение биномиальной случайной величины пуассоновской с доказательством. Укажите параметры биномиального распределения, когда данное приближение будет корректным: (i) $n = 100$, $p = 0.01$; (ii) $n = 10^6$, $p = 10^{-7}$; (iii) $n = 10$, $p = 0.4$

6. Непрерывная случайная величина. Плотность вероятности, ее свойства, функция распределения н.с.в., связь между ними. Пусть f и g – плотности распределений. Могут ли являться плотностями некоторых случайных величин функции $2f$, $f + g$, $\frac{f+g}{2}$, $f - g$, $\frac{2}{3}f + \frac{1}{3}g$? Объяснить.

7. Среднее значение случайной величины X (дискретный и непрерывный случай). Докажите, что $E(aX + b) = aE(X) + b$. Среднее значение функции случайной величины. Может ли среднее $E(X)$ не равняться ни одному возможному значению дискретной с.в.? Если нет — доказать, если да – привести пример. Привести пример с.в., среднее значение которой не существует.

8. Дисперсия и стандартное отклонение случайной величины X (дискретный и непрерывный случай). Чему равно $\mathrm{Var}(aX + b)$, где a и b константы? Докажите.

9. Равномерная и экспоненциальная случайная величина, определение, сфера применимости данных распределений. Вывод среднего и дисперсии для каждого распределения.

10. Нормальное распределение и его свойства. Правило трех сигм. Пусть $X \sim N(\mu, \sigma^2)$, докажите, что $\mathrm{Var}(X) = \sigma^2$. Как распределена с.в. $Y = \frac{X-\mu}{\sigma}$? Объяснить. Функция распределения стандартного нормального закона и ее свойства.

11. Приближение биномиальной случайной величины с помощью нормальной. Формулировка утверждения, область параметров, в которой приближение работает. Поправка на непрерывность.

12. Совместный закон распределения двумерной дискретной с.в., определение. Процедура маржинализации. Переход к функции распределения. Условные законы распределения. Требуется сквозной пример, на котором демонстрируются все указанные понятия. Как посчитать условное среднее значение $E(X|Y = y)$? Что означают входящие в данное выражение буквы X,Y,y?

13. Совместная плотность вероятности двумерной непрерывной с.в., определение. Процедура маржинализации, что получается в результате. Переход к функции распределения. Условные плотности. Требуется сквозной пример, на котором демонстрируются все указанные понятия. Как посчитать условную дисперсию $\mathrm{Var}(X|Y = y)$? Что означают входящие в данное выражение буквы X,Y,y?

14. Ковариация и корреляция двух случайных величин в дискретном и непрерывном случае: определение и свойства (перечислить, доказать альтернативное определение ковариации и ограниченность коэффициента корреляции). Изобразить качественно диаграмму рассеяния (облако данных) для коэффицента корреляции, равного ± 1, ± 0.5, 0. Ковариационная матрица для n-мерной с.в. Что стоит на главной диагонали этой матрицы?

15. Независимость двух случайных величин. Пример зависимых с.в. с формально факторизованной совместной плотностью. Нулевая корреляция и независимость двух случайных величин, взаимосвязь (с док-вом).Пример зависимых с.в. с нулевой корреляцией. Среднее значение произведения независимых с.в. (док-во для двумерной с.в.).

16. Вывести выражение для дисперсии линейной комбинации с.в. $\sum_{i=1}^{n} w_i X_i$. Записать результат в векторно-матричном виде. Рассмотреть частный случай равных весов $w_i = \frac{1}{n}$ и одинаково и независимо распределенных с.в. X_i, $i = 1, ..., n$, таких, что $\mathrm{Var}(X_i) = \sigma^2$. Найти среднеквадратичное отклонение такой линейной комбинации.

17. Вывод плотности суммы двух независимых непрерывных с.в. Пример двух равномерно распределенных с.в.

18. Распределение линейной комбинации нормально распределенных случайных величин (без доказательства).

19. Сформулировать и доказать неравенства Маркова и Чебышева.

20. Сформулировать и доказать закон больших чисел (слабая и сильная форма, док-во для слабой). Показать, что из ЗБЧ следует сходимость относительной частоты события к его вероятности.

21. Сформулировать центральную предельную теорему, обсудить ее значение и следствия.

22. Основные элементы статистического подхода: генеральная совокупность, модель признака объекта в ген. совокупности, модель случайной выборки (свойства).

23. Понятие статистики как функции элементов случайной выборки. Найти распределение выборочного среднего (указать предположения, которые при этом делаются).

24. Понятие статистики как функции элементов случайной выборки. Найти функцию распределения выборочного максимума.

25. Статистические оценки (определение, отличие от произвольной статистики) и их свойства (определение): смещение, состоятельность. Найти смещение выборочного среднего $\bar{X} = \frac{X_1+...+X_n}{n}$. Определить, состоятельна ли эта оценка в каком-либо смысле.

26. Определение смещения оценки. Найти смещение оценок дисперсии $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ и $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

27. Определение среднеквадратичной ошибки $\mathrm{MSE}(\hat{\theta})$ оценки $\hat{\theta}$. Вывести разложение $\mathrm{MSE}(\hat{\theta})$ через смещение оценки и ее дисперсию. Определить, состоятельна ли оценка среднего $\bar{X} = \frac{X_1+...+X_n}{n}$ в среднеквадратичном смысле. Состоятельна ли она в слабом смысле?

28. Эмпирическая функция распределения и гистограмма (определение). К чему сходится эпмпирическая функция распределения и гистограмма в слабом смысле (с доказательством)?

29. Метод моментов построения оценок: общая схема и аргументация в пользу метода. Сколько уравнений требуется составить в рамках данного метода?

30. Метод максимального правдоподобия посроения оценок, схема метода. Всегда ли оценка макс. правдоподобия единственна (пример)?

31. χ^2-распределение и распределение Стьюдента: конструкция этих случайных величин, эскизы графиков плотностей, предельный переход при $n \to \infty$.

