ПРИМЕНЕНИЕ M – ОЦЕНКИ ХЬЮБЕРА ДЛЯ ВАРИОГРАММЫ СТАЦИОНАРНОГО СЛУЧАЙНОГО ПРОЦЕССА

Т.В. Цеховая
(БГУ, г. Минск)

Рассмотрим стационарный в широком смысле случайный процесс $X(s)$, $s \in Z$, с нулевым математическим ожиданием и неизвестной вариограммой [1]

$$2\gamma(h) = D(X(s+h) - X(s)),\ h \in Z.$$

Некоторые вопросы исследования оценок вариограммы можно найти, например, в статьях [1-4]. В данной работе обсуждается вопрос оценивания вариограммы стационарного случайного процесса с применением $м$ – оценки хьюбера и изучаются статистические свойства построенных оценок.

Обозначим $U(h)=(X(s+h)-X(s))^2$, $s, h \in Z$. Тогда $2\gamma(h)=MU(h)$, то есть оценивание вариограммы сводится к построению оценки математического ожидания процесса $U(h)$, $h \in Z$. Пусть $X(1), X(2),\dots, X(n)$ – n последовательных, полученных через равные промежутки времени наблюдений за процессом $X(s)$, $s \in Z$, и, следовательно, имеем $U(h)$, $h = \overline{0, n-1}$.

В качестве оценок вариограммы $2\gamma(h)$ рассмотрим M-оценки Хьюбера T_n, которые являются решением уравнения

$$\sum_{j=0}^{n-1} \varphi(U(j) - T_n) = 0,$$

где

$$\varphi(x) = \begin{cases} x, & |x| \le k, \\ \operatorname{sgn}(x), & |x| > k, \end{cases}$$

$0 < k < \infty$. Отметим, что предельные случаи $k \downarrow 0$ и $k \downarrow \infty$ приводят, соответственно, к выборочным медиане $\widetilde{U}$ и среднему $\overline{U}$. Тогда семейство рассматриваемых оценок можно представить в виде $T_n = t_n(F_n) = t_n(U_0, U_1,\dots,U_{n-1})$, где $U_0, U_1,\dots, U_{n-1}$ – порядковые статистики для независимых одинаково распределенных значений $U(0), U(1),\dots, U(n-1)$ с эмпирической функцией распределения $F_n = F_n(u)$. Для среднего, например, функциональная форма не зависит от n и $\overline{U} = t(F_n)$. В других случаях можно выбрать подходящий функционал $t(F_n)$ так, что для больших n разность $t_n(F_n) - t(F_n)$ становится пренебрежительно малой. Так, для выборочной медианы $\widetilde{U}$ соответствующий выбор приводит к функционалу вида $t(F_n) = F_n^{-1}(1/2)$.

Известно, что если $t(\cdot)$ – непрерывный функционал, то оценка $t(F_n)$ состоятельна для $t(F)$, где $F = F(u)$ – истинное распределение $U(0)$, $U(1)$,..., $U(n-1)$. Далее, следуя работе [5], можно показать, что статистика

$$\frac{\{T_n - t(F)\}\sqrt{n}}{\sigma_{t,F}},$$

где

$$\sigma_{t,F}^2 = \int d_{t,F}^2(u)dF(u), \tag{1}$$

$$d_{t,F}(y) = \varphi(y) / \int \varphi'(u)dF(u),$$

имеет асимптотически нормальное распределение с нулевым математическим ожиданием и дисперсией, равной единице.

Чтобы проиллюстрировать свойства M – оценок Хьюбера, приведем некоторые численные результаты, полученные для больших выборок при помощи асимптотической теории [5]. Численные результаты относятся к шести различным порождающим распределениям, а именно

A. Стандартное нормальное $N(0,1)$ - распределение;

B. Смесь нормальных $N(0,1)$ - и $N(0,9)$ - распределений с весами 0,998 и 0,002 соответственно;

C. Смесь нормальных $N(0,1)$ - и $N(0,9)$ - распределений с весами 0,95 и 0,05 соответственно;

D. Смесь нормальных $N(0,1)$ - и $N(0,9)$ - распределений с весами 0,8 и 0,2 соответственно;

E. Распределение Лапласа с плотностью $f(x) = 0{,}5e^{-|x|}$;

F. Распределение Коши с плотностью $f(x) = \{\pi(1 + x^2)\}^{-1}$.

Для каждой комбинации порождающего распределения и M – оценки Хьюбера при известном ряде значений k и симметричного загрязнения ε в таблице приводится значение дисперсии $\sigma_{t,F}^2$ предельного нормального распределения статистики $\sqrt{n}(T_n - t(F))$, вычисленной по формуле (1). Для сравнения в таблицу включены выборочные медиана и среднее. Выбор оценки математического ожидания с меньшей асимптотической дисперсией будет эквивалентен получению лучшей оценки. Столбец ε_{min} содержит значения симметричного загрязнения ε, которые для соответствующего k будут приводить к минимальной асимптотической дисперсии.

