Математические основы анализа данных

Солодушкин Святослав Игоревич

Кафедра вычислительной математики и компьютерных наук,
УрФУ имени первого Президента России Б.Н. Ельцина

апрель 2023

Задачи

1. Найти методом наибольшего правдоподобия оценку параметра p биномиального распределения $P_n(k) = C_n^k p^k (1 - p)^{n-k}$, если в n_1 независимых испытаниях событие A появилось m_1 раз и в n_2 независимых испытаниях событие A появилось m_2 раз.
2. Используя метод наибольшего правдоподобия, оценить параметры a и σ^2 нормального распределения, если в результате n независимых испытаний случайная величина X приняла значения $x_1, x_2, \ldots, x_n$.
3. Пусть генеральная совокупность X равномерно распределена на отрезке $[a, b]$. Используя метод наибольшего правдоподобия, оценить параметры a и b.

Задачи более сложные

5. Всегда ли оценка ММП, если она существует, единственна? Пусть генеральная совокупность X равномерно распределена на отрезке $[0, \theta]$, где $\theta > 0$. Используя метод наибольшего правдоподобия, оценить параметр θ.
6. Построить точный доверительный интервал для параметра $\theta > 0$ равномерного на $[\theta, 2\theta]$ распределения.

Составим функцию правдоподобия:

$$L = P_{n_1}(m_1)P_{n_2}(m_2) = C_{n_1}^{m_1} C_{n_2}^{m_2} p^{m_1+m_2}(1-p)^{n_1-m_1+n_2-m_2}.$$

Найдем логарифм функции правдоподобия:

$$\ln L = \ln\left(C_{n_1}^{m_1} C_{n_2}^{m_2}\right) + (m_1+m_2)\ln p + (n_1-m_1+n_2-m_2)\ln(1-p).$$

Вычислим первую производную по p:

$$\frac{d\ln L}{dp} = \frac{m_1+m_2}{p} + \frac{n_1-m_1+n_2-m_2}{1-p}.$$

Запишем уравнение правдоподобия, для чего приравняем найденную производную нулю:

$$\frac{m_1+m_2}{p} + \frac{n_1-m_1+n_2-m_2}{1-p} = 0.$$

Решение задачи № 1

Решаем полученное уравнение относительно и находим критическую точку

$$p = \frac{m_1 + m_2}{n_1 + n_2}.$$

В этой точке вторая производная отрицательна, значит, это точка максимума и, значит, ее надо принять в качестве оценки по методу наибольшего правдоподобия неизвестной вероятности p биномиального распределения.

Ответ: $p = \dfrac{m_1 + m_2}{n_1 + n_2}.$

Решение задачи № 2

Используя метод наибольшего правдоподобия, оценить параметры a и σ^2 нормального распределения, если в результате n независимых испытаний случайная величина X приняла значения $x_1, x_2, \ldots, x_n$.

Пусть случайная величина $X \in N(a, \sigma^2)$, а $\theta = (a, \sigma^2)$ — оцениваемые параметры распределения.

Составим функцию правдоподобия:

$$L = \prod_{i=1}^{n} p(x_i, , \sigma^2) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - a)^2}{\sigma^2}\right).$$

Найдем логарифм функции правдоподобия:

$$\ln L = -n\ln\sigma - n\ln\sqrt{2\pi} - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - a)^2}{\sigma^2}.$$

Решение задачи № 2

Вычислим частные производные по a и σ:

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - a).$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - a)^2.$$

Запишем уравнения правдоподобия, для чего приравняем найденные производные нулю:

$$\sum_{i=1}^{n} (x_i - a) = 0, \qquad \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - a)^2 = n.$$

Корни этой системы уравнений

$$a = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \qquad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Решение задачи № 3

Легко проверить, что это нерегулярный случай (область возможных значений исследуемого признака, в которой плотность положительная, зависит от оцениваемых по выборке параметров a и b). В связи с этим обычная техника, использующая уравнение максимального правдоподобия, здесь неприменима.
Однако задесь экстремальная задача может быть решена непосредственно.

$$L = \frac{1}{(b-a)^n}$$

$$a \leq \min\{x_i\}, \ i = 1, 2, \ldots, n,$$

$$b \geq \max\{x_i\}, \ i = 1, 2, \ldots, n.$$

Решение задачи № 3

Очевидно, решение экстремальной задачи

$$\max \frac{1}{(b-a)^n}$$

при заданных ограничениях определыяется соотношениями

$$a = \min\{x_i\},\ b = \max\{x_i\},\ i = 1, 2, \ldots, n.$$