РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

№ 3 1997

УДК 621.391 : 519.234.3

Г. И. Салов

(Новосибирск)

О МОЩНОСТИ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ ДЛЯ ОБНАРУЖЕНИЯ ПРОТЯЖЕННЫХ ОБЪЕКТОВ НА СЛУЧАЙНОМ ФОНЕ

Детализируется и усиливается ранее предложенный метод получения непараметрических критериев. Даются явные формулы для распределений статистик Манна — Уитни, позволяющие вычислять истинные уровни значимости, а в одном важном случае и мощность непараметрических критериев для обнаружения ряда объектов. Приведены результаты численного исследования этих критериев.

Введение. Настоящая работа примыкает к статье [1] и к заметке [2], которую она детализирует и развивает. Как и в [1, 2], рассматриваются k блоков измеряемых (наблюдаемых) случайных величин, в частности $\{\xi_{i1}, \ldots, \xi_{in}, \zeta_{i1}, \ldots, \zeta_{im}, \psi_{i1}, \ldots, \psi_{in}\}$, $i = 1, \ldots, k$, где величины $\zeta_{i1}, \ldots, \zeta_{im}$ принадлежат точкам области проверяемого положения объекта на изображении, а $\xi_{i1}, \ldots, \xi_{in}$ и $\psi_{i1}, \ldots, \psi_{in}$ — точкам области фона по разные стороны от объекта (более точные схемы наблюдений для некоторых задач обнаружения даны ниже). Предполагается, что если в поле зрения объект отсутствует, то все наблюдаемые величины независимы в совокупности и для $i = 1, \ldots, k$ величины внутри i-го блока имеют одну и ту же непрерывную функцию распределения вероятностей F_i, зависящую от i, когда изображение неоднородное. Если же на проверяемом положении объект присутствует, то величины $\zeta_{i1}, \ldots, \zeta_{im}$ имеют функцию распределения $G_i(x) < F_i(x)$ для всех x, т. е. будут стохастически больше $\xi_{i1}, \ldots, \xi_{in}$ и $\psi_{i1}, \ldots, \psi_{in}$.

Функции F_i и G_i по предположению не известны наблюдателю. Решение задачи обнаружения объекта состоит в построении непараметрического (не зависящего от F_i и G_i) критерия для проверки гипотезы H_0: наблюдаемые величины внутри каждого блока однородны (объект отсутствует) против альтернативы H_1: $G_i(x) < F_i(x)$, $i = 1, \ldots, k$ (объект присутствует).

В [1] предлагается взять в качестве статистики (статистик) критерия сумму (суммы) статистик Манна — Уитни, определяемых внутри каждого блока наблюдаемых величин.

Предлагаемый в [2] простой метод построения критериев состоит в «параметризации» задачи. Для этого в рассмотрение вводится полная группа событий A_i, B_i и $C_i = \overline{A_i \cup B_i}$. При $n = m = 1$ $A_i = \{\zeta_i > \xi_i, \zeta_i > \psi_i\}$, $B_i = \{\zeta_i < \xi_i, \zeta_i < \psi_i\}$. В общем случае эти события определяются значениями упомянутых статистик Манна — Уитни. В [2] проверяется гипотеза H_0^*: вероятности $p_A = \mathbf{P}(A_i)$ и $p_B = \mathbf{P}(B_i)$ равны при альтернативе H_1^*: $p_A > p_B$. После определения значений p_A^0, p_B^0, p_C^0 вероятностей p_A, p_B, p_C при H_0 гипотеза H_0^* превращается в простую гипотезу вида H_0^*: $p_A = p_B$, $p_C = p_C^0$ — или, касаясь переобозначенных параметров двухпараметрического экспоненциального распределения в [2], вида H_0^*: $\theta = \log(p_A/p_B) = 0$, $\psi = \log(p_C^0/p_B^0) = \psi^0$.

Приведенный в [2] обычный (нерандомизированный) критерий для проверки гипотезы H_0^*, основанный на числах ν_A, ν_B, ν_C осуществившихся событий A_i, B_i, C_i во всех блоках, является несмещенным и при большом числе k блоков равномерно наиболее мощным с уровнем значимости, близким к заданному, равному α. Короткое доказательство последнего факта опирается на обобщенный нерандомизированный вариант леммы Неймана — Пирсона и монотонное отношение правдоподобия экспоненциальных распределений [3, с. 281—282].

Между тем гипотеза H_j влечет гипотезу H_j^* $(j = 0, 1)$, однако, к сожалению, не совпадает с ней. Возникает естественный вопрос: если гипотезы H_j и H_j^* не совпадают, то будет ли критерий работы [2] более мощным, чем соответствующий критерий в [1]? Отложим ответ до конца введения.

Рандомизация критериев [2] по схеме, отвечающей теореме 3 из [4, гл. 4], обеспечивает возможность получать в точности заданный уровень значимости, равный α. При этом и условный уровень значимости при $T = t$ (или, что то же самое, при $\nu_C = t$) также будет равен α. В таком случае говорят [4], что критерий имеет неймановскую структуру. Однако поскольку проверяемые в [2] гипотезы простые, то полнота «семейства» распределений статистики T не имеет места, т. е. выполняются не все предпосылки упомянутой теоремы 3. Отсюда нельзя сделать вывод, что рандомизированные в соответствии с теоремой 3 критерии будут равномерно наиболее мощными среди всех несмещенных критериев.

