

КРАТКИЕ СООБЩЕНИЯ

УДК 519.219 : 519.233.5

А. Г. БУЙМОВ, Н. А. БУЙМОВА  
(Томск)

СТАТИСТИЧЕСКИЙ АНАЛИЗ КОРРЕЛЯЦИЙ  
В ПАЛЬМОВСКОМ ПОЛЕ

**1. Введение и постановка задачи.** Работа посвящена экспериментальному исследованию тестовой модели случайного числового поля, названного в [1] полем Пальма. Это поле представляет собой обобщение телеграфных сигналов [2] на случай произвольной размерности пространства состояний и координатного пространства. Двухточечное вероятностное описание пальмовского поля имеет вид [1]

$$\pi_{st} = \pi_s(\pi_t + (\delta_{st} - \pi_t)\rho) = \pi_s p_{st}, \quad s, t = 1, 2, \dots, N, \quad (1)$$

где  $s = s(r_1)$ ,  $t = t(r_2)$  — номера состояний (значений поля) в точках  $r_1$  и  $r_2$  координатного пространства,  $\rho = \rho(r_1, r_2)$  — корреляционная функция поля из класса неотрицательных монотонно убывающих выпуклых функций. Вероятности  $\pi_s$  являются финальными вероятностями марковской матрицы  $(p_{st})$ . Если  $\rho = \rho(\tau = |r_1 - r_2|)$  — экспонента, то процесс изменения значений поля вдоль прямолинейной траектории представляет собой цепь Маркова. Из структуры распределения (1) вытекает возможность независимого управления одномерными вероятностными ( $\pi_s$ ) и корреляционными ( $\rho$ ) свойствами поля, а также эквивалентность понятий статистической независимости и некоррелированности. Подробные аналитические исследования свойств рассматриваемого поля выполнены в [3]. Там же синтезированы эффективные оценки параметров распределения (1).

Возможность задания произвольных распределений  $\pi_s$ , независимость управления по параметрам  $\pi_s$  и  $\rho$  и известные предельные точности оценивания этих параметров позволяют рекомендовать поле Пальма в качестве тестового сигнала для проверки различных статистических фактов, ЦВМ-анализа и синтеза систем обнаружения, измерения, управления.

Целями данной работы являются экспериментальное подтверждение положения [3] о независимости корреляционной структуры пальмовского поля относительно его нелинейных преобразований; сравнение оценок корреляционных функций из [3] с известной оценкой  $r$ , вычисляемой путем усреднения произведений двухточечных замеров поля [4].

**2. Статистический анализ корреляций.** Если имеется  $n$ -мерная однородная независимая выборка из (1), то в обозначениях  $n_{st}$  — число одинаковых пар  $(s, t)$  из общего числа  $n = \sum n_{st}$  — матрица  $n_{st}$  представляет собой достаточную статистику для оценивания вероятностей  $\pi_{st}$ , а матрица  $(n_{ss}\delta_{st})$  — достаточную статистику для оценивания  $\rho$  [3]. Эффективная оценка  $\rho_1$  с дисперсией  $\sigma_1^2$  в явном виде через статистику  $(n_{ss}\delta_{st})$  не выражается. Поэтому более удобной для оценивания  $\rho$  является оценка  $\rho_2$  с дисперсией  $\sigma_2^2$ , построенная на статистике  $k = \sum n_{ss}$  и имеющая вид

$$\rho_2 = \frac{k/n - \sum \pi_s^2}{1 - \sum \pi_s^2}. \quad (2)$$

В случае равных вероятностей  $\pi_s = 1/N$  оценка  $\rho_2$  совпадает с эффективной оценкой  $\rho_1$ , а при  $\rho \approx 1$  и (или)  $N \gg 1$  мало от нее отличается. Эти выводы [3] иллюстрируются графиками рис. 1. Кривые 1 и 2 соответствуют зависимостям  $\sigma_1/\sqrt{n}$  и  $\sigma_2/\sqrt{n}$  при нормальном распределении поля, кривая 3 — при равномерном распределении. В обоих случаях  $N = 8$ . Из рис. 1 видно, что при  $\rho \geq 0,1$  дисперсии оценок  $\rho_1$  и  $\rho_2$  при нормальном распределении вероятностей  $\pi_s$  (кривые 1 и 2) отличаются менее чем в 1,2 раза. Для статистического обнаружения такого различия по критерию Фи-

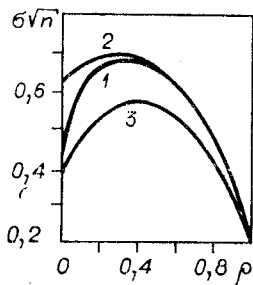


Рис. 1.

