

## Практикум 5. Доверительные интервалы.

1. Пусть  $X_1, \dots, X_n \sim R[0, \theta]$ . Исследуем как устроены доверительные интервалы в этом случае.

1.1 Строя асимптотический интервал, мы можем рассматривать различные состоятельные оценки дисперсии. Сравним насколько изменится интервал при использовании оценки  $S^2$  или оценки, основанной на подстановке в дисперсию вместо  $\theta$  оценки  $(n+1)X_{(n)}/n$ .

- Какая из оценок дисперсии точнее и насколько?
- Построить два асимптотических доверительных интервала, используя  $\bar{X}$  и оценивая дисперсию с помощью двух указанных оценок.
- Найти эмпирические доверительные вероятности интервалов (построить 1000 выборок, подсчитать долю тех, для которых интервал накрыл истинное значение параметра), сравнить между собой и с уровнем  $1 - \alpha$  (для разных  $n$ ,  $\alpha = 0.05, 0.01, 0.001$ ).

1.2 Для данной модели можно построить также точный интервал

$$(X_{(n)}, X_{(n)}/\sqrt[n]{1-\alpha})$$

- Моделировать выборку и построить точный доверительный интервал с помощью достаточной статистики.
  - Сравнить длины всех трех интервалов: моделировать 1000 выборок, для каждой найти длины доверительных интервалов и построить гистограммы их распределений для разных  $n$ ,  $\alpha = 0.05, 0.01, 0.001$
  - Сравнить средние длины интервалов в предыдущей задаче
- 1.3 Выберите из исследованных методов самый эффективный и примените его к следующей задаче. В файле *unif.txt* содержится информация о 150 выдаче датчика случайных чисел из некоторого отрезка  $[0, a]$ . Локализируйте  $a$  так, чтобы вероятность вашей ошибки была не более 1%.
- 1.4 \* Выборка из файла *Beta.txt* имеет бета-распределение.
- Оцените его параметры методом моментов;
  - Постройте для них 95% доверительный эллипс.

2. Пусть  $X_1, \dots, X_n \sim Poiss(\theta)$ . В дискретном случае построение точных интервалов достаточно громоздко.

2.1 Рассмотрим асимптотические доверительные интервалы, построенные на основе достаточной статистики методами подстановки оценки и стабилизации дисперсии.

- Построить указанные интервалы и изобразите их на прямой.
- Найти эмпирические доверительные вероятности этих интервалов: построить 1000 выборок, подсчитать долю тех, для которых интервал накрыл истинное значение параметра). Рассмотреть  $n=20, 50, 100$ .

2.2 Сравнить длины интервалов. Для этого моделировать 1000 выборок, для каждой найти длины двух доверительных интервалов.

- Построить гистограммы и сравнить распределения длин интервалов.
- Сравнить средние длины интервалов в обоих случаях.

2.3 В файле *poisson.txt* приведены результаты проявления редкой болезни в 115 городах примерно одинакового размера. В силу теоремы Пуассона данные хорошо описываются пуассоновским распределением. Постройте доверительный интервал для среднего числа больных в городе такого размера, выбрав более удачный из исследованных выше методов.

2.4 \* В файле *gamma.txt* приведены наблюдения за временем до появления болезни в каждом из городов, имеющим  $\text{Gamma}(\theta_1, \theta_2)$ .

- Построить для параметров ОММ.
- Построить на его основе асимптотическое доверительное множество.
- Согласно теории указанное время должно иметь  $\exp(1/3)$  распределение. Можем ли мы наверняка утверждать, что это не так?

3. Пусть  $X_1, \dots, X_n \sim \text{Geom}(\theta)$ . В дискретном случае построение точных доверительных интервалов достаточно громоздко.
- 3.1 Рассмотрим асимптотические доверительные интервалы, построенные на основе достаточной статистики методами подстановки оценки и стабилизации дисперсии.
- Построить указанные интервалы и изобразить их на прямой.
  - Найти эмпирические доверительные вероятности этих интервалов: построить 1000 выборок, подсчитать долю тех, для которых интервал накрыл истинное значение параметра). Рассмотреть  $n=20, 50, 100$ .
- 3.2 Сравнить длины интервалов. Для этого моделировать 1000 выборок, для каждой найти длины двух доверительных интервалов.
- Построить гистограммы и сравнить распределения длин интервалов.
  - Сравнить средние длины интервалов в обоих случаях.
- 3.3 В файле Inspection Data.txt приведены данные о пожарных инспекциях на различных объектах. Указан номер инспекции, после которой впервые были обнаружены нарушения. Построить доверительный интервал для вероятности нарушения, считая, что инспекции независимы и обнаруживают нарушение с вероятностью  $p$ . Используйте ваши знания из пунктов 1-2.
- 3.4 \* Сдающие тест по практикуму Александра Викторовича на мехмате правильно отвечают в тесте на 7 и более вопросов с вероятностью  $p_1$  (назовем их отличниками), на 5-6 вопросов с вероятностью  $p_2$  (хорошисты) и на меньшее число вопросов с вероятностью  $p_3 = 1 - p_1 - p_2$  (двоечники). Построить 95% асимптотический доверительный эллипс для параметров  $p_1, p_2$  — долей хорошистов и отличников на мехмате, используя данные из Test.csv. В качестве оценки использовать ОМП  $\nu_1/n, \nu_2/n$ , где  $\nu_i$  — количества отличников и хорошистов в приведенной группе.