Таким образом, поставленный выше вопрос приводит к вопросу о выгодности рандомизации критериев из [2]. Другой вопрос возникает из того факта, что простой выбор событий A_i (B_i) в [2], как оказывается, не всегда является наиболее подходящим. Более того, соответствующий выбор этих событий зависит от самой задачи обнаружения, числа наблюдаемых величин в блоках и т. д.

Цель данной работы — получить явные формулы для совместных распределений двух статистик Манна — Уитни при H_0, а в одном важном случае и при H_1, которые дают возможность вычислить истинные уровни значимости критериев и сравнить мощности критериев работы [1] и [2]. Исследование, проведенное в последнем разделе данной работы, посвящено также построению нового более мощного непараметрического критерия для обнаружения гребней.

Весьма важным и довольно часто встречающимся является случай, когда в изображении наиболее вероятны малые уровни яркости и менее вероятны — большие. При этом удобной аппроксимацией распределения вероятностей значений яркости является экспоненциальное распределение (см., например, [5, с. 8] и [6])

$$F_i(x) = 1 - \exp(-x/\sigma_i), \qquad x \geq 0, \tag{1}$$

в точках области фона и

$$G_i(x) = 1 - \exp(-(x - a_i)/\tau_i), \qquad a_i > 0, \quad x \geq a_i, \tag{2}$$

в точках области объекта. Для простоты положим $\sigma_i = \tau_i = 1$, $a_i = a$. Тогда при $x \geq a$

$$F_i(x) = 1 - \exp(-a)(1 - G_i(x)) = c + bG_i(x), \tag{3}$$

где $b = \exp(-a)$, $c = 1 - \exp(-a)$.

Ответ на поставленный выше в самой общей форме вопрос в этом случае оказывается отрицательным. Но лишь в одной задаче обнаружения контура объекта критерий из [2] уступает соответствующему критерию работы [1], в других же задачах критерии, построенные по методу [2], могут быть значительно мощнее критериев из [1].

Рассмотрим в отдельности каждую задачу обнаружения. Для краткости критерий, основанный на статистике μ (на паре статистик μ_1 и μ_2), будем называть просто μ-критерием (соответственно $\mu_1\mu_2$-критерием) и т. д.

О мощности критериев для обнаружения одного контура. Напомним вначале схему наблюдения. 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). Если в поле зрения объект отсутствует, то для $i = 1, ..., k$ величины $\zeta_{i1}, ..., \zeta_{im}, \xi_{i1}, ..., \xi_{in}$, составляющие i-й блок, и более однородные, чем все наблюдаемые величины в целом, имеют (по крайней мере, приближенно) одну и ту же функцию распределения вероятностей F_i, неизвестную наблюдателю. Если же на проверяемом положении интересующий наблюдателя объект присутствует, то величины ζ_{is}, $s = 1, ..., m$, имеют новую неизвестную функцию распределения G_i ($G_i(x) < F_i(x)$ для всех x).

Для обнаружения контура объекта достаточно проверить гипотезу H_0: для каждого $i = 1, ..., k$ величины ζ_{is} и ξ_{it}, $s = 1, ..., m$, $t = 1, ..., n$, стохастически равны (объект отсутствует) при альтернативной гипотезе H_1: ζ_{is} стохастически больше, чем ξ_{it} (объект присутствует).

Рассмотрим два критерия:

1. μ-критерий [1], основанный на сумме μ статистик μ_i^+ Манна — Уитни:

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

здесь и далее $I\{A\}$ — индикатор события A, равный 1 или 0, в зависимости от того, произошло или не произошло событие A. Критерий отклоняет гипотезу H_0 в пользу H_1, когда $\mu > \lambda$. Порог λ — наименьшее целое число, такое что $P(\mu > \lambda \mid H_0) \leq \alpha$.

2. Нерандомизированный ν-критерий [2], основанный на числах ν_A, ν_B, ν_C событий:

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

где C_i — здесь и в дальнейшем дополнительное событие, которое реализуется тогда и только тогда, когда не реализуются оба события A_i и B_i. Он отклоняет гипотезу H_0, когда $\nu_A > \lambda(k - \nu_C)$, при этом $\lambda(z)$ — наименьшее целое число, такое что

$$\sum_{i=\lambda+1}^{z} \binom{z}{i} 2^{-z} \leq \alpha^*, \tag{4}$$

где $\binom{z}{i}$ — здесь и в дальнейшем биномиальный коэффициент; α^* — гарантированный уровень значимости. Фактически достигаемый уровень значимости может оказаться намного меньше α^*. Последнее вынуждает выбирать значение $\alpha^* > \alpha$ так, чтобы достигаемый уровень значимости был по возможности близок к желаемому уровню α (но не превосходил α).

