РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 1 1995

УДК 621.391 : 519.234.3

Г. И. Салов
*(Новосибирск)*

# МЕТОД ПОЛУЧЕНИЯ РАВНОМЕРНО НАИБОЛЕЕ МОЩНЫХ КРИТЕРИЕВ ДЛЯ ОБНАРУЖЕНИЯ ПРОТЯЖЕННЫХ ОБЪЕКТОВ НА СЛУЧАЙНОМ ФОНЕ*

Предлагается простой метод построения непараметрических критериев, заключающийся в «параметризации» проверяемых в задачах обнаружения гипотез путем сведения их к гипотезе о параметрах полиномиального распределения, для которой существует равномерно наиболее мощный критерий.

**Введение.** В работе [1] для задачи обнаружения в заданной части пространства объекта известной формы, когда никаких сведений о типах распределений наблюдаемых (измеряемых) величин не имеется, приведены простейшие и слишком «осторожные» непараметрические критерии проверки статистических гипотез. Здесь предлагается простой метод получения равномерно наиболее мощных (РНМ) критериев, заключающийся в сведении непараметрических гипотез, проверяемых в задачах обнаружения, к гипотезе о параметрах полиномиального распределения, для проверки которой существует РНМ критерий. Полученные критерии являются непараметрическими, поскольку распределения используемых в них статистик при нулевой гипотезе об отсутствии сигнала не зависят от неизвестных наблюдателю распределений наблюдаемых величин.

Допускаем здесь, как и в [1], что анализируемое изображение может быть неоднородным и неизотропным случайным полем, и предполагаем, что все наблюдаемые величины имеют непрерывные функции распределения вероятностей и берутся в достаточно отдаленных друг от друга точках на изображении так, чтобы они могли рассматриваться как статистически независимые в совокупности, когда в поле зрения объект отсутствует. Будем рассматривать односторонние альтернативные гипотезы, полагая для определенности, что присутствие объекта характеризуется тем, что в точках области объекта интенсивность сигнала стохастически больше, чем в точках области фона, т. е. если $\zeta$ — величина, наблюдаемая в точке области объекта, а $\xi$ — в сравнительно близкой к ней точке области фона, то $P(\zeta > a) > P(\xi > a)$ при всех $a$.

Всюду в дальнейшем $\mu$ (с индексами) обозначает непараметрическую статистику Манна — Уитни [2—5].

**Критерий для обнаружения контуров.** Пусть для $i = 1, ..., k$ $\zeta_{i1}, ..., \zeta_{im}$ и $\xi_{i1}, ..., \xi_{in}$, $m, n \geq 1$, — величины, наблюдаемые в точках на линии $i$-й нормали к проверяемому положению контура объекта соответственно по одну сторону от этого положения (в точках предполагаемой области объекта) и по другую (в точках области фона). Будем считать, что если в поле зрения объект отсутствует, то для $i = 1, ..., k$ величины $\zeta_{i1}, ..., \zeta_{im}, \xi_{i1}, ..., \xi_{in}$, принадлежащие сравнительно близким точкам на изображении и составляющие $i$-й блок, имеют

* Работа выполнена при частичной финансовой поддержке грантом № NPA000 Международного научного фонда.

(приближенно) одну и ту же функцию распределения вероятностей $F_i$. Группировка наблюдаемых величин в блоки исключает (или ослабляет) расхождения, являющиеся следствием неоднородности и (или) неизотропности поля изображения. Если же на проверяемом положении интересующий наблюдателя объект присутствует, то величины $\zeta_{is}$ ($s = 1, ..., m$) по предположению стохастически больше, чем $\xi_{it}$ ($t = 1, ..., n$).

Таким образом, для обнаружения контура и, следовательно, объекта представляется естественным проверить гипотезу $H_0$: для каждого $i = 1, ..., k$ величины $\zeta_{is}$ и $\xi_{it}$ ($s = 1, ..., m$; $t = 1, ..., n$) стохастически равны (объект отсутствует) при альтернативной гипотезе $H_1$: $\zeta_{is}$ стохастически больше, чем $\xi_{it}$ (объект присутствует).

