Самостійна робота №3
з дисципліни "асимптотична статистика"
Варіант №4

Горбунова Даніела Денисовича
4 курс бакалаврату
група "комп'ютерна статистика"

8 березня 2021 р.

1 Вступ.

У даній роботі побудовано тест для перевірки простих гіпотез для вибірки фіксованого обсягу. Для цього використана теорія побудови та аналізу тестів відношення вірогідності.

2 Хід роботи.

2.1 Обчислення інформації за Фішером.

Для подальших викладок нам знадобиться явний вигляд інформації за Фішером для одного спостереження $\xi \sim Bin(m, \theta)$, де $\theta \in (0, 1)$ - невідома ймовірність успіху. Запишемо функцію вірогідності:

$$f_\theta(\xi_1) = C_m^{\xi_1} \theta^{\xi_1} (1-\theta)^{m-\xi_1}$$
$$\ln\left(f_\theta(\xi_1)\right) = \ln C_m^{\xi_1} + \xi_1 \ln\theta + (m - \xi_1)\ln(1-\theta)$$

Тоді функція впливу запишеться так:

$$u_\theta(\xi_1) = \frac{\partial}{\partial\theta} f_\theta(\xi_1) = \frac{\xi_1}{\theta} - \frac{m-\xi_1}{1-\theta} = \frac{\xi_1}{\theta(1-\theta)} - \frac{m}{1-\theta}$$

Неважко побачити, що вона буде центрованою:

$$\mathbb{M}\left[u_\theta(\xi_1)\right] = \frac{\mathbb{M}\left[\xi_1\right]}{\theta(1-\theta)} - \frac{m}{1-\theta} = \frac{m}{1-\theta} - \frac{m}{1-\theta} = 0$$

Тому простіше обчислити інформацію за Фішером використовуючи означення:

$$i(\theta) = \mathbb{D}\left[u_\theta(\xi_1)\right] = \mathbb{M}\left[u_\theta^2(\xi_1)\right] = \mathbb{M}\left[\left(\frac{\xi_1}{\theta} - m\right)^2\right]\frac{1}{(1-\theta)^2} = \mathbb{M}\left[\frac{\xi_1^2}{\theta^2} - \frac{2m}{\theta}\xi_1 + m^2\right]\frac{1}{(1-\theta)^2} =$$
$$= \left(\frac{1}{\theta^2}(\theta m + \theta^2 m(m-1)) - m^2\right)\frac{1}{(1-\theta)^2} = \frac{m}{(1-\theta)\theta}$$

2.2 Статистичні гіпотези. Статистика тесту відношення вірогідності.

Розглядається задача перевірки двох простих гіпотез про значення невідомої ймовірності успіху в біноміальному розподілі $Bin(m, p)$, $m = 2$, при фіксованому обсягу вибірки $n = 65$:

$$H_0 : p = 0.5 =: p_0$$
$$H_1 : p = 0.6 =: p_1$$

Різниця ймовірностей для різних гіпотез є незначною. Непогано було б розглянути запропоновані вище гіпотези з точки зору альтернатив, що зближуються. У такому разі, маємо одну із зафіксованих альтернатив при n:

$$H_1 : \ p = 0.6 = 0.5 + 0.1 = 0.5 + \frac{v}{\sqrt{n}}, \ \frac{v}{\sqrt{n}} = 0.1,$$

звідки можемо знайти $v = 0.1\sqrt{n} = 0.8062258$. Надалі це значення будемо використовувати для обчислення таких наближених значень тесту як критичне значення та потужність.

Застосувавши результати попереднього розділу, статистика тесту відношення вірогідності така:

$$\ln LR(\overline{\xi}) = \sum_{j=1}^{n} \xi_j \ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) + mn \ln\left(\frac{1-p_1}{1-p_0}\right), \ \overline{\xi} = (\xi_1, \ldots, \xi_n)$$

2.3 Визначення критичного значення тесту. Оцінка потужності тесту.

Побудуємо тест відношення вірогідності:

$$\pi_\alpha(\overline{\xi}) = \begin{cases} 0, & \ln LR(\overline{\xi}) \leqslant C(\alpha) \\ 1, & \ln LR(\overline{\xi}) > C(\alpha) \end{cases}$$

Потрібно відшукати порогове значення $C(\alpha)$. Одним із можливих способів є застосування нормальної апроксимації відношення вірогідності:

$$C(\alpha) \approx -\frac{v^2}{2} i(p_0) + \lambda_\alpha v \sqrt{i(p_0)}, \ \lambda_\alpha = Q^{N(0,1)}(1-\alpha)$$
$$\Leftrightarrow C(\alpha) \approx 1.150843, \ \alpha := 0.05$$

2.4 Визначення мінімального обсягу вибірки для виконання певних умов.

$$R(t) = \int_{\mathbb{R}} e^{i\lambda t} F(d\lambda), \ t \in \mathbb{R}, \ F \text{ - скінченна міра на } \mathbb{R}$$