Чтобы вычислить фактический уровень значимости и мощность критериев, предварительно нужно сосчитать вероятности событий $\{\mu_i^+ = u\}$ и $\{\mu_i^- = u\}$ при H_0 и H_1. При H_0 эти вероятности не зависят от распределений F_i и равны

$$\frac{m!n!}{(m+n)!} \sum_{p(n)} 1, \tag{5}$$

где сумма берется по всем представлениям (упорядоченным разбиениям) $p(n)$ числа n в виде суммы целых неотрицательных чисел

$$n = n_0 + n_1 + \ldots + n_m,$$

удовлетворяющих равенству

$$\sum_{j=0}^{m} (m - j)n_j = u \tag{6}$$

или

$$\sum_{j=0}^{m} jn_j = u \tag{7}$$

в соответствии с тем, вычисляется первая или вторая вероятность, т. е. сумма в (5) равна просто числу разбиений, удовлетворяющих условию (6) или соответственно (7) среди всех $(m + n)!/(m!n!)$ разбиений. Симметрия (5), а также (6) и (7) подтверждает, что при H_0 указанные события имеют одну и ту же вероятность.

При H_1 получение явных и простых формул для нужных вероятностей событий возможно, по-видимому, лишь в некоторых случаях. Во многих случаях (в частности, для нормального распределения) придется прибегать к численному интегрированию. Для рассматриваемого здесь случая с экспоненциальными распределениями (1)—(3) обе вероятности $\mathbf{P}\{\mu_i^+ = u \mid H_1\}$ и $\mathbf{P}\{\mu_i^- = u \mid H_1\}$ могут быть записаны в виде

$$m!n! \sum_{p(n)} \sum_{s=0}^{n_0} c^{n_0 - s} b^{n - n_0 + s} [(n_0 - s)!(m + n - n_0 + s)!]^{-1}, \tag{8}$$

где, как и выше, для первой вероятности суммирование проводится по всем разбиениям $p(n)$, удовлетворяющим условию (6), а для второй — условию (7), при этом n_0 — первый элемент этих разбиений.

Полученные значения мощности критериев при $m = n = 2$ представлены в табл. 1. Для каждого значения k первая строка принадлежит первому критерию, вторая — второму. Значения мощности при $a = 0$ здесь и в дальнейшем являются фактическими уровнями значимости критериев. Из сравнения приведенных результатов видно, что с точки зрения мощности μ-критерий лучше, чем ν-критерий.

О мощности критериев для совместного обнаружения противолежащих контуров. Другой способ обнаружения протяженного объекта, в котором при отклонении гипотезы H_0 вероятность того, что имеет место именно интересующий наблюдателя объект, по-видимому, больше, чем в предыдущем, — это одновременное обнаружение двух противолежащих (левого и правого или внешнего и внутреннего) контуров. Пусть для $i = 1, \ldots, k$ $\zeta'_{i1}, \ldots, \zeta'_{im}, \xi'_{i1}, \ldots, \xi'_{in}$ — величины, наблюдаемые в окрестности проверяемого положения первого контура, а $\zeta''_{i1}, \ldots, \zeta''_{im}, \xi''_{i1}, \ldots, \xi''_{in}$ — параллельно в окрестности противоле-

Таблица 1

k	a								
	0,0	0,1	0,2	0,3	0,5	0,7	1,0	1,5	2,0
10	0,0556	0,1295	0,2413	0,3785	0,6529	0,8432	0,9647	0,9982	$0,9^4 37$
	0,0646	0,1474	0,2625	0,3948	0,6472	0,8244	0,9503	0,9958	$0,9^3 72$
20	0,0342	0,1221	0,2871	0,4973	0,8379	0,9682	0,9986	$0,9^5 75$	
	0,0377	0,1339	0,3018	0,5038	0,8253	0,9587	0,9972	$0,9^4 87$	

жащего контура ($m, n \geq 1$), и пусть $\mu_i'^{+}$, $\mu_i'^{-}$, а также $\mu_i''^{+}$, $\mu_i''^{-}$ определяются аналогично μ_i^{+}, μ_i^{-} в предыдущем разделе.

Сделаем одно замечание по поводу выбора событий A_i и B_i в этой задаче и в последующих. Если следовать [2], то для настоящей задачи получим

$$A_i = \left\{\mu_i'^{+} > \frac{mn}{2},\ \mu_i''^{+} > \frac{mn}{2}\right\}, \qquad B_i = \left\{\mu_i'^{-} > \frac{mn}{2}, \mu_i''^{-} > \frac{mn}{2}\right\}. \tag{9}$$

Однако такие события могут оказаться слишком редкими (маловероятными), что может привести к снижению мощности критерия.

Для выбора подходящих событий A_i (B_i) все возможные события вида $\{\mu_i'^{+} = u_i,\ \mu_i''^{+} = v_i\}$ (соответственно $\{\mu_i'^{-} = u_i,\ \mu_i''^{-} = v_i\}$) полезно расположить в квадратную матрицу, столбцы и строки которой перенумерованы от 0 до $m \times n$ так, чтобы пара (u_i, v_i) отождествлялась с адресом события в этой матрице. Тогда максимальное допустимое событие A_i (аналогично B_i) есть объединение тех событий (элементов) этой матрицы, которые расположены ниже ее побочной диагонали. В частности, при $n \times m = 1$ всегда $A_i = \{\mu_i'^{+} = 1,\ \mu_i''^{+} = 1\}$ (соответственно $B_i = \{\mu_i'^{-} = 1,\ \mu_i''^{-} = 1\}$), а при $m \times n = 2$ максимальное событие, например A_i, состоит из следующих событий: $\{\mu_i'^{+} = 1,\ \mu_i''^{+} = 2\}$, $\{\mu_i'^{+} = 2,\ \mu_i''^{+} = 1\}$, $\{\mu_i'^{+} = 2,\ \mu_i''^{+} = 2\}$. В общем случае наиболее подходящее событие A_i (B_i) является подмножеством максимального допустимого события. Его элементы «удаляются» от побочной диагонали указанной матрицы по мере того, как произведение $m \times n$ возрастает.

