Нейман, Пирсон и Бернулли

Опубликовал
sobodv
Автор или источник
sobopedia
Предмет
Математическая Статистика (/Subjects/Details?id=5)
Тема
Теория проверки статистических гипотез (/Topics/Details?id=35)
Раздел
Лемма Неймана-Пирсона (/SubTopics/Details?id=127)
Дата публикации
13.02.2022
Дата последней правки
14.02.2022
Последний вносивший правки
sobodv
Рейтинг
★★☆

Условие

По выборке из распределения Бернулли, включающей два наблюдения $X = (X_1, X_2)$, тестируется гипотеза $H_0 : p = 0.2$ против альтернативы $H_1 : p = 0.6$ с использованием теста, полученного с помощью леммы Неймана-Пирсона.

1. Найдите распределение тестовой статистики.

2. Начиная с данного пункта предположим, что объем выборки возрос до 100 наблюдений $X = (X_1, \dots, X_{100})$. Найдите асимптотическое распределение тестовой статистики при условии верной нулевой гипотезы.

3. Используя асимптотическое распределение тестовой статистики запишите критическую область теста при уровне значимости, равном 5%.

4. Рассчитайте мощность теста при 5%-м уровне значимости. Для этого рассмотрите фиксированный объем выборки $n = 100$ предполагая, что распределение тестовой статистики совпадает с асимптотическим распределением.

5. Опишите, как изменится тестовая статистика и критическая область при изменении альтернативной гипотезы.

Решение

1. Тестовая статистика имеет вид:

$$T(X) = \frac{0.6^{X_1+X_2}(1-0.6)^{2-X_1-X_2}}{0.2^{X_1+X_2}(1-0.2)^{2-X_1-X_2}}$$

Логарифмируя получаем:

$$T_2(X) = \ln(T(X)) = (X_1 + X_2)\ln(0.6) + (2 - X_1 - X_2)\ln(0.4) - (X_1 + X_2)\ln(0.2) - (2 - X_1 - X_2)\ln(0.8) \approx$$
$$\approx 1.8(X_1 + X_2) - 1.4$$

За счет положительного линейного преобразования получаем:

$$T_3(X) = X_1 + X_2 \sim B(2, p)$$

2. Действуя по аналогии получаем:

$$T(X) = X_1 + \ldots + X_{100} \sim B(100, p)$$

Совершая положительные линейные преобразования над тестовой статистикой и применяя ЦПТ получаем:

$$T_2(X) = \frac{\overline{X}_n - 0.2}{\sqrt{\frac{0.2(1-0.2)}{n}}}, \qquad T_2(X)|H_0 \xrightarrow{d} \mathcal{N}(0, 1)$$

В результате с небольшой погрешностью мы можем предположить:

$$T_2(X) = \frac{\overline{X}_n - 0.2}{\sqrt{\frac{0.2(1-0.2)}{100}}}, \qquad T_2(X)|H_0 \dot{\sim} \mathcal{N}(0, 1)$$

3. В результате критическая область принимает вид $\mathcal{T}_{0.05} = (1.645, \infty)$.

4. Обратим внимание, что:

$$\overline{X}_n \dot{\sim} \mathcal{N}\left(100p, \frac{p(1-p)}{100}\right)$$

Отсюда следует, что:

$$\overline{X}_n|H_1 \dot{\sim} \mathcal{N}\left(0.6, \frac{0.6(1-0.6)}{100}\right) = \mathcal{N}(0.6, 0.0024)$$

Пользуясь свойствами нормального распределения получаем:

$$T_2(X)|H_1 = \frac{\overline{X}_n - 0.2}{\sqrt{\frac{0.2(1-0.2)}{100}}}|H_1 = \frac{\overline{X}_n - 0.2}{\sqrt{0.0016}}|H_1 \dot{\sim} \mathcal{N}\left(\frac{0.6 - 0.2}{\sqrt{0.0016}}, \frac{0.0016}{0.0024}\right) = \mathcal{N}(10, 1.5)$$

Используя полученный результат посчитаем вероятность ошибки второго рода:

$$\beta = P(T_2(X) < 1.645|H_1) = \Phi\left(\frac{1.645 - 10}{\sqrt{1.5}}\right) \approx 0$$

В результате мощность данного теста крайне велика, поскольку близка к 1.

5. Нетрудно показать, что независимо от вида альтернативной гипотезы $H_1 : p = p_1$, то есть при любом $p_1 \in (0, 1)$, тестовая статистика останется прежней. При этом если $p_1 > 0.6$, то критическая область окажется левосторонней. Полученный результат объясняет интуицию, исходя из которой тестирование при сложной альтернативной гипотезе осуществляется с помощью аналогичной тестовой статистики.

Показать решение