32. Доверительный интервал для среднего в нормально распределенной ген. совокупности при известной дисперсии признака: вывод формулы, анализ ширины доверит. интервала при изменении параметров задачи.

33. Доверительный интервал для среднего в нормально распределенной ген. совокупности при неизвестной дисперсии признака: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи.

34. Доверительный интервал для среднего в произвольно распределенной ген. совокупности при неизвестной дисперсии признака и большой выборке: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи.

35. Доверительный интервал для разности средних в нормально распределенных ген. совокупностях при известных дисперсиях: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи. Какой вывод можно сделать с помощью такого доверит. интервала?

36. Доверительный интервал для доли в ген. совокупности при большом объеме выборки: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи.

37. Доверительный интервал для разности долей в ген. совокупностях при больших объемах выборок: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи. Какой вывод можно сделать с помощью такого доверит. интервала?

38. Доверительный интервал для дисперсии в нормально распределенной ген. совокупности: вывод формулы, анализ ширины доверит. интервала в зависимости от параметров задачи.

39. Тестирование гипотезы о среднем в нормально распределенной ген. совокупности при известной дисперсии признака: основная (простая) и альтернативные гипотезы, z-статистика теста, уровень значимости теста, критическая область в увязке с альтернативной гипотезой, правило отвержения/неотвержения нулевой (основной) гипотезы, p-значение теста в увязке с альтернативной гипотезой и его использование для процедуры тестирования.

40. Тестирование гипотезы о среднем в нормально распределенной ген. совокупности при неизвестной дисперсии признака: основная (простая) и альтернативные гипотезы, t-статистика теста, уровень значимости теста, критическая область в увязке с альтернативной гипотезой, правило отвержения/неотвержения нулевой (основной) гипотезы, p-значение теста в увязке с альтернативной гипотезой и его использование для процедуры тестирования.

41. Тестирование гипотезы о доле в ген. совокупности при большом объеме выборки: основная (простая) и альтернативные гипотезы, z-статистика теста, уровень значимости теста, критическая область в увязке с альтернативной гипотезой, правило отвержения/неотвержения нулевой (основной) гипотезы, p-значение теста в увязке с альтернативной гипотезой и его использование для процедуры тестирования.

42. Тестирование гипотезы о разности долей в ген. совокупностях при больших объемах выборок: основная (простая) и альтернативные гипотезы, z-статистика теста, уровень значимости теста, критическая область в увязке с альтернативной гипотезой, правило отвержения/неотвержения нулевой (основной) гипотезы, p-значение теста в увязке с альтернативной гипотезой и его использование для процедуры тестирования. Какой вывод можно сделать с помощью такого теста?

43. Тестирование гипотезы о разности средних в нормально распределенных ген. совокупностях при известных дисперсиях: основная (простая) и альтернативные гипотезы, z-статистика теста, уровень значимости теста, критическая область в увязке с альтернативной гипотезой, правило отвержения/неотвержения нулевой (основной) гипотезы, p-значение теста в увязке с альтернативной гипотезой и его использование для процедуры тестирования. Какой вывод можно сделать с помощью такого теста?

44. Ошибки I и II второго рода при тестировании гипотез. Мощность теста.

Литература:

0. Лекции текущего года

0+. Видео лекций прошлых лет, размещенные на Smart LMS. Иметь в виду, что там много сходного с циклом текущих лекций, но каких-то вещей там может не быть, или они рассказаны по-другому. Отсылки на отсутствие тех или иных вопросов в данных видеозаписях не принимаются. 100% актуальными считаются лекции текущего года.

0++. Есть смысл проделать тесты на смарт лмс.

1. Шведов А.С. Теория вероятностей и мат. статистика, изд-во ВШЭ (несколько версий и изданий)
2. Гмурман, Теория вероятностей и мат. статистика, учебник для вузов
3. Ю.А. Кацман Теория вероятностей и мат. статистика, учебник для вузов
4. Любые другие известные учебники (Вентцель, Феллер, Гнеденко, Ширяев)

3 Примеры экзаменационного билета

Вариант 1

1. (**2 балл**) С.в. X распределена с плотностью

$$f(x) = \begin{cases} C(4-x^2), & |x| < 2 \\ 0, & \text{иначе} \end{cases}$$

Найти фунцию распределения X (на всей области определения $-\infty < x < \infty$) и с помощью нее вероятность $P(X > 0)$, а также 1-й квартиль.

Ответ: $C = 3/32$, $F(x) = -1/32\,x^3 + 1/2 + 3/8\,x$, $x = [-2, 2]$, $P(X > 0) = 1/2$

2. (**1 балл**) Имеется 40 батарей, 20 штук типа A и 20 штук типа B. Батарея типа A работает в среднем 50 часов, с.к.о. ее срока службы равно 15; батарея типа B служит в среднем 30 часов, с.к.о. равно 6. Приближенно найти вероятность, что всех этих батарей при последовательной работе (одна садится, ее тут же заменяют другой) хватит не менее, чем на 1700 часов работы.

Ответ: в рамках ЦПТ срок службы батарей типа A: $S_A \sim N(1000, 4500)$, батарей типа B: $S_B \sim N(600, 720)$. В силу нормальности и независимости двух с.в. срок службы всех 40 батарей $S = S_A + S_B \sim N(1000 + 600, 4500 + 720)$ Ответ: 0.084

3. (**2 балла**) Случайная выборка X_1 размера $n = 1$ взята из генеральной совокупности следующей конструкции: дискретная с.в. X в этой совокупности может принимать лишь три значения, -1, 0 и 1, с вероятностями α, $1-3\alpha$ и 2α соответственно. Рассматриваются две оценки параметра α: $\hat{\alpha}_1 = X_1$ $\hat{\alpha}_2 = X_1^2/3$. Рассчитать среднеквадратичную ошибку этих оценок и на этой основе выбрать более эффективную.