Перейдем теперь к формулировке критериев. Здесь будут рассмотрены следующие критерии:

1. $\mu'\mu''$-критерий, основанный на суммах μ' и μ'' статистик $\mu_i'^{+}$ и $\mu_i''^{+}$ соответственно. Он отклоняет гипотезу H_0, когда одновременно $\mu' > \lambda$ и $\mu'' > \lambda$, где порог λ — наименьшее целое число, такое что $P(\mu' > \lambda,\ \mu'' > \lambda \mid H_0) \leq \alpha$.

2. Рандомизированный по теореме 3 из [4, гл. 4] ν-критерий, основанный на числах ν_A, ν_B, ν_C событий A_i, B_i, C_i. Он достоверно отклоняет гипотезу H_0 при $\nu_A > \lambda(k - \nu_C)$ и отклоняет ее случайно с вероятностью $\gamma(k - \nu_C)$ при $\nu_A = \lambda(k - \nu_C)$, где $\lambda(z)$ и $\gamma(z)$ определяются равенствами: $\lambda(0) = 0, \gamma(0) = \alpha$,

$$\sum_{i=\lambda+1}^{z} \binom{z}{i} 2^{-z} + \gamma(z)\binom{z}{\lambda} 2^{-z} = \alpha \qquad (z = 1, \ldots, k).$$

Отсюда условная вероятность отклонения гипотезы H_0 при данном $T = t$ (или, что то же, при $\nu_C = t$) равна α при всех t, т. е. критерий имеет неймановскую структуру относительно T, и, следовательно, его уровень значимости также равен α.

Требование наличия неймановской структуры здесь нельзя согласовать с другими естественными статистическими требованиями, в частности, с требованием, чтобы критерий не отклонял гипотезу H_0 (с вероятностью 1), если $\nu_C = k$, т. е. чтобы $\gamma(0) = 0$.

3. Нерандомизированный ν-критерий, отклоняющий гипотезу H_0, когда $\nu_A > \lambda(k - \nu_C)$, где ν_A и ν_C определяются точно так же, как в предыдущем критерии, а пороги $\lambda(z)$ находятся по формуле (4).

Все критерии полностью характеризуются так называемой критической функцией. В частности, для ν-критериев она имеет вид:

$$\varphi(u, t) = \begin{cases} 1 & \text{при}\ \ u > \lambda(k - t), \\ \gamma(k - t) & \text{при}\ \ u = \lambda(k - t), \\ 0 & \text{при}\ \ u < \lambda(k - t). \end{cases}$$

Для нерандомизированного критерия $\gamma(z) = 0$ при всех z. Отсюда фактический уровень значимости и мощность ν-критериев определяются соответственно равенствами

$$\mathbf{M}_0\varphi(\nu_A, \nu_C) = \sum_{u+s+t=k} \varphi(u, t)\mathbf{P}\{\nu_A = u, \nu_B = s, \nu_C = t \mid H_0\},$$

$$\mathbf{M}_1\varphi(\nu_A, \nu_C) = \sum_{u+s+t=k} \varphi(u, t)\mathbf{P}\{\nu_A = u, \nu_B = s, \nu_C = t \mid H_1\},$$

где суммирование проводится по всем разбиениям $u + s + t = k$ (порядок следования чисел u, s, t важен). Вероятности под знаком суммы являются элементами полиномиального распределения с параметрами p_A, p_B, p_C (вероятностями событий A_i, B_i, C_i), которые в силу предположения независимости наборов наблюдаемых величин легко определяются, исходя из соотношений (5) и (8).

В табл. 2 для каждого значения k в первой строке приведены полученные значения мощности первого критерия, во второй — второго и т. д. При $n = m = 2$ в качестве подходящего события A_i (аналогично B_i) оказалось максимальное допустимое событие, но без элементов матрицы (событий) с

