1. Теоретический минимум

1. Дайте определение нормально распределённой случайной величины. Укажите диапазон возможных значений, функцию плотности, ожидание, дисперсию. Нарисуйте функцию плотности.
2. Дайте определение хи-квадрат распределения. Укажите диапазон возможных значений, выражение через нормальные распределения, математическое ожидание. Нарисуйте функцию плотности при разных степенях свободы.
3. Дайте определение распределения Стьюдента. Укажите диапазон возможных значений, выражение через нормальные распределения. Нарисуйте функцию плотности распределения Стьюдента при разных степенях свободы на фоне нормальной стандартной функции плотности.
4. Дайте определение распределения Фишера. Укажите диапазон возможных значений, выражение через нормальные распределенеия. Нарисуйте возможную функцию плотности.

Для следующего блока вопросов предполагается, что имеется случайная выборка X_1, X_2, ..., X_n из распределения с функцией плотности $f(x, \theta)$, зависящей от от параметра θ. Дайте определение каждого понятия из списка или сформулируйте соответствующую теорему:

5. Выборочное среднее и выборочная дисперсия;
6. Формула несмещённой оценки дисперсии;
7. Выборочный начальный момент порядка k;
8. Выборочный центральный момент порядка k;
9. Выборочная функция распределения;
10. Несмещённая оценка $\hat{\theta}$ параметра θ;
11. Состоятельная последовательность оценок $\hat{\theta}_n$;
12. Эффективность оценки $\hat{\theta}$ среди множества оценок $\hat{\Theta}$;
13. Неравенство Крамера–Рао для несмещённых оценок;
14. Функция правдоподобия и логарифмическая функция правдоподобия;
15. Информация Фишера о параметре θ, содержащаяся в одном наблюдении;
16. Оценка метода моментов параметра θ при использовании первого момента, если $\mathrm{E}(X_i) = g(\theta)$ и существует обратная функция g^{-1};
17. Оценка метода максимального правдоподобия параметра θ;

Для следующего блока вопросов предполагается, что величины $X_1, X_2, ..., X_n$ независимы и нормальны $\mathcal{N}(\mu; \sigma^2)$.

18. Укажите закон распределения выборочного среднего, величины $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, величины $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$, величины $\frac{\hat{\sigma}^2(n-1)}{\sigma^2}$;
19. Укажите формулу доверительного интервала с уровнем доверия $(1-\alpha)$ для μ при известной дисперсии, для μ при неизвестной дисперсии, для σ^2;

2. Задачный минимум

1. Рост в сантиметрах (случайная величина X) и вес в килограммах (случайная величина Y) взрослого мужчины является нормальным случайным вектором $Z = (X, Y)$ с математическим ожиданием $\mathrm{E}(Z) = (175, 74)$ и ковариационной матрицей

$$\mathrm{Var}(Z) = \begin{pmatrix} 49 & 28 \\ 28 & 36 \end{pmatrix}$$

Лишний вес характеризуется случайной величиной $U = X - Y$. Считается, что человек страдает избыточным весом, если $U < 90$.

   а) Определите вероятность того, что рост мужчины отклоняется от среднего более, чем на 10 см.

   б) Укажите распределение случайной величины U. Выпишите её плотность распределения.

   в) Найдите вероятность того, что случайно выбранный мужчина страдает избыточным весом.

2. Рост в сантиметрах, случайная величина X, и вес в килограммах, случайная величина Y, взрослого мужчины является нормальным случайным вектором $Z = (X, Y)$ с математическим ожиданием $\mathrm{E}(Z) = (175, 74)$ и ковариационной матрицей

$$\mathrm{Var}(Z) = \begin{pmatrix} 49 & 28 \\ 28 & 36 \end{pmatrix}$$

   а) Найдите средний вес мужчины при условии, что его рост составляет 170 см.

   б) Выпишите условную плотность распределения веса мужчины при условии, что его рост составляет 170 см.

   в) Найдите условную вероятность того, что человек будет иметь вес, больший 90 кг, при условии, что его рост составляет 170 см.

3. Для реализации случайной выборки $x = (1, 0, -1, 1)$ найдите:

   а) выборочное среднее,

   б) неисправленную выборочную дисперсию,

   в) исправленную выборочную дисперсию,

   г) выборочный второй начальный момент,

   д) выборочный третий центральный момент,

4. Для реализации случайной выборки $x = (1, 0, -1, 1)$ найдите:

   а) вариационный ряд,

   б) первый член вариационного ряда,

   в) последний член вариационного ряда,

   г) график выборочной функции распределения.

5. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из дискретного распределения, заданного с помощью таблицы

x	-3	0	2
$\mathbb{P}(X_i = x)$	$2/3 - \theta$	$1/3$	θ

Рассмотрите оценку $\hat{\theta} = \dfrac{\bar{X} + 2}{5}$.

а) Найдите $\mathrm{E}[\hat{\theta}]$.

б) Является ли оценка $\hat{\theta}$ несмещенной оценкой неизвестного параметра θ?

6. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения с плотностью распределения

$$f(x, \theta) = \begin{cases} \dfrac{6x(\theta - x)}{\theta^3} & \text{при } x \in [0; \theta], \\ 0 & \text{при } x \notin [0; \theta], \end{cases}$$

где $\theta > 0$ — неизвестный параметр распределения и $\hat{\theta} = \bar{X}$.

а) Является ли оценка $\hat{\theta} = \bar{X}$ несмещенной оценкой неизвестного параметра θ?