Ответ: $\text{Bias}(\hat{\alpha}_1) = 0$, $\text{Var}(\hat{\alpha}_1) = \alpha(3-\alpha) = MSE_1$, $\text{Bias}(\hat{\alpha}_2) = 0$, $\text{Var}(\hat{\alpha}_2) = \frac{1}{3}\alpha(1-3\alpha) = MSE_2$, $MSE_1 > MSE_2$

4. (**2 балла**) В случайной выборке объема 49 из нормально распределенной ген. совокупности выборочное среднее $\bar{x} = 25$, генеральное (известное) с.к.о. $\sigma = 9$. Протестируйте гипотезу о генеральном среднем $H_0 : \mu = 29$ против гипотезы $H_1 : \mu \neq 29$ на 1% уровне значимости. Сделайте вывод. Посчитайте p-значение теста и сделайте вывод.

Ответ: z-статистика $\frac{25-29}{9/\sqrt{49}} = -3.11$, критическая область $|z| > 2.57$; отвергаем. p - значение $= 2P(z > 3.11) = 2(1 - 0.99906)$

5. (**3 балла**) Биномиальное распределение:

   (a) смысл значений биномиальной с.в. и его параметров (**0.25**)

   (b) пусть $X \sim \mathrm{Bin}(n,p)$, выведите выражение для дисперсии X (**1.5**).

   (c) нарисуйте эскиз распределения биномиальной с.в., если (i) $n = 100$, $p = 0.2$ ii) $n = 100$, $p = 0.5$. Нанесите на ось x необходимую разметку (**0.5**)

   (d) в какой точке достигается максимум вероятности с.в. $X \sim \mathrm{Bin}(n,p)$ (вывод формулы!) (**0.5**)?

   (e) укажите все возможные значения с.в. $X \sim \mathrm{Bin}(n,p)$ (**0.25**)

Вариант 2

1. (**2 балла**) Одна серия экспериментов состоит из 90 опытов, каждый из которых заканчивается успехом с малой вероятностью 0.01. Ученый проводит пять серий таких экспериментов. Найти вероятность, что будут по крайней мере три серии, в каждой из которых успех наступит по крайней мере два раза. Указать вид и параметр(ы) апроксимирующего распределения.

   Ответ: 0.0757 здесь приближать Пуассоном, а не нормальным, т.к. параметр $np(1-p) < 5$. $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$, $P(X = k) \approx e^{-\lambda}\lambda^k/k!$, $\lambda = 0.01 \cdot 90$

2. (**2 балла**) Совместная плотность вероятности случайных величин X,Y имеет вид $f_{XY}(x,y) = cxy^2$, $0 < x < y$, $0 < y < \infty$. Найти $E(X^2|Y = y)$. Независимы ли X,Y?

3. (**2 балла**) Пусть X_i, $i = 1, ..., N$ – выборка из генеральной совокупности, распределенной по закону Пуассона с параметром λ. Найти оценку параметра λ методом максимального правдоподобия, найти ее смещение, указать, состоятельна ли она в среднеквадратичном смысле.

   Ответ: $\hat{\lambda} = \frac{1}{N}\sum_{i=1}^{N} X_i$, $\mathrm{Bias}(\hat{\lambda}) = 0$, $\mathrm{MSE}(\hat{\lambda}) = 0 + \mathrm{Var}(\hat{\lambda}) = \lambda/N \to 0$ при $N \to 0$

4. (**1 балл**) Из нормально распределенной ген. совокупности взята выборка размера 16, выборочное среднее равно 15, <u>выборочное</u> с.к.о. равно 3. Найти 90% доверительный интервал для среднего в ген. совокупности

   Ответ: использовать t-распределение

5. (**2.5 - 3 балла**) Теор. вопрос

4 Общие рекомендации по решению задач

Темы, на что обратить внимание:

- Одномерная дискретная случайная величина. Знать и уметь:
  - строить и работать с таблицей (законом) распределения
  - функция распределения дискретной случайной величины: определение через вероятность, вычисление по таблице распределения и обратно (т.е. дана $F(x)$, построить таблицу).

  - среднее, дисперсия, с.к.о.: прямое вычисление через таблицу распределения, формула $\mathrm{Var}(X) = E(X^2) - E^2(X)$.
  - свойства среднего и дисперсии

- Биномиальное распределение и распределение Пуассона:

  - соответствующие формулы для вычисления вероятностей, среднего, дисперсии.
  - понимать из текста задачи, какое распределение использовать (может быть не сказано в задаче намеренно)
  - пуассоновское приближение биномиального закона: уметь применять и знать условие применимости приближения

- Одномерная непрерывная с.в.:

  - плотность вероятности: нормировка, правило вычисления вероятностей, среднего, дисперсии, среднего квадратичного отклонения, среднего произвольной функции с.в., вычисление условных вероятностей для н.с.в.
  - частный случай предыдущего пункта: равномерно распределенная с.в.,экспоненциальная с.в.: выражение для плотности, среднее, дисперсия
  - связь плотности с функцией распределения (интегральная и дифференциальная), вычисление вероятностей через функцию распределения
  - расчет медианы, квартилей с помощью функции распределения
  - нормальный закон: стандартизация нормально распределенной с.в. с произвольными параметрами, вычисление вероятностей, работа с таблицей функции распределения стандартного нормального закона $\Phi(x)$ (табл. **??**), формула для $\Phi(-x)$, распределение линейного преобразования с.в. $aX + b$, где $X \sim N(\mu, \sigma^2)$, a,b – константы
  - применение т. Муавра-Лапласа для вычисления вероятностей биномиальной с.в., условие применимости данного приближения, отличие от “пуассоновского” предела; поправка на непрерывность
  - поиск плотности вероятности функции случайной величины

- Многомерная дискретная с.в.

  - по заданному совместному распределению двухмерной с.в. (таблице) уметь рассчитывать вероятности (например, $P(X + Y > 1)$), одномерные и условные распределения X и Y, различные средние (безусловные и условные), дисперсии/с.к.о.(безусловные и условные), ковариацию и коэф-т корреляции.
  - делать вывод о зависимости X и Y (факторизация соответствующих вероятностей и/или на языке условных распределений)

- Многомерная непрерывная с.в.