Таблица 2

m	k	a								
		0,0	0,1	0,2	0,3	0,5	0,7	1,0	1,5	2,0
1	11	0,0753 0,0675 0,0675	0,1527 0,1406 0,1443	0,2570 0,2428 0,2525	0,3769 0,3643 0,3809	0,6122 0,6144 0,6366	0,7887 0,7993 0,8225	0,9286 0,9418 0,9537	0,9911 0,9952 0,9969	$0{,}9^{3}07$ $0{,}9^{3}71$ $0{,}9^{3}86$
	20	0,0634 0,0598 0,0598	0,1646 0,1646 0,1670	0,3155 0,3275 0,3345	0,4874 0,5165 0,5277	0,7747 0,8212 0,8330	0,9230 0,9552 0,9609	0,9887 0,9968 0,9975	$0{,}9^{3}72$ $0{,}9^{4}85$ $0{,}9^{5}14$	
	50	0,0103 0,0094 0,0094	0,0754 0,0778 0,0795	0,2572 0,2835 0,2907	0,5207 0,5829 0,5946	0,8965 0,9470 0,9519	0,9885 0,9979 0,9983	$0{,}9^{3}81$ $0{,}9^{5}67$ $0{,}9^{5}78$		
2	9	0,0682 0,0680 0,0680	0,1794 0,1792 0,1839	0,3425 0,3376 0,3499	0,5221 0,5104 0,5289	0,8037 0,7871 0,8061	0,9369 0,9262 0,9372	0,9913 0,9884 0,9912	$0{,}9^{3}78$ $0{,}9^{3}58$ $0{,}9^{3}77$	
	20	0,0293 0,0293 0,0290	0,1497 0,1545 0,1574	0,3902 0,4005 0,4108	0,6545 0,6669 0,6812	0,9395 0,9471 0,9538	0,9940 0,9957 0,9967	$0{,}9^{4}00$ $0{,}9^{4}54$ $0{,}9^{4}73$		
	50	0,0108 0,0103 0,0103	0,1724 0,1816 0,1837	0,5937 0,6273 0,6330	0,8973 0,9237 0,9267	0,9986 $0{,}9^{3}60$ $0{,}9^{3}65$	$0{,}9^{5}52$ $0{,}9^{6}72$ $0{,}9^{6}78$			
3	10	0,0507 0,0507 0,0463	0,1906 0,1696 0,1781	0,4215 0,3672 0,4061	0,6576 0,5874 0,6495	0,9247 0,8856 0,9294	0,9891 0,9793 0,9917	$0{,}9^{3}63$ $0{,}9^{3}00$ $0{,}9^{3}78$		
	20	0,0416 0,0416 0,0405	0,2548 0,2385 0,2460	0,6109 0,5835 0,6058	0,8660 0,8530 0,8725	0,9937 0,9942 0,9962	$0{,}9^{3}87$ $0{,}9^{4}26$ $0{,}9^{4}67$			

адресами (1,4) и (4,1), а при $n = m = 3$ — множество элементов матрицы (событий), расположенных ниже побочной диагонали и двух непосредственно с ней смежных. Численные результаты показывают, что нерандомизированный ν-критерий имеет преимущество по сравнению с двумя другими. При $n = m = 3$ и $k = 10$ оно не является столь явным. С увеличением k мощность рандомизированного ν-критерия становится лишь незначимо меньше мощности нерандомизированного. Поэтому, когда желаемое значение α находится между частными значениями уровня значимости нерандомизированного ν-критерия, равными α_1 и α_2, наблюдатель (исследователь) при планировании эксперимента (или системы обнаружения) может решить принять нижний уровень α_1 или применить рандомизацию критерия, или же принять α_2, если α_2 не слишком сильно отличается от α, что приведет к максимальной вероятности обнаружения.

О мощности критериев для обнаружения линий. Пусть, как и в [1, 2], $\xi_{i1}, ..., \xi_{in}, \zeta_i, \psi_{i1}, ..., \psi_{in}$ — величины, наблюдаемые на линии i-й нормали к проверяемому положению линии ($i = 1, ..., k$), причем ζ_i — в точке на самом проверяемом положении, а $\xi_{i1}, ..., \xi_{in}$ и $\psi_{i1}, ..., \psi_{in}$ — по разные стороны от последнего. Будем считать, что для всех i при H_0 величины $\xi_{i1}, ..., \xi_{in}, \zeta_i, \psi_{i1}, \psi_{i2}, ..., \psi_{in}$ стохастически равны, а при H_1 величина ζ_i будет уже стохастически больше, чем величины $\xi_{i1}, ..., \xi_{in}$ и $\psi_{i1}, ..., \psi_{in}$.

Для каждого i введем исходные статистики:

$$\mu_{i1}^+ = \sum_{t=1}^{n} I\{\zeta_i > \xi_{it}\}, \qquad \mu_{i2}^+ = \sum_{t=1}^{n} I\{\zeta_i > \psi_{it}\},$$

$$\mu_{i1}^- = \sum_{t=1}^{n} I\{\zeta_i < \xi_{it}\}, \qquad \mu_{i2}^- = \sum_{t=1}^{n} I\{\zeta_i < \psi_{it}\}.$$

В данной ситуации применимы критерии, идентичные одноименным критериям предыдущего раздела:

1. $\mu_1\mu_2$-критерий из [1], основанный на суммах μ_1 и μ_2 статистик μ_{i1}^+ и μ_{i2}^+ соответственно. Он отклоняет гипотезу H_0, когда одновременно $\mu_1 > \lambda$ и $\mu_2 > \lambda$.

2. Рандомизированный по теореме 3 из [4, гл. 4] ν-критерий, основанный на числах ν_A, ν_B, ν_C событий A_i, B_i, C_i.

3. Нерандомизированный ν-критерий.

Пороговые уровни $\lambda(z)$ и вероятности $\gamma(z)$ определяются здесь точно так же, как и выше для одноименных критериев.