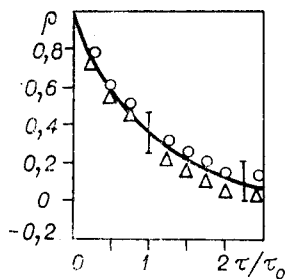


Рис. 2.

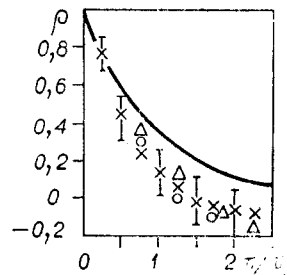


Рис. 3.

шера при уровне значимости 0,05 требуется не менее 200 раз повторить эксперимент по генерации поля и вычислению  $\rho_1$  и  $\rho_2$ . Значительно меньший объем эксперимента ( $\approx 30$  повторов) требуется для различения точности оценок  $\rho_2$  при разных распределениях (кривые 2 и 3). С учетом этих соображений для экспериментального анализа точности оценок корреляционных функций выбрана оценка  $\rho_2$  и для сравнения — известная оценка  $r$ . Оценки вычислялись по реализациям одномерных полей с нормальным и равномерным распределениями  $\pi_s$  и следующими исходными данными:  $N = 8$ ,  $\rho(\tau) = e^{-\tau/\tau_0}$ , длина реализации  $T = 8\tau_0$ , а величина дискрета равна  $\tau_0/4$ . Статистическая обработка оценок выполнена по ансамблям из 30 реализаций на ЭВМ ЕС-1033.

На рис. 2 приведены усредненные  $\rho_2$  для нормального (кружки) и равномерного (треугольники) распределений. Значения  $\rho_2$  вычислены по формуле (2) при известных  $\pi_s$ . Доверительные интервалы построены в предположении нормальности распределения оценки и соответствуют вероятности 0,95. Сплошной линией проведена теоретическая (заданная) корреляционная функция. Из рис. 2 видно, что при известных  $\pi_s$  оценки  $\rho_2$  являются несмещенными. Статистическая незначимость их различий для равномерных и нормальных  $\pi_s$  подтверждает тезис о нечувствительности корреляционной структуры пальмовского поля к нелинейным преобразованиям: в эксперименте нормальность распределения  $\pi_s$  обеспечена путем нелинейного преобразования поля с равномерным распределением.

На рис. 3 приведены те же оценки, что и на рис. 2, но при условии, что вероятности  $\pi_s$  неизвестны. Вместо заданных  $\pi_s$  в формулу (2) подставлялись их максимально правдоподобные оценки. Крестиками обозначена автокорреляционная функция  $r$ . Для нее по  $z$ -критерию Фишера [4] построены 95%-ные доверительные интервалы. Из рис. 3 видно, что при неизвестных  $\pi_s$  оценки  $r$  и  $\rho_2$  оказываются смещенными. В наших обозначениях  $r = \hat{K}(\tau)/\hat{K}(0)$ , где  $\hat{K}(\tau)$  — оценка ковариационной функции поля  $K(\tau)$ :

$$\hat{K}(\tau) = \frac{1}{T-\tau} \sum_{i=1}^{T-\tau} \left( x_{i+\tau} - \frac{1}{T} \sum_{i=1}^T x_i \right) \left( x_i - \frac{1}{T} \sum_{i=1}^T x_i \right).$$

Математическое ожидание этой оценки равно

$$M\{\hat{K}(\tau)\} \approx K(0) [\rho(\tau) - 2\tau_0/T], \quad (3)$$

где  $\tau_0 = \int_0^{\infty} \rho(\tau) d\tau$ . Отсюда следует, что смещенные оценки  $r$  зависят от отношения

$T/2\tau_0$  и что при оценивании  $\rho$  по коррелированной выборке целесообразно вводить поправку на смещение и вычислять

$$r' = r(1 - 2\tau_0/T) + 2\tau_0/T. \quad (4)$$

Теперь заметим, что для максимально правдоподобных оценок  $\hat{\pi}_s$  вероятностей  $\pi_s$ , полученных по  $T/2\tau_0$ -мерным однородным независимым выборкам, математическое ожидание суммы  $\sum \hat{\pi}_s^2$  равно  $M\{\sum \hat{\pi}_s^2\} \approx \sum \pi_s^2 + 2\tau_0/T$ . Поэтому при неизвестных  $\pi_s$  вместо оценки (2) имеет смысл использовать

$$\rho'_2 = \rho_2(1 - 2\tau_0/T) + 2\tau_0/T. \quad (5)$$

Исправление рис. 3 по формулам (4) и (5) сразу же приводит его к виду рис. 2. На основании этих результатов можно сделать вывод, что величина  $T/2\tau_0$  имеет смысл числа степеней свободы коррелированной выборки и, следовательно, получение выборочных значений чаще, чем через  $2\tau_0$ , не может дать существенного увеличения точности оценок.