  - по заданной совместной плотности рассчитывать вероятности (например, $P(X + Y > 1)$), одномерные и условные плотности, различные средние (безусловные и условные), дисперсии/с.к.о.(безусловные и условные), ковариацию и коэф-т корреляции. Возможен вариант с трехмерной с.в. (а не только с двухмерной) с несложной для расчетов плотностью – все вышеперечисленное.

- делать вывод о зависимости X и Y (факторизация соответствующих плотностей и/или на языке условных плотностей)

- Линейная комбинация нормально распределенных величин
  - знать закон распределения для $X = a_1X_1 + a_2X_2 + ... + a_nX_n$, если $X_i \sim N(\mu_i, \sigma_i^2)$ при отсуствии и наличии парной корреляции между X_i. Уметь пользоваться этим приемом для расчета вероятностей типа $P(X_1 - 3X_2 > 5 - X_3)$.

- Свойства ковариации и коэф-та корреляции
  - знать свойства ковариации (билинейность, "переход" в дисперсию,...)
  - знать свойства коэф-та корреляции (формула, диапазон значений, интерпретация предельных значений ± 1)
  - уметь вычислять дисперсию $\mathrm{Var}(a_1X_1 + a_2X_2 + ... + a_nX_n)$ и записывать результат в векторно-матричном виде

- Неравенства Маркова, Чебышева и центральная предельная теорема
  - ЦПТ – возможны задачи на суммы экспоненциальных, биномиальных, пуассоновских и равномерных с.в., а также с.в. с произвольным распределением с известным или требующим расчета средним/дисперсией. Возможны "комбо"-задачи с дальнейшей манипуляцией полученными в результате применения ЦПТ нормальными с.в. Неравенства Маркова и Чебышева также могут встретиться в качестве отдельной задачи.

- Для успешного решения задач в билете необходимо помнить следующие виды распределений: нормальное, пуассоновское, Бернулли, биномиальное, экспоненциальное, равномерное. Помнить их плотность, среднее и, желательно, дисперсию.

- Различные типы сходимости (четыре вида)

- t, χ^2 - распределения
  - знать, из чего устроены указанные распределения (например, с.в. по χ^2 = сумма квадратов стандартной нормальной с.в.)
  - уметь работать с таблицами для указанных распределений, особенно в контексте доверительных интервалов и тестирования гипотез
  - понимать, как определить число степеней свободы в конкретной задаче

- Распределение для выборочного среднего, выборочного макс/мин

- Точечные оценки:
  - смещение, дисперсия, среднеквадратичная ошибка – уметь считать. Связь между указанными величинами.
  - состоятельность оценки в слабом и среднеквадратичном смысле
  - выбор наиболее эффективной оценки

- несмещенная оценка дисперсии – знать формулу (использовать при необходимости в доверит. интервалах и тестировании гипотез)

- Метод максимального правдоподобия и метод моментов
  - в методе макс. правдоподобия возможны варианты как с дискретной, так и с непрерывной с.в. (задействуется плотность).

- Доверительные интервалы (формулы надо помнить или выводить, все они устроены структурно одинаково, кроме формулы для с.к.о.):
  - для генерального среднего: с.к.о. σ в генеральной совокупности известна/не известна
  - для доли в генеральной совокупности
  - для дисперсии в нормальной генеральной совокупности
  - для разности средних в двух ген. совокупностях (с.к.о. σ_1 и σ_2 известны).
  - для разности долей в двух ген. совокупностях
  - вывод по построенному доверит. интервалу для разности (средних, долей), что больше/меньше/вывод нельзя сделать

- Тестирование гипотез (формулы надо помнить или выводить):
  - различные альтернативные гипотезы. Могут быть как односторонними, так и двусторонними.
  - тестирование гипотезы как на языке критических областей, так и на языке p-значений (p-value) теста
  - p-значение зависит от альтернативной гипотезы

При построении доверит. интервалов и тестировании гипотез для среднего значения и дисперсии в нормально распределенной ген. совокупности может быть случай с известной дисперсией (используем таблицы для нормального закона) и случай неизвестной дисперсии, замененной вычисленной несмещенной оценкой дисперсии – используем таблицы t-распределения с $n-1$ степенями свободы. В этой связи также надо помнить, что при больших $n > 30$ распределение Стюдента переходит в нормальное, и в таком пределе можно снова использовать z-таблицы.

5 Чем можно пользоваться для подготовки (практические задачи)

5.1 Задачи для тренировки

Ниже приведены примеры задач, уверенное решение которых сильно повышает шансы хороших результатов на экзамене. Иногда в этом списке встречаются несколько более сложные задачи. Номера даны из задачника, приведенного в конце данного раздела (доступен в интернете).

1. В ящике находится 20 деталей, 4 из которых бракованные. Из ящика берется выборка из трех деталей. Найти среднее значение бракованных деталей в выборке такого размера.

2. Кость подбрасывается до тех пор, пока единица не выпадет в 10-й раз. Найти дисперсию числа бросков, требуемых для этого.

3. Пусть с.в. принимает целые значения $X = 0, 1, ..., 10$ и пусть $P(X = i) = c(i + 1)$, $i = 0, 1..., 10$. Найти $E(X)$.

4. Ежедневное среднее число аварий на определенном шоссе равно 3. Найти вероятность, что сегодня произойдет не менее трех аварий. Тот же вопрос при условии, что по крайней мере одна авария уже произошла.

   Ответ: 0.5768; 0.6070

5. Прямым вычислением продемонстрировать качество аппроксимации биномиального распределения пуассоновским в следующих случаях: (а) $P(X = 2), n = 8, p = 0.1$ (б) $P(X = 9), n = 10, p = 0.95$ (в) $P(X = 0), n = 10, p = 0.1$ (г) $P(X = 4), n = 9, p = 0.2$

6. Вы покупаете по одному билету в пятидесяти лотереях, в каждой из которых шанс выиграть равен 0.01. Чему равна приближенная (через Пуассона) вероятность, что вы выиграете (а) по меньшей мере, один раз (б) только один раз (в) не меньше двух разю Можно ли использовать это приближение в данной задаче?