Как и раньше, для выбора событий A_i (B_i) удобно ввести матрицу всех возможных событий вида $\{\mu_{i1}^+ = u_i,\ \mu_{i2}^+ = v_i\}$ (соответственно вида $\{\mu_{i1}^- = u_i,\ \mu_{i2}^- = v_i\}$), где столбцы и строки пронумерованы от 0 до n. При этом по построению (u_i, v_i) будет также адресом возможного события в этой матрице, а максимальное допустимое событие A_i (аналогично B_i) представляет собой объединение событий, расположенных ниже побочной диагонали матрицы.

Для вычисления уровней значимости и мощности критериев прежде нужно определить вероятности событий $\{\mu_{i1}^+ = u,\ \mu_{i2}^+ = v\}$ и $\{\mu_{i1}^- = u,\ \mu_{i2}^- = v\}$ при H_0 и H_1. При H_0 обе вероятности равны [1]

$$\frac{(n!)^2}{(2n+1)!}\binom{u+v}{u}\binom{2n-u-v}{n-u}.$$

Из полученного далее более общего результата следует, что при H_1 для случая с распределениями (1)—(3)

$$\mathbf{P}\{\mu_{i1}^+ = u,\ \mu_{i2}^+ = v \mid H_1\} = \mathbf{P}\{\mu_{i1}^- = n-u,\ \mu_{i2}^- = n-v \mid H_1\} =$$

$$= (n!)^2 \binom{u + \mathrm{v}}{u} \binom{2n - u - \mathrm{v}}{n - u} \sum_{s=0}^{u+\mathrm{v}} c^{(u+\mathrm{v}-s)} v^{(2n-u-\mathrm{v}+s)} \times$$

$$\times \; [(u + \mathrm{v} - s)!(2n - u - \mathrm{v} + s + 1)!\,]^{-1}.$$

При $n = 1$ и $n = 2$ в качестве наиболее подходящих событий A_i и B_i оказались максимальные допустимые события, а при $n = 3$ — в точности события вида (9).

В табл. 3 представлены некоторые результаты вычислений мощности рассматриваемых здесь критериев (как и выше, столбец с $a = 0$ дает фактические уровни значимости критериев). Там, где одному и тому же значению k соответствуют три строки, первая из них принадлежит первому критерию и т. д., где лишь две строки, первая принадлежит второму критерию, вторая — третьему (в этом случае вычисление значений мощности первого критерия по точным формулам требует чрезмерно много машинного времени). Как можно увидеть из таблицы, и в этой задаче нерандомизированный ν-критерий имеет значительно лучшие результаты. Конечно, следует иметь в виду, что и ран-

Таблица 3

n	k	a								
		0,0	0,1	0,2	0,3	0,5	0,7	1,0	1,5	2,0
1	10	0,0559	0,1004	0,1634	02433	0,4329	0,6195	0,8231	0,9631	$0{,}9^{2}37$
		0,0432	0,0862	0,1499	0,2327	0,4329	0,6303	0,8407	0,9733	$0{,}9^{2}66$
		0,0432	0,0858	0,1504	0,2360	0,4467	0,6539	0,8655	0,9828	$0{,}9^{2}85$
	20	0,0362	0,0898	0,1814	0,3078	0,5962	0,8167	0,9605	$0{,}9^{2}83$	
		0,0314	0,0891	0,1943	0,3423	0,6669	0,8819	0,9857	$0{,}9^{3}85$	
		0,0314	0,0911	0,2004	0,3537	0,6839	0,8937	0,9884	$0{,}9^{4}00$	
	50	0,0040	0,0277	0,1110	0,2845	0,7275	0,9476	0,9982		
		0,0037	0,0318	0,1401	0,3636	0,8405	0,9864	$0{,}9^{4}36$		
		0,0037	0,0317	0,1398	0,3641	0,8427	0,9871	$0{,}9^{4}44$		
2	10	0,0274	0,0561	0,1050	0,1783	0,3905	0,6281	0,8743	0,9892	$0{,}9^{3}41$
		0,0255	0,0542	0,1064	0,1878	0,4227	0,6701	0,9002	0,9926	$0{,}9^{3}63$
		0,0255	0,0536	0,1053	0,1874	0,4312	0,6903	0,9199	0,9963	$0{,}9^{4}07$
	20	0,0181	0,0556	0,1373	0,2739	0,6323	0,8850	0,9904	$0{,}9^{4}55$	
		0,0176	0,0548	0,1425	0,2963	0,6918	0,9284	0,9971	$0{,}9^{5}73$	
		0,0176	0,0557	0,1461	0,3045	0,7052	0,9355	0,9977	$0{,}9^{5}84$	
	50	0,0011	0,0120	0,0792	0,2890	0,8651	0,9960	$0{,}9^{5}89$		
		0,0011	0,0121	0,0804	0,2930	0,8696	0,9963	$0{,}9^{6}14$		
3	10	0,0407	0,0827	0,1529	0,2554	0,5283	0,7802	0,9592	$0{,}9^{3}09$	
		0,0374	0,0781	0,1503	0,2589	0,5454	0,7962	0,9634	$0{,}9^{3}14$	
		0,0374	0,0798	0,1560	0,2714	0,5730	0,8243	0,9741	$0{,}9^{3}60$	
	20	0,0139	0,0467	0,1307	0,2883	0,7102	0,9455	0,9988	$0{,}9^{6}73$	
		0,0139	0,0464	0,1299	0,2884	0,7162	0,9500	$0{,}9^{3}06$	$0{,}9^{6}89$	
	50	0,0008	0,0103	0,0772	0,3036	0,9004	0,9986	$0{,}9^{7}44$		
		0,0008	0,0103	0,0775	0,3051	0,9030	0,9987	$0{,}9^{7}58$		

доминизированный ν-критерий оказался здесь более мощным, чем $\mu_1\mu_2$-критерий.

