1 Постановка задачи

Дано сообщение — упорядоченный набор символов — $M = (s_0, \ldots, s_n)$, $s_i \in \Sigma$, $\Sigma = \{c_0, \ldots, c_k\}$. 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 $f\colon \Sigma \to \Sigma$ исходных символов новыми: $M = (f(M_0[i]), \ldots, f(M_0[n]))$.

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: $F = \{f | f\colon \Sigma \to \Sigma\}$. 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 $M' = (f_j^{-1}(M[0]), \ldots, f_j^{-1}(M[n]))$, 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 $n \to \infty$ имеет χ^2-распределение с $k-1$ степенями свободы.

Теперь для выбранного уровня значимости α, сравним полученную величину χ^2 с критической величиной $\chi^2_{cr}(k, \alpha)$ (табличная величина): если $\chi^2 < \chi^2_{cr}(k, \alpha)$, 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 из [1], [2], [3], [4].

Список литературы

[1] Wikipedia: Chi-square distribution. http://en.wikipedia.org/wiki/Chi-square_distribution.

[2] Wikipedia: Pearson's chi-square test. http://en.wikipedia.org/wiki/Pearson's_chi-square_test.

[3] Wikipedia: Квантили распределения хи-квадрат. http://ru.wikipedia.org/wiki/%D0%9A%D0%B2%D0%B0%D0%BD%D1%82%D0%B8%D0%BB%D0%B8_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F_%D1%85%D0%B8-%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82.

[4] Н.И. Чернова. Лекции по математической статистике. http://www.nsu.ru/mmf/tvims/chernova/ms/lec/node46.html, 2006.