   Ответ: 0.3935; 0.3033; 0.0902

7. Посетители заходят в казино со средней скоростью 1 человек в две минуты. Найти вероятность, что (а) никто не войдет в интервал времени 12:00-12:05 (б) в этот интервал войдут по меньшей мере 4 человека.

   Ответ: 0.0821; 0.2424

8. (Пуассон+биномиальный) За период в месяц один человек из 100000 заболевает редкой болезнью. Какова вероятность, что в городе с населением 600000 человек в течение года будет по крайней мере 2 месяца с числом заболевших (за месяц) более 5?

9. Приведите пример с.в. с конечным средним, но с несуществующей конечной дисперсией.

10. Случайная величина X распределена с плотностью

$$f(x) = \begin{cases} c(1 - x^2), & -1 < x < 1 \\ 0, & |x| \geqslant 1 \end{cases}$$

   Найти константу c и функцию распределения $F(x)$.

   Ответ: 3/4; $F(x) = 0$, $x < -1$; $F(x) = -x^3/4 + 1/2 + 3x/4$, $|x| < 1$; $F(x) = 1$, $x > 1$.

11. Пусть

$$F(x) = \begin{cases} 1 - e^{-2x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

   Найти $\lim\limits_{\Delta x \to 0} \frac{P(X \in [1, 1+\Delta x])}{\Delta x}$

12. С.в. $X \sim \mathrm{Unif}[5, 20]$. Найти дисперсию с.в. $Y = 10X + 50$

13. С.в. $X \sim \mathrm{Unif}[0, 1]$. Найти дисперсию с.в. $Y = X^{7/4}$

14. Может ли функция

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x-1}{3}\right)^3}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

быть функцией распределения какой-либо с.в.?

15. Время (в месяцах) безотказной работы системы X случайно и распределено с плотностью

$$f(x) = \begin{cases} Cxe^{-x/2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Найти вероятность, что система проработает не меньше пяти месяцев.

Ответ: $3.5e^{-5/2}$

16. Функция распределения с.в. X имеет вид:

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ 0.4x^{3/2} + 0.6x, & 0 < x \leqslant 1 \\ 1, & x > 1 \end{cases}$$

Найти $E(X)$ и $P(X < 9/16 \,|\, X > 1/4)$.

Ответ: 0.54; 0.3828

17. С.в. X распределена с плотностью $f(x) = 3e^{-3x}$, $x > 0$. Пусть $Y = e^X$. Найти $E(Y)$,$\mathrm{Var}(Y)$.

Ответ: 1.5; 0.75

18. Пусть $X \sim \mathrm{Exp}(\lambda)$ – непрерывная с.в., распределенная по экспоненциальному закону с плотностью $f(x) = \lambda \exp(-\lambda x)$, $x \geq 0$ и $f(x) = 0$ в противном случае. Найти α%-квантиль, т.е. значение x с.в. X, такое, что $P(X \leq x) = \alpha/100$. Сравнить со средним значением $E(X)$.

Ответ: $-\frac{1}{\lambda}\ln(1 - \alpha/100)$; $1/\lambda$

19. Пассажир прибывает на остановку в 10 утра, зная, что автобус прибудет равновероятно в любой момент из промежутка 10:00–10:30. (а) Найти вероятность, что пассажир прождет автобус более 10 минут (б) Пусть на 10:15 автобус еще не прибыл. Найти вероятность, что придется ждать еще по крайней мере 10 минут.

Ответ: 2/3; 1/3

20. Пусть $X \sim N(2, 3^2)$. Найти $P(X^3 > 27)$.

21. Пусть $X \sim N(\mu, \sigma^2)$, где $\mu = 10$, $\sigma = 6$. Найти[1] $P(X > 5)$, $P(4 < X < 16)$, $P(X < 8)$, $P(X < 20)$, $P(X > 16)$.

Ответ: 0.7977; 0.6827; 0.3695; 0.9522; 0.1587

[1]К этой и следующим задачам нужна таблица функции распределения $\Phi(x)$ стандартной нормальной величины $Z \sim N(0, 1)$, см. рис. **??**

22. Пусть $X \sim N(12, 4^2)$. Найти c, такое, что $P(X > c) = 0.1$.

    Ответ: 14.56

23. Толщина (в см) детали, выходящей из-под пресса, нормально распределена со средним 0.9 и с.к.о. 0.003. Стандарт на деталь есть 0.900 ± 0.005. Каков процент брака? Найти максимально возможное с.к.о., при котором процент брака составляет менее 1%.

    Ответ: 9.5; 0.0019

24. Пусть $X \sim N(\mu, \sigma^2)$, где $\mu = 1$, $\sigma = 2$. Найти среднее и дисперсию с.в. $Y = 3X + 1$. По какому закону распределена Y?

    Ответ: 4; 36

25. 12% населения левши[2]. Найти приближенно вероятность, что в коллективе из 200 человек по крайней мере двадцать – левши.

    Ответ: 0.8363

26. Модель динамики цены акции предполагает, что если сейчас акция стоит S, то через определенный фиксированный период времени (минута, 5 минут, день...) она будет стоить $u \cdot S$ с вероятностью p и $d \cdot S$ с вероятностью $1 - p$. Предполагая, что движения цены в разные периоды независимы, приближенно найти вероятность, что акция будет стоить дороже по меньшей мере на 30% через 1000 периодов, если $u = 1.012$, $d = 0.990$, $p = 0.52$.

    Ответ: 0.9993

27. Время работы микрочипа нормально распределено со средним $\mu = 1.4 \times 10^6$ часов и $\sigma = 3 \times 10^5$ часов. Чему приближенно равна вероятность, что в партии из 100 чипов по крайней мере 80 отработают меньше 1.8×10^6 часов?

