Метод «нуль-пространства» в Scilab.

Поскольку Scilab является открытым программным продуктом, то для проведения ответственных вычислений он нуждается в средствах проверки и диагностики вычислений. Одним из методов проверки является метод «нуль-пространства», который был предложен в документах Национальной физической лаборатории (Англия) http://www.npl.co.uk/ . Этот метод позволяет любому пользователю ПО создать свой набор "эталонных" данных для тестирования ПО, и является наиболее удобным способом их создания. Конечно, есть у этого метода недостаток, который заключается в необходимости наличия хорошего программного генератора псевдослучайных чисел. Но, несмотря на это, этот метод широко используется из-за простоты программной реализации и появления всё более качественных алгоритмов генерации псевдослучайных чисел. Рассмотрим идеи и алгоритм, лежащий в основе этого метода, и затем создадим генератор набора "эталонных" данных для Scilab.

Основная идея этого алгоритма заключается в том, что входные «эталонные» данные формируются таким образом, чтобы «эталонные» выходные данные оставались неизменными — так называемый метод «нуль-пространства». В этом случае различным наборам (векторам) входных «эталонных» данных соответствует единственный выходной набор (вектор) «эталонных» данных. Рассмотрим его применение на примере задачи линейной регрессии. Предположим, что решается задача линейной регрессии: задан вектор

$$y = \begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix}^T$$

результатов наблюдений, проводившихся в дискретные моменты времени x_i

$i=\overline{1,m}$, при этом считаем, что результаты y_i определяются линейной зависимостью $y_i = b_1 + b_2 x_i + r_i$, $i=\overline{1,m}$, где $b_{1,2}$ — параметры, подлежащие оценке, r_i — случайные ошибки, возникающие в процессе измерения. Перепишем систему уравнений в матричной форме:

$$y = Ab + r, \tag{2.2}$$

где A — матрица наблюдений, y — вектор результатов наблюдения, b — вектор параметров модели, r — вектор остатков;

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Известно, что решение задачи линейной регрессии по методу наименьших квадратов характеризуется тем, что

$$A^T r = 0.$$

Следовательно,

$$\sum_{i=1}^{m} r_i = 0, \quad \sum_{i=1}^{m} x_i r_i = 0.$$

Пусть N — матрица, столбцы которой являются базисными векторами нуль-пространства матрицы A^T, тогда $A^T N = 0$. Пусть $r = Nu$, тогда для различных векторов результатов наблюдения y и $y+r$ получаем одинаковые значения параметров b_i. Действительно, $A^T A b = A^T y$ и $A^T\left(y+r\right) = A^T y + A^T N u = A^T y$.

Распишем этапы построения тестовых наборов данных для данной задачи:

1. Вычисляется вектор результатов наблюдения $y_0 = Ab$;

2. Строится матрица N, столбцы которой представляют собой базис нуль-пространства для матрицы A^T;

3. Формируется вектор остатков r = N u, где элементы вектора u генерируются при помощи датчика случайных чисел;

4. Проводится нормировка вектора r таким образом, чтобы он и его компоненты имели распределение с заданным средним значением и СКО;

5. Формируется вектор результатов наблюдения y в соответствии с формулой $y = y_0 + r$.

Целью таких тестов является проверка выбранных математических (статистических) функций конкретного программного продукта для определения степени точности расчетов в различных диапазонах значений входных данных и исполнительных параметров. В случае отсутствие «эталонного» программного обеспечения описанные методы тестирования могут быть использованы для оценки качества алгоритмов обработки измерительной информации в ПО средств измерений и измерительных систем.

Генератор "эталонных" данных на основе метода "нуль-пространства".

Для примера, в Scilab был создан генератор эталонных данных, который можно использовать для тестирования процедуры линейной регрессиии. Здесь была взята модель с линейной зависимостью: y=a+b*x и при конкретных значениях a и b били сгенерированы эталонные данные. Ниже приведён текст такой программы.

```
x=[-1:0.1: 1];//диапазон изменения данных
y=5+2*x';// модель для фитирования
k=size(x);
a=ones(k(2),1);
J=[a x'];
N=kernel(J');// нуль-пространство
mult=N/max(abs(N));
l=size(N);
ro=grand(l(2),1,'nor', 0,1)//генератор чисел с нормальным
//распределением со средним 0 и СКО 1
r=mult*ro;
y1=y+r;// эталонные данные
```

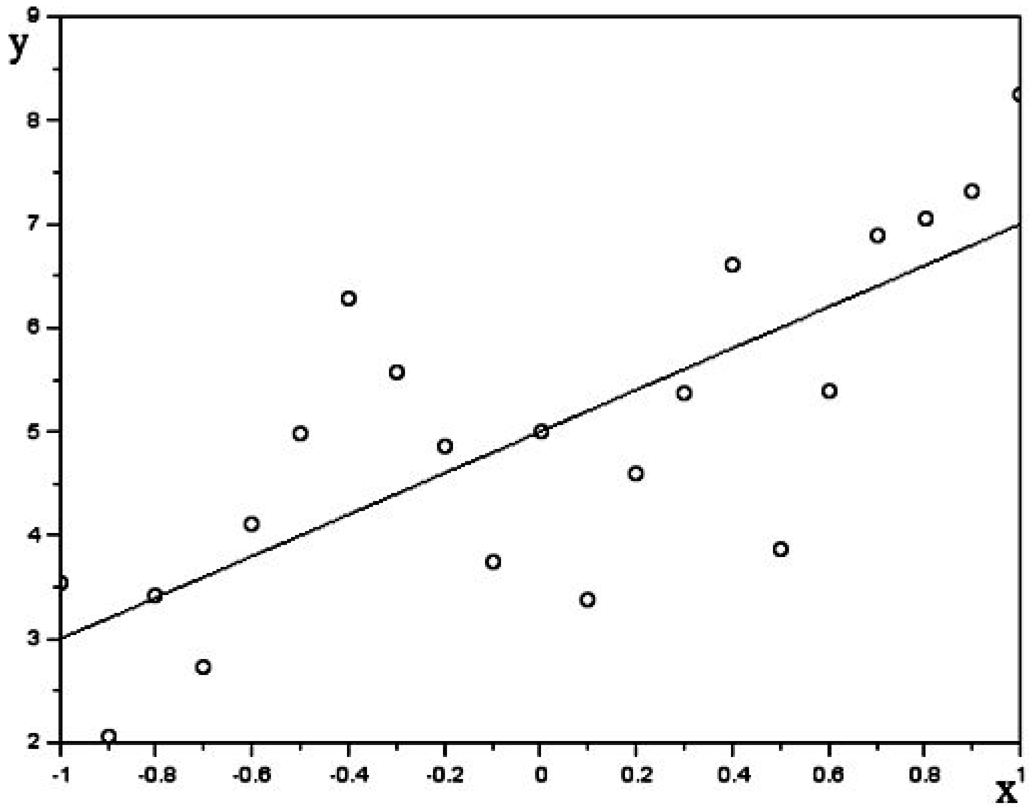

Рис.1 0-эталонные данные, полученные на основе зависимости y=5+2x

На рисунке 1 представлены "эталонные" данные, полученные на основе эталонной зависимости y=5+2x. Выполнив процедуру поиска параметров данной линейной зависимости, исходя из полученных "эталонных" данных, получим параметры нашей тестовой зависимости.

a=2.00000000000000444D+00 b=5.00000000000000000D+00

Относительная ошибка точности вычисления в каждой точке составляет меньше чем 10^{-16}.