Сосчитаем количество $\mu_i^+$ и $\mu_i^-$ положительных и отрицательных разностей в $i$-м блоке:

$$\mu_i^+ = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} - \xi_{it} > 0\}, \qquad \mu_i^- = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} - \xi_{it} < 0\},$$

здесь и далее $I\{A\}$ — индикатор события $A$, равный 1 или 0, в зависимости от того, произошло или не произошло событие $A$. Величины $\mu_i^+$ и $\mu_i^-$ принимают все целые значения от 0 до $mn$ и симметрично распределены около точки $mn/2$, когда гипотеза $H_0$ верна.

Рассмотрим события

$$A_i = \left\{\mu_i^+ > \frac{mn}{2}\right\}, \qquad B_i = \left\{\mu_i^- > \frac{mn}{2}\right\},$$

а также дополнительное событие $C_i = \overline{A_i \cup B_i}$, которое реализуется тогда и только тогда, когда не реализуются оба события $A_i$ и $B_i$. Если $mn$ — нечетное число, как, например, в случае $m = n = 1$, то в силу предположения непрерывности распределений $F_i$ вероятность события $C_i$ равна 0. Однако на практике вследствие округления возможны совпадения значений наблюдаемых величин, так что событие $C_i$ возможно и в этом случае. Более того, полученный ниже результат справедлив для каждого случая. Симметричность распределений $\mu_i^+$ и $\mu_i^-$ означает, что при $H_0$ события $A_i$ и $B_i$ имеют одну и ту же вероятность. Отсюда следующие два предположения (или утверждения) эквивалентны:

1) для каждого $i$ все величины $\zeta_{is}$ и $\xi_{it}$ одинаково распределены,

2) для каждого $i$ $P(A_i) = P(B_i)$.

С другой стороны, стохастическое возрастание $\zeta_{is}$ при $H_1$ скажется на значениях вероятностей $P(A_i)$ и $P(B_i)$ по-разному. Распределение $\mu_i^+$ ($\mu_i^-$) будет «сдвигаться» вправо (влево), и, значит, $P(A_i)$ будет возрастать, а $P(B_i)$ — убывать. Следовательно, альтернативная гипотеза эквивалентна предположению $P(A_i) > P(B_i)$.

Таким образом, переходя к эквивалентным гипотезам, которые можно снова обозначить через $H_0$ и $H_1$, проблему обнаружения контура (объекта) можно трактовать как задачу проверки гипотезы $H_0$: $P(A_i) = P(B_i)$, $i = 1, ..., k$ (объект отсутствует) при альтернативе $H_1$: для всех $i$ $P(A_i) > P(B_i)$ (объект присутствует).

Пусть $\nu_A$, $\nu_B$, $\nu_C$ обозначают количество реализовавшихся событий $A_i$, $B_i$, $C_i$ соответственно во всех $k$ блоках:

$$\nu_A = \sum_{i=1}^{k} I\{A_i\}, \qquad \nu_B = \sum_{i=1}^{k} I\{B_i\}, \qquad \nu_C = \sum_{i=1}^{k} I\{C_i\}. \tag{1}$$

Откажемся (по крайней мере на время) от допущения зависимости $P(A_i)$, $P(B_i)$ и $P(C_i)$ при $H_1$ от $i$ и положим

$$p_A = P(A_i), \qquad p_B = P(B_i), \qquad p_C = P(C_i).$$

Тогда [4—7] совместное распределение $\nu_A, \nu_B$ и $\nu_C$ будет полиномиальным (триномиальным) распределением:

$$P(\nu_A = x_A,\ \nu_B = x_B,\ \nu_C = x_C) = \frac{k!}{x_A!x_B!x_C!} p_A^{x_A} p_B^{x_B} p_C^{x_C}. \tag{2}$$