28. Три шара изымаются из ящика (без возвращения), в котором находится 5 белых и 8 красных шаров. Пусть $X_i = 1$, если i-й взятый шар белый и $X_i = 0$ в противном случае. Найти совместный закон распределения величин (а) X_1, X_2 (б) X_1, X_2, X_3.

    Ответ: (а) 14/39; 10/39; 10/39; 5/39 (б) 84; 70; 70; 70; 40; 40; 40; 15 все числа поделены на 429.

29. Совместная плотность с.в. X и Y равна

$$f(x, y) = c(y^2 - x^2)e^{-y}, \quad -y \le x \le y,\, 0 < y < \infty$$

    Найти (а) константу c; (б) одномерные плотности $f_X(x)$, $f_Y(y)$; (в) E(X); (г) условную плотность с.в. X при заданном $Y = y$; (д) независимы ли с.в. X и Y? Учесть результат (г); (е) найти условное среднее и дисперсию $E(X|Y = y)$, $\mathrm{Var}(X|Y = y)$, сравнить $E_{X|Y=y}$ с $E(X)$.

    Ответ: $c = 1/8$; $E[X] = 0$; $f_{X|Y}(x|y) = \frac{3}{4y}(1 - x^2/y^2)$

[2] Задачи 25-27 на применение формулы Муавра-Лапласа. Проверьте в каждом случае, выполняется ли правило $np(1-p) > 5$ применимости этого приближения.

30. Совместная плотность с.в. X и Y равна:

$$f(x,y) = \frac{6}{7}(x^2 + xy/2), \quad 0 < x < 1,\, 0 < y < 2$$

(а) Показать, что эта функция действительно может быть совместной плотностью вероятности; (б) найти одномерную плотность с.в. X; (в) $P(X > Y)$; (г) $E(X)$, $\mathrm{Var}(X)$, $E(Y)$; (д) независимы ли с.в. X,Y?

Ответ: $(12x^2 + 6x)/7$; $15/56$; $5/7$; $23/490$; $8/7$

31. Совместная плотность с.в. X и Y равна:

$$f(x,y) = \begin{cases} xe^{-x-y}, & x > 0, y > 0 \\ 0, & \text{иначе} \end{cases}$$

Независимы ли с.в. X,Y? Тот же вопрос для совместной плотности

$$f(x,y) = \begin{cases} 2, & y > x > 0,\ 1 > y > 0 \\ 0, & \text{иначе} \end{cases}$$

32. С.в. A, B, C распределены равномерно на отрезке $[0,1]$ независимо друг от друга. Найти совместную функцию распределения (не плотность!).

Ответ: $F(x,y,z) = xyz$ при $(x,y,z) \in [0,1]^3$

33. Двумерная с.в. распределена равномерно на треугольнике $x > 0$, $y > 0$ $2x + y < 4$. Найти $E(X)$.

34. (сумма независимых нормальных с.в.) Недельные продажи небольшой компании распределены нормально со средним \$2200 и с.к.о. \$230. В предположении независимости продаж в каждую неделю найти вероятность, что общая сумма продаж за две последующие недели превысит \$5000.

Ответ: 0.0326

35. По статистике 25.2% мужчин и 23.6% женщин не завтракают. Случайным образом выбрано 200 мужчин и 200 женщин. Найти приближенно[3] вероятность, что по крайней мере 110 из этих 400 человек не едят завтрак. Также найти приближенно вероятность, что в указанной выборке число незавтракающих женщин не меньше, чем число незавтракающих мужчин.

Ответ: 0.0829; 0.3766

36. Совместная плотность задана выражением:

$$f_{X,Y}(x,y) = \begin{cases} 1/y, & 0 < y < 1,\ 0 < x < y \\ 0, & \text{иначе} \end{cases}$$

Найти $E(XY)$,$E(X)$,$E(Y)$. Выполняется ли $E(XY) = E(X)E(Y)$? Почему?

Ответ: $1/6$; $1/4$; $1/2$

[3] т.е. через т. Муавра-Лапласа

37. Если X, Y – независимо и одинаково распределенные с.в. со средним μ и дисперсией σ^2, найти $E[(X-Y)^2]$.

    Ответ: $2\sigma^2$

38. Пусть X, Y – одинаково, но не обязательно независимо распределенные величины. Найти $\mathrm{Cov}(X+Y, X-Y)$.

    Ответ: 0

39. Совместная плотность задана выражением:

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-2x}/x, & 0 < x < \infty,\ 0 < y < x \\ 0, & \text{иначе} \end{cases}$$

    Найти $\mathrm{Cov}(X, Y)$.

    Ответ: 1/8

40. Совместная плотность задана выражением:

$$f_{X,Y}(x,y) = \begin{cases} e^{-(y+x/y)}/y, & x > 0, y > 0 \\ 0, & \text{иначе} \end{cases}$$

    Найти $\mathrm{Cov}(X, Y)$.

    Ответ: 1

41. Ковариационная матрица для трех с.в. X_i, $i = 1, 2, 3$, имеет вид $V = \begin{pmatrix} 10 & 6 & -1 \\ 6 & 6 & 2 \\ -1 & 2 & 7 \end{pmatrix}$.

    Найти дисперсию с.в. $X_1 - 2X_2 + 3X_2$.

    Ответ: $\vec{w}^T V \vec{w}$, $\vec{w} = (1, -2, 3)^T$

42. Пусть X_i, $i = 1, 2, 3, 4$ – с.в. с нулевой попарной корреляцией, каждая со нулевым средним и единичной дисперсией. Найти коэффициент корреляции $\rho(X_1+X_2, X_2+X_3)$, $\rho(X_1 + X_2, X_3 + X_4)$.

    Ответ: 1/2; 0

43. Пусть $Y = aX + b$, где a,b – константы. Найти коэф-т корреляции $\rho(X, Y)$.

    Ответ: $+1$, если $a > 0$, -1, если $a < 0$

44. Пусть $Z \sim N(0, 1)$, а $Y = a + bZ + cZ^2$. Найти $\rho(Y, Z)$.

