

Школа лингвистики, 2023-24 уч. год**Теория вероятностей****Задачи по статистике. Точечные оценки. (12.04.2024)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Задача 1. Пусть дана выборка $\{x_1, \dots, x_n\}$ и известно что это результаты независимых экспериментов, а соответствующая случайная величина ξ распределена с параметрами $E\xi = \mu$ и $D\xi = d$. Проверьте несмещённость и состоятельность следующих оценок для μ :

(a) $\hat{\mu}_1(x_1, \dots, x_n) = \frac{2x_1 + 3x_2 + x_3}{6};$

(b) $\hat{\mu}_2(x_1, \dots, x_n) = 0.21x_1 + 0.38x_2 + 0.7x_3;$

(c) $\hat{\mu}_3(x_1, \dots, x_n) = \frac{1}{\lfloor \frac{n}{3} \rfloor} \sum_{k=1}^{\lfloor \frac{n}{3} \rfloor} x_k;$

Задача 2. Пусть дана конкретная выборка $\{2, 4, 1, 3\}$ для ситуации из предыдущей задачи. Вычислите несмещённую оценку для d .

Задача 3. Зная как выражаются математическое ожидание и дисперсия через параметры, предложите оценки для параметров следующих распределений:

- (a) Биномиальное;
- (b) распределение Пуассона;
- (c) Показательное;
- (d) Равномерное на отрезке.

Задача 4. В первой четверти XXI века возник феномен “твиттеризации мышления”, когда люди перестают не только писать, но и воспринимать длинные тексты. Для исследования этого факта выбираются законченные куски текстов, написанных пользователями в Интернете и исследуется их длина. Для простоты будем предполагать, что распределение длины текста описывается равномерным распределением на отрезке $[0; a]$ ¹. Пусть дана выборка $\{x_1, \dots, x_n\}$ из такой случайной величины.

- (a) Какой будет оценка для параметра a исходя из несмещённой оценки для математического ожидания?;
- (b) предложите другую естественную оценку для этого параметра;
- (c) является ли оценка из предыдущего пункта несмещённой? (Подсказка: здесь необходимо воспользоваться тем, что вы знаете как теоретически распределена искомая случайная величина, явно записать функцию распределения для случайной величины, получаемой из оценки и вычислить необходимые числовые характеристики). Если нет, как её изменить, чтобы она стала несмещённой?;
- (d) А состоятельной?;
- (e) А какая оценка для параметра a более эффективна - первая или вторая?

Дополнительные задачи

Задача 5. Предложите статистический способ оценки числа рыб в пруду.

¹Несмотря на то, что длина текста это целое число, бывает удобно описывать распределение равномерной случайной величиной (часто число символов делится на какой-то фактор, чтобы полученные значения были меньше и удобнее в работе). 0 символов может соответствовать сообщению, состоящему из знака препинания или смайлика ;-).