Ввиду тождества $\nu_A + \nu_B + \nu_C = k$ одну компоненту в левой части (2) можно опустить:

$$P(\nu_A = x_A,\ \nu_C = x_C) = \frac{k!}{x_A!x_C!(k - x_A - x_C)!} p_A^{x_A} p_C^{x_C} p_B^{(k - x_A - x_C)}. \tag{3}$$

Это распределение принадлежит [5, 6] к так называемому $(s + 1)$-параметрическому экспоненциальному семейству распределений вида

$$p(x; \Theta, \theta) = g(x)\exp\left[\Theta U(x) + \sum_{j=1}^{s} \theta_j T_j(x)\right],$$

поскольку правую часть (3) всегда можно переписать в форме

$$\left[\frac{p_A}{p_B}\right]^{x_A}\left[\frac{1 - p_A - p_B}{p_B}\right]^{x_C} p_B^k\, g(x),$$

где $x = (x_A, x_C)$, что сводится к

$$g(x)\exp\{x_A\log(p_A/p_B) + x_C\log[(1 - p_A - p_B)/p_B]\}. \tag{4}$$

Это есть двухпараметрическое $(s = 1)$ семейство с

$$U(x) = x_A, \qquad T(x) = x_C, \qquad \Theta = \log(p_A/p_B), \qquad \theta = \log[(1 - p_A - p_B)/p_B].$$

Отсюда роль достаточных статистик $U$ и $T$ будут играть $\nu_A$ и $\nu_C$ соответственно, так что (4) является одновременно совместным распределением $U$ и $T$. Гипотеза $H_0$ эквивалентна гипотезе $\Theta = 0$, параметр $\theta$ является «мешающим».

С целью получения критерия для проверки гипотезы $H_0$: $\Theta = 0$ против альтернативы $H_1$: $\Theta > 0$ можно воспользоваться теоремой 3 из [6, гл. 4]. Она утверждает, что РНМ критерий в классе всех несмещенных критериев (обладающих естественным свойством: вероятность отвергнуть гипотезу $H_0$, когда она верна, не превосходит вероятности отвергнуть $H_0$, когда она неверна) определяется на условном распределении $U$ при фиксированном значении $T = t$, т. е. на условном распределении $\nu_A$ при $\nu_C = t$. Исходя из совместного распределения (3), нетрудно получить, что условное распределение $\nu_A$ при $\nu_C = t$ является биномиальным $B(p^*; k - t)$ с параметром $p^* = p_A/(p_A + p_B)$ и числом испытаний, равным $k - t$. Проверка гипотезы $H_0$ сводится к проверке гипотезы $p^* = 1/2$ (против альтернативы $p^* > 1/2$), для которой, как известно (см., например, [5, 6]), существует РНМ несмещенный критерий, отвергающий ее, когда $\nu_A > \lambda(t)$. При этом порог $\lambda = \lambda(t)$ — наименьшее целое число, такое что

$$\sum_{i=\lambda+1}^{k-t} \binom{k-t}{i} 2^{-(k-t)} \le \alpha, \tag{5}$$

здесь и далее $\binom{k-t}{i}$ — биномиальный коэффициент, $\alpha$ — заданный допустимый уровень вероятности отклонить гипотезу $H_0$, когда она верна, — уровень вероятности ложного обнаружения при одной проверке гипотезы

$H_0$. Этот критерий делает максимальной условную вероятность отклонения гипотезы $H_0$, когда объект действительно присутствует. Отсюда безусловную вероятность максимизирует критерий

$$\nu_A > \lambda(\nu_C) \tag{6}$$

(эквивалентный критерий: $\nu_A - \nu_B > \lambda(\nu_C) - (k - \nu_C)/2 = \lambda^*(\nu_C)$).

Сформулируем полученный результат.