б) Подберите константу c так, чтобы оценка $\tilde{\theta} = c\bar{X}$ оказалась несмещенной оценкой неизвестного параметра θ.

7. Пусть $X = (X_1, X_2, X_3)$ — случайная выборка из распределения Бернулли с неизвестным параметром $p \in (0, 1)$. Какие из следующих ниже оценкой являются несмещенными? Среди перечисленных ниже оценок найдите наиболее эффективную оценку:

- $\hat{p}_1 = \dfrac{X_1 + X_3}{2}$,
- $\hat{p}_2 = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3$,
- $\hat{p}_3 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$.

8. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения с плотностью

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}\, e^{-\frac{x}{\theta}} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где $\theta > 0$ — неизвестный параметр. Является ли оценка $\hat{\theta}_n = \dfrac{X_1 + \ldots + X_n}{n+1}$ состоятельной?

9. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения с плотностью распределения

$$f(x, \theta) = \begin{cases} \dfrac{6x(\theta - x)}{\theta^3} & \text{при } x \in [0; \theta], \\ 0 & \text{при } x \notin [0; \theta], \end{cases}$$

где $\theta > 0$ — неизвестный параметр распределения. Является ли оценка $\hat{\theta}_n = \frac{2n+1}{n}\bar{X}_n$ состоятельной оценкой неизвестного параметра θ?

10. Пусть $X = (X_1, \dots, X_n)$ — случайная выборка из распределения с плотностью распределения

$$f(x,\theta) = \begin{cases} \dfrac{6x(\theta - x)}{\theta^3} & \text{при } x \in [0;\theta], \\ 0 & \text{при } x \notin [0;\theta], \end{cases}$$

где $\theta > 0$ — неизвестный параметр распределения. Используя центральный момент 2-го порядка, при помощи метода моментов найдите оценку для неизвестного параметра θ.

11. Пусть $X = (X_1, \dots, X_n)$ — случайная выборка. Случайные величины $X_1, \dots, X_n$ имеют дискретное распределение, которое задано при помощи таблицы

x	-3	0	2
$\mathbb{P}(X_i = x)$	$2/3 - \theta$	$1/3$	θ

Используя второй начальный момент, при помощи метода моментов найдите оценку неизвестного параметра θ. Для реализации случайной выборки $x = (0, 0, -3, 0, 2)$ найдите числовое значение найденной оценки параметра θ.

12. Пусть $X = (X_1, \dots, X_n)$ — случайная выборка из распределения с плотностью распределения

$$f(x,\theta) = \begin{cases} \frac{2x}{\theta}\, e^{-\frac{x^2}{\theta}} & \text{при } x > 0, \\ 0 & \text{при } x \le 0, \end{cases}$$

где $\theta > 0$. При помощи метода максимального правдоподобия найдите оценку неизвестного параметра θ.

13. Пусть $X = (X_1, \dots, X_n)$ – случайная выборка из распределения Бернулли с параметром $\mathbb{P} \in (0;1)$. При помощи метода максимального правдоподобия найдите оценку неизвестного параметра $\mathbb{P}$.

14. Пусть $X = (X_1, \dots, X_n)$ — случайная выборка из распределения с плотностью

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}\, e^{-\frac{x}{\theta}} & \text{при } x \ge 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где $\theta > 0$ — неизвестный параметр. Является ли оценка $\hat{\theta} = \bar{X}$ эффективной?

15. Стоимость выборочного исследования генеральной совокупности, состоящей из трех страт, определяется по формуле $TC = c_1 n_1 + c_2 n_2 + c_3 n_3$, где c_i — цена одного наблюдения в i-ой страте, а n_i — число наблюдений, которые приходятся на i-ую страту. Найдите n_1, n_2 и n_3, при которых дисперсия стратифицированного среднего достигает наименьшего значения, если бюджет исследования 8000 и имеется следующая информация:

Страта	1	2	3
Среднее значение	30	40	50
Стандартная ошибка	5	10	20
Вес	25%	25%	50%
Цена наблюдения	1	5	10

Ответы:

1. а) ≈ 0.15
   б) $U \sim \mathcal{N}(101, 29)$, $f(u) = \frac{1}{\sqrt{2\pi \cdot 29}} e^{-\frac{1}{2}\frac{(u-101)^2}{29}}$
   в) ≈ 0.02
2. а) 71.14
   б) $f(y|x = 170) = \frac{1}{\sqrt{2\pi \cdot 20}} e^{-\frac{1}{2}\frac{(y-71.14)^2}{20}}$
   в) ≈ 0
3. а) 0.25
   б) 0.6875
   в) $0.91(6)$
   г) 0.75
   д) -0.28125
4. а) $-1, 0, 1, 1$
   б) -1
   в) 1
   г) $f(x) = \begin{cases} 0, & x < -1 \\ 0.25, & -1 \leq x < 0 \\ 0.5, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$
5. а) θ
   б) да
6. а) нет, оценка смещена
   б) $c = 2$
7. а) все оценки несмещенные
   б) $\hat{p}_3$ наиболее эффективная
8. да
9. да
10. $\hat{\theta}_{MM} = \sqrt{\frac{\sum_{i=1}^{n}(X - X_i)^2 \cdot 20}{n}}$
11. $\hat{\theta}_{MM} = \frac{1}{5}\left(6 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right)$, $\hat{\theta}_{MM} = 0.68$
12. $\hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} x_i^2}{n}$
13. $\hat{p}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n}$
14. да
15. $n_1 \approx 260$, $n_2 \approx 232$, $n_3 \approx 658$