

Санкт-Петербургский государственный политехнический университет

Лабораторная работа

по курсу «Стохастические модели»

«Проверка гипотезы о присутствии закодированного текста в сообщении»

Студент:	Руцкий В. В.
Группа:	5057/2
Преподаватель:	Иванков А. А.

Санкт-Петербург 2011

1 Постановка задачи

Дано сообщение — упорядоченный набор символов — $M = (s_0, \dots, s_n)$, $s_i \in \Sigma$, $\Sigma = \{c_0, \dots, c_k\}$ — алфавит. Необходимо проверить гипотезу о том, что в сообщении закодирован текст M_0 на английском языке, при условии, что кодирование было произведено переобозначением с помощью какой-то биекции $f: \Sigma \rightarrow \Sigma$ исходных символов новыми: $M = (f(M_0[0]), \dots, f(M_0[n]))$.

2 Решение

Формализация гипотезы Рассмотрим сообщение M_0 как n наблюдений случайной величины S , принимающей значения из Σ . Для английского языка определённой стилистики S подчиняется некоторому закону распределения, который можно считать известным.

Проверка гипотезы Рассмотрим все биекции на Σ : $F = \{f | f: \Sigma \rightarrow \Sigma\}$. Предположим, что $f_j \in F$ — биекция, которой было закодировано сообщение M_0 . Гипотеза H_0 которую необходимо проверить состоит в том, что случайная величина S' , для которой наблюдается выборка $M' = (f_j^{-1}(M[0]), \dots, f_j^{-1}(M[n]))$, подчиняется такому же закону распределения, что и S .

Проверим гипотезу с помощью критерия согласия Пирсона:

$$\chi^2 = \sum_{i=1}^k \frac{(v_i - np_i)^2}{np_i},$$

p_i — теоретическая вероятность наблюдения c_i в S , v_i — число наблюдений c_i в выборке M' . Выборочная характеристика χ^2 при $n \rightarrow \infty$ имеет χ^2 -распределение с $k - 1$ степенями свободы.

Теперь для выбранного уровня значимости α , сравним полученную величину χ^2 с критической величиной $\chi_{cr}^2(k, \alpha)$ (табличная величина): если $\chi^2 < \chi_{cr}^2(k, \alpha)$, то гипотеза принимается с уровнем значимости α .

Если хотя бы для одного f_i принимается H_0 , то в сообщении закодирован текст на английском языке, иначе — нет.

В ходе решения были использованы материалы из [1].

Список литературы

- [1] Г.И. Ивченко and Ю.И. Медведев. *Математическая статистика*. М: Высш. шк., 1984.