    Ответ: $\frac{b}{\sqrt{b^2+2c^2}}$

45. Найти $P(X - Y > 0)$, если $X \sim N(2, 3^2)$,$Y \sim N(1, 2^2)$

46. Независимые с.в. $X_i \sim \mathrm{Unif}[-1, 1]$, $i = 1, 2, 3, 4$. Найти $E((X_1 + X_2 + X_3 + X_4)^2)$

47. [4]Студент в среднем сдает экзамен на 75 баллов из 100. Оценить сверху вероятность, что оценка студента превысит 85 баллов. Далее предположим, что также известна дисперсия баллов студента, она равна 25. Оценить вероятность, что студент сдаст на балл между 65 и 85 (использовать нер-во Чебышева).

    Ответ: $\leq 15/17$; $\geq 3/4$

48. 50 чисел округляются к ближайшему целому и затем суммируются. Если ошибки округления равномерно распределены в интервале $(-0.5, 0.5)$, приближенно найти вероятность, что указанная сумма отличается от точной больше, чем на три.

    Ответ: 0.1416

49. Кость последовательно подбрасывается до тех пор, пока сумма очков не превысит 300. Приближенно найти вероятность, что потребуется по меньшей мере 80 бросков.

    Ответ: 0.9431

50. Имеется 100 ламп со временем работы, распределенным по экспоненциальному закону со средним 5 часов. Лампы используются последовательно, сгоревшая сразу же заменяется новой. Приближенно найти вероятность, что свет еще будет гореть после 525 часов.

    Ответ: 0.3085

51. Некоторый критический для системы компонент мгновенно заменяется новым при выходе из строя. Среднее время работы этого компонента равно 100 часов, с.к.о. = 30 часов. Сколько таких устройств надо иметь в запасе, чтобы вероятность бесперебойной работы системы на протяжении 2000 часов была не менее 95%?

    Ответ: ≥ 23

52. §1 задачи 3,4,5,15,18(а,б),19,20(а,б),22

53. §2 задачи 12

54. §3 задачи 1,3,6(а),8,9,24

55. §4 задачи 1,2,3,4,5,6,17,20,28

56. §5 задачи 1,2,6

57. §6 задачи 1(в среднеквадратичном смысле),2,6,8,10,11,19,23,29

58. §8 задачи 1,5,6,9(а)

59. Пусть $X_i \sim N(0,3)$, $i = 1,2,3,4$ и все X_i независимы. Найти такое k, что[5]:

    (а) $P(X_1^2 + X_2^2 + X_3^2 + X_4^2 < k) = 0.99$ (χ^2)

    (б) $P(X_1 < k\sqrt{X_2^2 + X_3^2}) = 0.1$ (Стьюдент)

[4]Задачи 47-51 на тему “вероятностные нер-ва и ЦПТ”. Их надо изучить повнимательнее.

[5]использовать, что: (i) если независимые $Z_i \sim N(0,1)$, $i = 1...n$, то сумма $Z_1^2 + Z_2^2 + ... + Z_n^2 \sim \chi_n^2$, т.е. следует распределению χ^2 с n степенями свободы; (ii) если $Z \sim N(0,1)$, $X \sim \chi_n^2$, то $T = \frac{Z}{\sqrt{X/n}}$ следует распределению Стьюдента с n степенями свободы

60. Пусть независимые с.в. $X_i \sim N(0,1)$, $i = 1, ..., 5$. Найти такую константу c, что с.в. $\frac{c(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$ следует распределению Стьюдента.

    Ответ: $\sqrt{3/2}$

61. Взята выборка из 9 упаковок сахара из большой партии товара. Веса упаковок оказались равны 812.0, 786.7, 794.1, 791.6, 811.1, 797.4, 797.8, 800.8, 793.2 грамм. Предположить, что вес упаковки в ген. совокупности распределен нормально. (а) Построить руками 95% доверительный интервал для среднего веса упаковки в партии. (б) Переделать задачу, если известно с.к.о. в генеральной совокупности, $\sigma = 8.5$ г. (в) без вычислений сказать, что произойдет с доверительными интервалами из пп. а,б, если доверительную вероятность поднять до 99%. (г) понизить до 90%?

    Ответ: (791.74, 804.86); (792.75, 803.85);

62. На складе произведена выборка 50 ед. товара и произведена оценка их остаточной стоимости. Среднее по выборке оказалось равным 320.41, выборочное с.к.о. равно 40.60. (а) Найти 95% доверительный интервал для средней остаточной стоимости в генеральной совокупности (б) Полученный в (а) интервал кажется слишком широким: хочется уменьшить его ширину в два раза при неизменной доверительной вероятности. Какой *дополнительный* объем выборки потребуется для этого?

    Ответ: (309.16, 331.66); 150

63. В выборке из 200 студентов ун-та A 30 человек сказали, что любят предмет ТВиМС. Найти 95% доверительный интервал для доли студентов ун-та A, любящих статистику. Далее, в ун-те B в выборке из 20 студентов 8 человек сказали, что любят ТВиМС. Найти 95% доверительный интервал для доли фанатов ТВиМС в ун-те B. Далее, найти 95% доверительный интервал для разности долей любящих ТВиМС в двух ун-тах и сделать вывод. Наконец, изменить доверительную вероятность в последнем задании до 99% и сделать вывод. Ответ: $(0.100512425802026, 0.199487574197974)$; $(0.185292757457975, 0.614707242542025)$; $(0.029663393872012, 0.470336606127988)$; $(-0.040034920311331, 0.540034920311331)$

64. Взята выборка из 9 упаковок сахара из большой партии товара. Веса упаковок оказались равны 812.0, 786.7, 794.1, 791.6, 811.1, 797.4, 797.8, 800.8, 793.2 грамм. Предположить, что вес упаковки в ген. совокупности распределен нормально. Найти 90% доверительный интервал для с.к.о.