Все перечисленные исследования были повторены при  $T = 32$ ;  $N = 2, 4, 8$ ;  $\tau_0 = 2, 4, 8$ ;  $\rho(\tau) = e^{-\tau/\tau_0}$  и  $\rho(\tau) = 1 - \tau/2\tau_0$ ;  $\tau \leq 2\tau_0$ , и отклонений от изложенного варианта обнаружено не было.

#### ЛИТЕРАТУРА

1. Буймов А. Г., Решетников М. Т. Случайное поле Пальма.— Труды X Всесоюз. школы-семинара по статистической гидроакустике (Сухуми).— Киев: изд. Киевского политехн. ин-та, 1978.
2. Тихонов В. И., Миронов М. А. Марковские процессы. М.: Сов. радио, 1977.
3. Буймов А. Г. К статистике пальмовских полей.— Автометрия, 1981, № 6.
4. Митропольский А. К. Техника статистических вычислений. М.: ФМ, 1964.

Поступило в редакцию 11 декабря 1979 г.

УДК 621.391.19

Ю. К. КЛОКОВ, В. Н. СИДЕЛЬНИКОВ, Р. Р. ХАМНТОВ

(Москва)

#### ПРЕОБРАЗОВАНИЕ, МИНИМИЗИРУЮЩЕЕ ВЕРХНЮЮ ГРАНИЦУ ВЕРОЯТНОСТИ ОШИБКИ КЛАССИФИКАЦИИ ПУАССОНОВСКИХ НАБЛЮДЕНИЙ

Задача классификации, решаемая на ЭВМ, связана со значительными вычислительными затратами, если велика размерность  $N_0$  вектора наблюдений. Один из эффективных путей сокращения этих затрат — проведение классификации в преобразованном пространстве образов меньшей размерности  $N \ll N_0$ .

Пусть наблюдаемый образ  $\tilde{J}(s)$  является совокупностью (поток) точек

$$\tilde{J}(s) = \sum_{i=1}^n \delta(s - s_i) \quad (1)$$

( $s$  — пространственная или временная переменная) и описывается системой плотностей распределения вероятностей

$$\begin{aligned} p[\tilde{J}(s) | J_k(s)] &= p(n; s_1, s_2, \dots, s_n | J_k(s)) = \\ &= \frac{1}{n!} \prod_{i=1}^n (J_k(s_i) + J_\Phi) \exp \left\{ - \int (J_k(s) + J_\Phi) ds \right\}, \end{aligned} \quad (2)$$

где интенсивность потока  $J_k(s) + J_\Phi$  представлена в виде суммы полезного (эталонного) процесса и постоянного фона. Известны априорные вероятности  $p[J_k(s)] = p_k$ , и при классификации требуется отнести наблюдаемый образ  $\tilde{J}(s)$  к одному из эталонов  $J_k$ ,  $k = 1, 2, \dots, m$ .

В работе \* показано, что при оптимальной байесовской классификации для верхней границы вероятности ошибки справедливо

$$\begin{aligned} P_e &\leq \sum_k^m \sum_l^m P_k P_l \int p(X | \omega_k) \ln p(X | \omega_l) dX - \sum_k^m P_k \int p(X | \omega_k) \ln p(X | \omega_k) dX + C = \\ &= \sum_k^m P_k \left\{ \sum_l^m P_l \overline{\ln p(X | \omega_l)} - \overline{\ln p(X | \omega_k)} \right\} + C, \end{aligned} \quad (3)$$

где усреднение ведется по всем наблюдениям  $X$  относительно распределения  $p(X | \omega_k)$ ,  $\omega_k$  — эталонный образ.

\* Babu C. Chitti. On Feature Extraction in Pattern Recognition.— In: Proc. 5-th Hawaii International Conference on System Science, Honolulu, 1972, p. 11—13.