**Утверждение 1.** Пусть статистики $\nu_A$, $\nu_B$ и $\nu_C$ определяются по (1). Тогда критерий уровня $\alpha$, определяемый по (5) и (6), будет РНМ несмещенным критерием для проверки гипотезы $H_0$: $P(A_i) = P(B_i)$, $i = 1, ..., k$ (объект отсутствует) при альтернативной гипотезе $H_1$: для всех $i$ $P(A_i) > P(B_i)$ (объект присутствует).

При $m = n = 1$ этот критерий сводится к известному критерию знаков.

**Критерий для обнаружения линий и полос.** Пусть $\xi_{i1}, ..., \xi_{in}$, $\zeta_{i1}, ..., \zeta_{im}$, $\psi_{i1}, ..., \psi_{in}$ $(i = 1, ..., k)$ — величины, наблюдаемые на линии $i$-й нормали к средней линии проверяемого положения полосы, причем $\zeta_{i1}, ..., \zeta_{im}$ — в точках, расположенных равномерно поперек самого проверяемого положения (полосы), а $\xi_{i1}, ..., \xi_{in}$ и $\psi_{i1}, ..., \psi_{in}$ — по разные стороны от последнего. При $m = 1$ имеем случай обнаружения линии. Будем считать, что при $H_0$ величины $\xi_{i1}, ..., \xi_{in}, \zeta_{i1}, ..., \zeta_{im}, \psi_{i1}, ..., \psi_{in}$ $(i = 1, ..., k)$ имеют (приближенно) одну и ту же функцию распределения вероятностей $F_i$.

Для каждого $i$ введем следующие статистики:

$$\mu_{i1}^+ = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} > \xi_{it}\}, \qquad \mu_{i1}^- = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} < \xi_{it}\},$$

$$\mu_{i2}^+ = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} > \psi_{it}\}, \qquad \mu_{i2}^- = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} < \psi_{it}\}.$$

Эквивалентными гипотезам $H_0$ и $H_1$: величины $\zeta_{i1}, ..., \zeta_{im}$ стохастически больше, чем $\xi_{i1}, ..., \xi_{in}$ и $\psi_{i1}, ..., \psi_{in}$, являются гипотезы, связанные с событиями

$$A_i = \left\{\mu_{i1}^+ > \frac{mn}{2},\ \mu_{i2}^+ > \frac{mn}{2}\right\}, \qquad B_i = \left\{\mu_{i1}^- > \frac{mn}{2},\ \mu_{i2}^- > \frac{mn}{2}\right\},$$
$$C_i = \overline{A_i \cup B_i}. \tag{7}$$

(Здесь всегда $P\{C_i \mid H_0\} > 0$.) Для доказательства достаточно заметить, что совместное распределение статистик $\mu_{i1}^+$ и $\mu_{i2}^+$ ($\mu_{i1}^-$ и $\mu_{i2}^-$) симметрично при $H_0$ (см. формулу (11) в [1] и [3]):

$$P\{\mu_{i1} = u,\ \mu_{i2} = v \mid H_0\} = P\{\mu_{i1} = mn - u,\ \mu_{i2} = mn - v \mid H_0\}.$$

**Утверждение 2.** Пусть статистики $\nu_A$, $\nu_B$ и $\nu_C$ определяются по формулам (1) и (7). Тогда критерий уровня $\alpha$, определяемый по формулам (5) и (6), будет РНМ несмещенным критерием для проверки гипотезы $H_0$: $P(A_i) = P(B_i)$, $i = 1, ..., k$ (полоса отсутствует) при альтернативной гипотезе $H_1$: для всех $i$ $P(A_i) > P(B_i)$ (полоса присутствует).

Доказательство точно такое же, как для утверждения 1.