    Ответ: (6.127, 14.595)

65. Имеются две выборки размера $n_1 = 64$ и $n_2 = 36$ из двух нормально распределенных совокупностей с параметрами $\bar{x}_1 = 400$, $\sigma_1 = 20$ и $\bar{x}_2 = 360$, $\sigma_2 = 25$. Построить 90% доверительный интервал для разности средних в указанных генеральных совокупностях. Указать, можно ли с принятой доверительной вероятностью заключить, что среднее в одной ген. совокупности больше среднего значения в другой?

    Ответ: (32.031032410378089, 47.968967589621911)

66. Среднее время работы одного прибора в выборке из 100 приборов равно 1570 часов, с.к.о. известно и равно 120 часов. Протестировать гипотезу, что среднее время работы

во всей партии равно 1600 часов (двусторонний тест на 5% уровне значимости). Найти p-значение теста. Отвергается ли H_0 на 1% уровне значимости?

Ответ: z-статистика=-2.5; отвергнуть H_0; p-value=0.0124; нет

67. Выборка из семи аккумуляторов взята из большой партии данного товара. Напряжение в этих семи приборах равно 12.9, 11.6, 13.5, 13.9, 12.1, 11.9, 13.0 В при номинале в 12 В. Протестировать гипотезу, что среднее напряжение в партии равно 12 В, на 5% уровне значимости. Также протестировать гипотезу, что среднее напряжение меньше 12 В.

    Ответ: t-статистика=2.16, не отвергаем H_0; не отвергаем H_0.

68. Исследуется влияние инноваций в инвестициях в банках. Утверждается, что средняя доходность на инвестиции всех банков в некоторой стране равна 10.2% годовых. Имеется выборка доходностей 26 банков, которые используют особо продвинутые стратегии инвестирования (в %): 10.00, 11.90, 9.90, 10.09, 10.31, 9.96, 10.34, 10.30, 10.50, 10.23, 10.72, 11.54, 10.81, 10.15, 9.04, 11.55, 10.81, 8.69, 10.74, 10.31, 10.76, 10.92, 11.26, 11.21, 10.20, 10.76. Известное генеральное с.к.о. равно 0.7138. Подтверждают ли эти данные повышение доходности для банков-инноваторов? *Указание*: в качестве альтернативной гипотезы принять $H_1 : \mu > 10.2\%$ (почему?) Рассчитать p-значение теста.

    Ответ: на 5% уровне значимости H_0 отвергаем; p-value ≈ 0.02

69. 68 из 150 клиентов сказали, что им понравился новый продукт. Альтернативные масштабные исследования говорят, что 40% клиентов полюбили данный продукт. Протестировать двусторонню гипотезу о доле на 5% уровне значимости. Найти p-значение теста.

    Ответ: не можем отвергнуть H_0 на 5% уровне значимости

70. В выборке[6] из 200 студентов ун-та А 30 человек сказали, что любят предмет ТВиМС. В ун-те В в выборке из 20 студентов 8 человек сказали, что любят ТВиМС. Проверить двустороннюю гипотезу H_0 о равенстве долей любящих ТВиМС в указанных ун-тах на 5% и на 1% уровнях значимости.

    Ответ: отврегаем H_0 для 5%, но не для 1%

71. Имеются две выборки размера $n_1 = 64$ и $n_2 = 36$ из двух нормально распределенных совокупностей с параметрами $\bar{x}_1 = 400$, $\sigma_1 = 20$ и $\bar{x}_2 = 360$, $\sigma_2 = 25$. Протестировать двустороннюю гипотезу о равенстве средних в генеральных совокупостях на 10% уровне значимости.

    Ответ: отвергаем H_0

72. Предполагается, что машина нарезает провод на куски в 22 см. С.к.о. длины куска составляет 0.08 см. Взята выборка из 50 нарезанных кусков. Согласно протоколу машина работает правильно, т.е. режет на куски 22 см (нулевая гипотеза), если выборочное среднее лежит в интервале (21.97, 22.03) см. Иными словами, если выборочное среднее лежит в указанном интервале, нулевая гипотеза не отвергается, в противном случае – отвергается. Вопросы: (а) найти вероятность ошибки первого рода; (б) найти вероятность ошибки второго рода, если на самом деле машина режет на куски в 22.05 см; (в)

[6] В задачах 70,71 обратить внимание на совпадение выводов с задачами 63,65. Доверительные интервалы и двусторонние гипотезы – это, грубо говоря, одно и то же.

найти вероятность отвергнуть нулевую гипотезу, если на самом деле машина режет на куски 22.01 см (мощность теста).

Ответ: $\alpha = 0.008$; $\beta = 0.0384$; 0.0386

73. Производитель утверждает, что леска выдерживает нагрузку в 7 кг, с.к.о. известно и равно 0.25 кг. Рассматривается тест $H_0 : \mu = 7$ vs. $H_1 : \mu < 7$. Найти размер выборки, такой, что мощность теста равна 0.9, если истинная разрывная нагрузка равна 6.95 кг.

    Ответ: $n > 214$

74. Пусть с.в. X распределена экспоненциально с параметром β. Тестируется гипотеза $H_0 : \beta \geq 1$ против гипотезы $H_1 : \beta < 1$. Тест устроен так, что H_0 отвергается, если $X \geq 1$. Найти мощность теста как функцию β.

    Ответ: $e^{-\beta}$

6 Процедура экзамена

- Иметь при себе калькулятор
- Стат. таблицы будут предоставлены (такие же, как на Smart LMS). Обратить внимание и уметь пользоваться таблицами 5 и 10.
- Работа выполняется на скрепленных листах A4 (будут предоставлены). Раскреплять листы запрещено.
- Разговоры и общение между студентами недопустимы: предупреждение, затем удаление.
- За списывание удаление
- В течение экзамена выходить из аудитории запрещено
- Рассадка в аудитории может быть не произвольной.

Список литературы

[1] Д. А. Коршунов, Н. И. Чернова *Сборник задач и упражнений по математической статистике* (2004)