Таблица

Асимптотическая дисперсия статистики $\sqrt{n}(T_n - t(F))$ для M – оценок Хьюбера T_n и шести порождающих распределений.

k	ε_{min}	A $\varepsilon=0$	B $\varepsilon=0.002$	C $\varepsilon=0.05$	D $\varepsilon=0.2$	E	F
0.0	1.00000	1.5708	1.5750	1.6810	2.0913	1.0000	2.4674
0.1	0.87529	1.4923	1.4964	1.5999	2.0012	1.0333	2.3786
0.5	0.44171	1.2625	1.2664	1.3650	1.7486	1.1653	2.2881
1.0	0.14283	1.1073	1.1113	1.2142	1.61366	1.3226	2.5465
1.5	0.03761	1.0371	1.0418	1.1601	1.6137	1.4653	2.9927
2.0	0.00842	1.0104	1.0161	1.1600	1.7000	1.5890	3.5208
2.5	0.00160	1.0023	1.0094	1.1864	1.8315	1.6917	4.0888
3.0	0.00025	1.0004	1.0090	1.2221	1.9757	1.7740	4.6785
∞		1.0000	1.0160	1.4000	2.6000	2	∞

Как видно из таблицы, выборочное среднее приемлемо при стандартном $N(0,1)$ – порождающем распределении, но совершенно не пригодно в любой ситуации, где возможно появление резко выделяющихся наблюдений. В случае выборочной медианы, резко выделяющиеся наблюдения оказывают известное влияние. Из полученных результатов следует, что M – оценки Хьюбера являются наиболее стабильными для всех распределений и остаются лучшими при распределениях с относительно небольшим загрязнением. Но при распределениях, порождающих большие доли резко выделяющихся наблюдений, и при распределении Коши их дисперсия существенно возрастает.

Описанный выше метод оценивания вариограммы стационарного случайного процесса является дальнейшим развитием некоторых результатов, приведенных в книге Ф. Хампеля, Э. Рончетти, П. Рауссеу, В. Штаэля [5].

Литература.

1. N. Cressie, D. M. Hawkins. Robust Estimation of the Variogram: I // Mathematical Geology, Vol. 12, No. 2, 1980. – p. 115 – 125.

2. N. Cressie. Fitting Variogram Models by Weighted Least Squares // Mathematical Geology, Vol. 17, No. 5, 1985. - p. 563 – 586.

3. Н.Н. Труш, Т.В. Цеховая. Исследование статистических свойств оценок вариограммы и ковариационной функции // Весці НАН Беларусі.Сер. фіз.-мат. навук.- 2001.- №2.- С.24-29.
4. Т.В. Цеховая Асимптотическое поведение семиинвариантов высших порядков оценки вариограммы. // Тезисы докл. VIII международной математической конференции Ч. 2, с. 178.– 2000г.
5. Ф. Хампель, Э. Рончетти, П. Рауссеу, В. Штаэль. Робастность в статистике. М.: Мир, 1989.

ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ И ПРОЦЕССОВ ПО КОРРЕЛЯЦИОННОМУ ОБРАЗУ

В.Н. Шуть, И.Г. Прожерин
(БГТУ, Брест)

1. Введение

Объектом исследования является группа случайных точек на плоскости отражающая не жестко, неявно некоторый объект либо физический процесс. Такую группу не жестко связанных точек на плоскости назовем дискретной системой. Определим ее следующим образом. Пусть d и D - произвольные положительные числа. Систему точек на плоскости назовем простой (d,D)-системой, если выполняются следующие два условия:

1. расстояние между любыми двумя точками системы не меньше d (d-условие);
2. где бы на плоскости ни нарисовать окружность радиуса D, внутри или на ней самой лежит хотя бы одна точка системы (D-условие).

Непересекающиеся между собой круги одного и того же радиуса D на плоскости можно располагать бесконечно. Поскольку по D-условию в каждом из них должна содержаться, по крайней мере одна точка (d,D-системы), то такая система точек бесконечна.

С другой стороны в любой ограниченной области может содержаться лишь конечное число точек d,D-системы. Это непосредственно вытекает из d-условия. Таким образом, в нашей системе нет ни больших разряжений точек, ни чересчур плотных скоплений. D-условие является притягивающей (собирающей) силой, а r-условие действует как отталкивающая сила. Плотность распределения точек колеблется в некоторых пределах, зависящих, разумеется, от параметров d и D. Введение параметров посредством условий 1 и 2 неоднозначно связывает с ними данную систему. В самом деле, если сис-