**Критерий для обнаружения хребтов.** Рассмотрим простейший вариант задачи обнаружения хребтов — с минимальным числом наблюдений на области склонов хребта. Пусть $\xi_{i1}, \xi_{i2}, \zeta_i, \psi_{i1}, \psi_{i2}$ $(i = 1, ..., k)$ — величины, наблюдаемые в точках на линии $i$-й нормали к проверяемому положению гребня хребта, причем $\zeta_i$ — в точке на самом проверяемом положении гребня, а $\xi_{i1}$, $\xi_{i2}$ и $\psi_{i1}, \psi_{i2}$ — по разные стороны от последнего (на склонах). Вновь будем считать, что при отсутствии объекта (при $H_0$) для каждого $i$ величины $\xi_{i1}, \xi_{i2}$, $\zeta_i, \psi_{i1}, \psi_{i2}$ имеют (приближенно) одну и ту же функцию распределения вероятностей $F_i$ (стохастически равны), а при наличии хребта (при $H_1$) — стохастически $\zeta_i > \xi_{i1} > \xi_{i2}$ и $\zeta_i > \psi_{i1} > \psi_{i2}$.

Утверждение 3. Пусть для $i = 1, ..., k$

$$A_i = \{\zeta_i > \xi_{i1} > \xi_{i2},\ \zeta_i > \psi_{i1} > \psi_{i2}\},$$

$$B_i = \{\zeta_i < \xi_{i1} < \xi_{i2},\ \zeta_i < \psi_{i1} < \psi_{i2}\}, \qquad C_i = \overline{A_i \cup B_i}.$$

Тогда, если статистики $\nu_A$, $\nu_B$ и $\nu_C$ определяются по формуле (1), то критерий уровня $\alpha$, определяемый по формулам (5) и (6), будет РНМ несмещенным критерием для проверки гипотезы $H_0$: $P(A_i) = P(B_i)$, $i = 1, ..., k$ (хребет отсутствует) при альтернативной гипотезе $H_1$: для всех $i$ $P(A_i) > P(B_i)$ (хребет присутствует).

Для доказательства нужно только показать, что $P\{A_i \mid H_0\} = P\{B_i \mid H_0\}$. Однако последнее очевидно в силу симметрии событий $A_i$ и $B_i$.

Заключение. Построенные непараметрические критерии просты и являются равномерно наиболее мощными при рассматриваемых альтернативах. Дополнительное их преимущество — легко вычислять критические значения (пороги) $\lambda$ по желаемому уровню $\alpha$.

В данном методе вместо статистики Манна — Уитни $\mu$ можно взять другую подходящую статистику с симметричным распределением. В частности, когда внутри каждого блока наблюдаемые величины имеют распределения, близкие к гауссовым с одинаковыми (приближенно) дисперсиями, можно ожидать, что мощность первых двух полученных критериев увеличится, если вместо $\mu$-статистики взять известную [6, с. 233] $t$-статистику одностороннего двухвыборочного $t$-критерия (который в этом случае сам по себе является РНМ несмещенным критерием для проверки гипотезы однородности внутри блока), полагая, например, для критерия обнаружения линий и полос:

$$A_i = \{t(\xi_i, \zeta_i) > 0,\ t(\psi_i, \zeta_i) > 0\},$$

$$B_i = \{t(\xi_i, \zeta_i) < 0,\ t(\psi_i, \zeta_i) < 0\}.$$

## СПИСОК ЛИТЕРАТУРЫ


1. Салов Г. И. Непараметрические критерии для обнаружения контуров, линий, полос и хребтов заданной формы на случайном фоне // Автометрия. 1994. № 1.
2. Mann H. B., Whitney D. R. On a test of whether one of two random variables is stochastically larger than the other // Ann. Math. Stat. 1947. 18. P. 50
3. Whitney D. R. A bivariate extension of the U statistic // Ann. Math. Stat. 1951. 22. P. 274.
4. Уилкс С. Математическая статистика. М.: Наука, 1967.
5. Кендалл М. Дж., Стюарт А. Статистические выводы и связи. М.: Наука, 1973.
6. Леман Э. Проверка статистических гипотез. М.: Наука, 1964.
7. Кендалл М. Дж., Стюарт А. Теория распределений. М.: Наука, 1966.


*Поступила в редакцию 22 сентября 1994 г.*