О мощности критериев для обнаружения полос. Пусть, как и в [1, 2], $\xi_{i1}, ..., \xi_{in}, \zeta_{i1}, ..., \zeta_{im}, \psi_{i1}, ..., \psi_{in}$ — величины, наблюдаемые на линии i-й нормали к средней линии проверяемого положения полосы ($i = 1, ..., k$), 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. Будем считать, что при H_0 для всех i величины $\xi_{i1}, ..., \xi_{in}$, $\zeta_{i1}, ..., \zeta_{im}, \psi_{i1}, ..., \psi_{in}$ имеют (по крайней мере, приближенно) одну и ту же функцию распределения вероятностей F_i (неизвестную наблюдателю), а при H_1 величины $\zeta_{i1}, ..., \zeta_{im}$ будут иметь новую функцию распределения G_i, причем будут стохастически больше, чем величины $\xi_{i1}, ..., \xi_{in}$ и $\psi_{i1}, ..., \psi_{in}$, имеющие, быть может, ту же функцию распределения F_i, что и при H_0.

Чтобы увидеть, что именно получается в этой более общей ситуации, и составить окончательное мнение о критериях, вновь возьмем статистики Манна — Уитни [1, 2]:

$$\mu_{i1}^+ = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} > \xi_{it}\}, \qquad \mu_{i2}^+ = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} > \psi_{it}\},$$

$$\mu_{i1}^- = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} < \xi_{it}\}, \qquad \mu_{i2}^- = \sum_{s=1}^{m} \sum_{t=1}^{n} I\{\zeta_{is} < \psi_{it}\}$$

— и рассмотрим те же самые критерии, что и в двух предшествующих разделах:

1. $\mu_1\mu_2$-критерий, основанный на суммах μ_1 и μ_2 статистик μ_{i1}^+ и μ_{i2}^+ соответственно и отклоняющий гипотезу H_0, когда одновременно $\mu_1 > \lambda$ и $\mu_2 > \lambda$.

2. Рандомизированный ν-критерий, основанный на числах ν_A, ν_B, ν_C событий A_i, B_i, C_i.

3. Нерандомизированный ν-критерий.

Здесь также для выбора событий A_i (B_i) удобно ввести матрицу всех возможных событий вида $\{\mu_{i1}^+ = u_i, \mu_{i2}^+ = v_i\}$ (соответственно вида $\{\mu_{i1}^- = u_i, \mu_{i2}^- = v_i\}$), где столбцы и строки пронумерованы от 0 до $m \times n$. При этом по построению (u_i, v_i) будет также адресом события, а максимальное допустимое событие A_i (аналогично B_i) представляет собой объединение событий, расположенных ниже побочной диагонали этой матрицы.

Получим вероятности событий $\{\mu_{i1}^+ = u_i, \mu_{i2}^+ = v_i\}$ и $\{\mu_{i1}^- = u_i, \mu_{i2}^- = v_i\}$ при H_0 и H_1, используя порядковые статистики.

Расположим элементы $\zeta_{i1}, ..., \zeta_{im}$ в порядке возрастания и обозначим полученные упорядоченные величины через $\zeta_1^*, ..., \zeta_m^*$ (для простоты индекс i здесь и далее опущен).

Пусть $x_1 < x_2 < ... < x_m$. Тогда условные вероятности $P\{\mu_{i1}^+ = u, \mu_{i2}^+ = v \mid \zeta_1^* = x_1, ..., \zeta_m^* = x_m\}$ и $P\{\mu_{i1}^- = u, \mu_{i2}^- = v \mid \zeta_1^* = x_1, ..., \zeta_m^* = x_m\}$ можно записать в виде

$$(n!)^2 \sum_{p_1(n)} \sum_{p_2(n)} \left(\prod_{j=0}^{m} n_j'!\, n_j''! \right)^{-1} \left[F(x_1)\right]^{(n_0' + n_0'')} \times$$
$$\times \left[F(x_2) - F(x_1)\right]^{(n_1' + n_1'')} \times ... \times \left[F(x_m) - F(x_{m-1})\right]^{(n_{m-1}' + n_{m-1}'')} \times$$
$$\times \left[1 - F(x_m)\right]^{(n_m' + n_m'')}, \tag{10}$$

где суммирование на этот раз выполняется по всем парам $p_1(n)$, $p_2(n)$ упорядоченных разбиений числа n в виде суммы целых неотрицательных чисел

$$n = n_0' + n_1' + \dots + n_m', \qquad n = n_0'' + n_1'' + \dots + n_m'',$$

удовлетворяющих равенствам

$$\sum_{j=0}^{m} (m-j) n_j' = u, \qquad \sum_{j=0}^{m} (m-j) n_j'' = \text{v} \tag{11}$$

