

А. Ю. Кулай (Москва, ООО «Стэл-КС»). **Условная вероятность двух реализаций случайной последовательности с неизвестными параметрами распределения.**

Пусть X и O — реализации случайной последовательности, некоторые параметры распределения которой неизвестны. В ряде приложений, например, при распознавании образов [1], необходимо вычислять вероятность X при условии наблюдения O .

Как и в байесовском подходе к оцениванию параметров [2], рассмотрим неизвестные параметры $\bar{\theta}$ как случайную величину с некоторой плотностью распределения. При отсутствии иной информации будем считать, что на параметрическом пространстве Θ задано равномерное распределение с плотностью $\mu(\bar{\theta}) = \text{const}$. Тогда

$$\mathbf{P}\{X|O\} = \int_{\bar{\theta} \in \Theta} \mathbf{P}\{X|\bar{\theta}\} d\mathbf{P}\{\bar{\theta}|O\},$$

$$d\mathbf{P}\{\bar{\theta}|O\} = \frac{\mathbf{P}\{O|\bar{\theta}\}\mu(\bar{\theta})d\bar{\theta}}{\int_{\bar{\theta}' \in \Theta} \mathbf{P}\{O|\bar{\theta}'\}\mu(\bar{\theta}')d\bar{\theta}'} = \frac{\mathbf{P}\{O|\bar{\theta}\}d\bar{\theta}}{\int_{\bar{\theta}' \in \Theta} \mathbf{P}\{O|\bar{\theta}'\}d\bar{\theta}'},$$

$$\mathbf{P}\{X|O\} = \int_{\bar{\theta} \in \Theta} \mathbf{P}\{X|\bar{\theta}\} \frac{\mathbf{P}\{O|\bar{\theta}\}d\bar{\theta}}{\int_{\bar{\theta}' \in \Theta} \mathbf{P}\{O|\bar{\theta}'\}d\bar{\theta}'} = \frac{\int_{\bar{\theta} \in \Theta} \mathbf{P}\{X|\bar{\theta}\}\mathbf{P}\{O|\bar{\theta}\}d\bar{\theta}}{\int_{\bar{\theta}' \in \Theta} \mathbf{P}\{O|\bar{\theta}'\}d\bar{\theta}'}$$

В случае полиномиального распределения вычисление интегралов в последнем равенстве приводит к следующему результату.

Теорема. Если X и O — реализации выборки из неизвестного полиномиального распределения с n исходами на параметрическом пространстве Θ (симплексе) задано равномерное распределение, то

$$\mathbf{P}\{X|O\} = \prod_{j=1}^n \frac{A_{\nu_X(j)+\nu_O(j)}^{\nu_X(j)}}{A_{N_X+N_O+n-1}^{N_X}}$$

где A_n^m — количество размещений из n по m , $\nu_Y(j)$ — частота встречаемости j -го символа в последовательности Y , N_Y — длина последовательности Y , $Y \in \{O, X\}$.

СПИСОК ЛИТЕРАТУРЫ

1. Фукунага К. Введение в статистическую теорию распознавания образов. М.: Наука, 1979.
2. Боровков А. А. Математическая статистика. М.: Наука, 1984.