Лабораторная работа 1

Титов Вадим, 676 группа, titov.vn@phystech.edu

Задача 1.3

Каждая пара состоит из оригинального (original_id.bmp) изображения и искаженного (modified_id_phrase.bmp). Также задана пара контрольных изображений original_test, modified_test.

В 99% случаев искажение заключается в добавлении белого шума. В 1% случаев искажение заключается в добавлении к изображению скрытого сообщения. Алгоритм заключается в следующем:

1. У исходной фразы берутся порядковые номера всех символов в порядке английского алфавита (abcz $\to$ 0,1,2,25).
2. Полученный вектор домножается на неизвестный коэффициент α и складывается с вектором картинки (image = image.flatten() + $\alpha \cdot v$ + шум).
3. Если фраза слишком короткая, искажение продолжается периодически.

Требуется раскодировать фразу из контрольной пары.

NB: предполагается, что вы найдете искаженные изображения без шума с применением статистических моделей, а не перебором.

Постановка задачи:

Рассмотрим для каждой пары разности модифицированного вытянутого изображения и вытянутого оригинального. Если в модифицированном изображении нет закодированной фразы, то выборка представляет из себя набор нормальных величин со средним 0. Тогда постановка задачи примет следующий вид:

$$\begin{aligned}
\text{Выборки}: & X^i = (X_1^i, ..., X_p^i), \qquad i = \overline{1, ..., n} \\
\text{Нулевая гипотеза, } & H_0 : medX_j = 0, \\
\text{Альтернатива, } & H_1 : medX_j \neq 0 \\
\text{Статистика}: & W(X^j) = \sum_{i=1}^{p} \text{rank}(|X_i^j|) \cdot \text{sign}(X_i^j) \\
\text{Нулевое распределение}: & \text{табличное}
\end{aligned}$$

В ходе решения поправка на проверку множества гипотез была применена, что значительно увеличило мощность.

Задача 2.1

Рассматривается задача тестирования вакцины от некоторого вируса. Производство вакцины очень дорогое и затратное по времени, поэтому в день может быть произведена только одна ампула.

Требуется проверить, что вакцина помогает (вероятность заразиться меньше у человека с вакциной чем у человека без вакцины).

Эксперимент ставится следующим образом: каждый день есть два идентичных по здоровью человека. Один из людей принимает вакцину, а второй нет, после чего обоих ставят в одну среду с вирусом. В конце для проверяют кто заразился. (В таблице: s — sick; h — healthy)

Весь мир ждет вакцину от данного вируса, поэтому к руководству института постоянно приходят запросы о сроках завершения тестирования образца. Руководство поручило Вам оценить среднее время, которое понадобится на тестирования данной вакцины. А также провести анализ полученных данных на уровне значимости $\alpha = 0.05$ и при ошибке второго рода $\beta = 0.2$.

Требуется:

1. Записать задачу формально;

2. Выполнить оценку среднего количества дней для принятия решения (учесть что истинная вероятность заразиться с вакциной и без равны 0.2 и 0.5 соответственно);

3. Выполнить анализ данных и выяснить работает ли вакцина или нет.

Все выкладки должны быть сделаны аналитически, без использования компьютера.

Решение

Так как ошибки фиксированы и необходимо оценить среднее количество дней, то воспользуемся последовательным критерием Вальда:

Постановка задачи:

Нулевая гипотеза: Доля заболевших людей без вакцинации не меньше доли заболевших пациентов после вакцинации
Альтернатива: Доля заболевших людей без вакцинации меньше доли заболевших пациентов после вакцинации

Формально:

$$
\begin{aligned}
\text{Выборки}: X_1^{30} &= (X_{1,1}, ..., X_{1,30}),\\
X_2^{30} &= (X_{1,1}, ..., X_{2,30}).\\
\text{Нулевая гипотеза},\ H_0: u &> u_U,\\
\text{Альтернатива},\ H_1: u &\leq u_L\\
\text{Статистика}: d_m(X_1^{30}, X_2^{30}) &= \sum_{i=1}^{30}(1 - X_{1,i})X_{2,i}
\end{aligned}
$$

Относительный риск:

$$
u = \frac{k_1}{k_2} = \frac{p_1(1-p_2)}{p_2(1-p_1)} = [\ p_1 = 0.5,\ p_2 = 0.2\ ] = 4
$$

Среднее количество итераций:

$$
\mathbb{E}_u(n) = \frac{L(u)\ln B + (1 - L(u))\ln A}{\frac{u}{u+1}\ln\frac{u_U(1+u_L)}{u_L(1+u_U)} + \frac{1}{u+1}\ln\frac{1+u_L}{1+u_U}} \Bigg/ (p_1(1-p_2) + p_2(1-p_1))
$$

Где $L(u) = \dfrac{A^h - 1}{A^h - B^h}$, а h определяется из условия

$$\frac{u}{u+1} = \frac{1 - \left(\frac{1+u_L}{1+u_U}\right)^h}{\left(\frac{u_U(1+u_L)}{u_L(1+u_U)}\right)^h - \left(\frac{1+u_L}{1+u_U}\right)^h}$$

Решение, при $u_L = 0.8$, $u_U = 1.2$: (Wolfram) $h \approx -6.954475537$
Тогда:

$$L(u) = \frac{A^{-7} - 1}{A^{-7} - B^{-7}} \approx 0.302651$$

Таким образом:

$$\mathbb{E}_u \approx 15$$

Проведем анализ данных для приведенных данных, исключая из рассмотрения пары с одинаковыми измеренями: $r_m = 0.494915m + 6.83805$, $a_m = 0.494915m - 3.84286$

Данные:

С вакциной	Без вакцины
0	0
0	1
0	1
0	0
0	0
0	1
0	0
0	1
0	1
0	1
1	1
0	0
0	1
0	0
0	0
0	1
0	0
1	0
0	1
0	0
0	1
0	0
0	0
0	0
1	1
0	1
1	1
0	1
0	0
1	1

m	a_m	r_m	d_m
1	-3.347945	7.3	0
2	-2.85303	7.8	0
3	-2.3581	8.3	0
4	-1.8632	8.8	0
5	-1.368285	9.3	0
6	-0.87337	9.8	0
7	-0.37845	10.3	0
8	0.11646	10.8	0

Таблица 2:

Таким образом, принимаем 0ю гипотезу о том, что при отсутсвии вакцинации шанс заразиться выше, чем при вакцинации, то есть вакцина полезна. Решение принято на 15й день, удивительно, что совпало с мат.ожиданием среднего количества шагов.