или

$$\sum_{j=1}^{m} j n_j' = u, \qquad \sum_{j=1}^{m} j n_j'' = \text{v} \tag{12}$$

в соответствии с тем, вычисляется первая или вторая вероятность. Формулы (10)—(12) вытекают из следующего. Поскольку $x_1 < \dots < x_m$, то вероятность того, что среди всех n наблюдений ξ_s ровно n_0' наблюдений, а среди всех n наблюдений ψ_t ровно n_0'' будут меньше x_1, кроме того, ровно n_j' и соответственно n_j'' наблюдений будут заключены между x_j и x_{j+1} $(j = 1, \dots, m-1)$, наконец, ровно n_m' и соответственно n_m'' наблюдений будут больше x_m, при этом автоматически будем иметь:

$$\mu_{i1}^{+} = \sum_{j=0}^{m} (m-j) n_j', \qquad \mu_{i2}^{+} = \sum_{j=0}^{m} (m-j) n_j'',$$

$$\mu_{i1}^{-} = \sum_{j=1}^{m} j n_j', \qquad \mu_{i2}^{-} = \sum_{j=1}^{m} j n_j'',$$

равна (двумерное полиномиальное распределение)

$$\frac{(n!)^2}{n_0'! \dots n_m'! n_0''! \dots n_m''!} \left[F(x_1)\right]^{(n_0' + n_0'')} \left[F(x_2) - F(x_1)\right]^{(n_1' + n_1'')} \times \dots$$

$$\dots \times \left[F(x_m) - F(x_{m-1})\right]^{(n_{m-1}' + n_{m-1}'')} \left[1 - F(x_m)\right]^{(n_m' + n_m'')}.$$

Если к величинам ζ_j^* применить преобразование $\varepsilon_j = F(\zeta_j^*)$, то условные вероятности $P\{\mu_{i1}^{+} = u, \mu_{i2}^{+} = \text{v} \mid \varepsilon_1 = y_1, \dots, \varepsilon_m = y_m\}$ и $P\{\mu_{i1}^{-} = u, \mu_{i2}^{-} = \text{v} \mid \varepsilon_1 = y_1, \dots, \varepsilon_m = y_m\}$, где $y_j = F(x_j)$, также будут определяться формулой (10) с подстановкой y_j вместо $F(x_j)$. Элемент совместного распределения величин $\varepsilon_1, \dots, \varepsilon_m$ имеет вид $m! dy_1 \dots dy_m$ [7, 8], а само распределение сосредоточено в области R_m, для которой $y_j \le y_{j+1}$. Интеграл от (10) (с подстановкой y_j вместо $F(x_j)$), взятый по этой области, дает формулу как для $P\{\mu_{i1}^{+} = u, \mu_{i2}^{+} = \text{v} \mid H_0\}$, так и для $P\{\mu_{i1}^{-} = u, \mu_{i2}^{-} = \text{v} \mid H_0\}$ $(n_j = n_j' + n_j'',\ j = 0, \dots, m)$:

$$m!(n!)^2 \sum_{p_1(n)} \sum_{p_2(n)} \left(\prod_{j=0}^{m} n_j'!\, n_j''!\right)^{-1} \int_0^1 \left[1 - y_m\right]^{n_m} dy_m \times$$

$$\times \int_0^{y_m} \left[y_m - y_{m-1}\right]^{n_{m-1}} dy_{m-1} \dots \int_0^{y_3} \left[y_3 - y_2\right]^{n_2} dy_2 \times$$

$$\times \int_0^{y_2} \left[y_2 - y_1\right]^{n_1} y_1^{n_0}\, dy_1 = \frac{m!(n!)^2}{(m+2n)!} \sum_{p_1(n)} \sum_{p_2(n)} \prod_{j=0}^{m} \binom{n_j' + n_j''}{n_j'},$$

где суммирование проводится так же, как и выше, по всем парам $p_1(n)$, $p_2(n)$ упорядоченных разбиений числа n, удовлетворяющим равенствам (11) или (12) в соответствии с тем, вычисляется первая вероятность или вторая.

Чтобы найти вероятности $P\{\mu_{i1}^+ = u, \mu_{i2}^+ = \mathrm{v} \mid H_1\}$ и $P\{\mu_{i1}^- = u, \mu_{i2}^- = \mathrm{v} \mid H_1\}$, следует воспользоваться преобразованием $\varepsilon_j = G(\zeta_j^*)$. К сожалению, при этом во многих случаях (в частности, для нормального распределения) интегралы будут включать в себя «неэлементарно интегрируемые» функции.

Для случая с экспоненциальными распределениями (см. формулы (1)—(3)) на основании (10) и формулы бинома Ньютона оказалось возможным получить следующее выражение для искомых вероятностей ($n_j = n_j' + n_j''$, $j = 0, \ldots, m$):

$$m!(n!)^2 \sum_{p_1(n)} \sum_{p_2(n)} \left(\prod_{j=1}^{m} n_j'! n_j''! \right)^{-1} \sum_{s=0}^{n_0} \binom{n_0}{s} c^{n_0 - s} b^{2n - n_0 + s} \times$$

$$\times \int_0^1 \left[1 - y_m\right]^{n_m} dy_m \int_0^{y_m} \left[y_m - y_{m-1}\right]^{n_{m-1}} dy_{m-1} \times \ldots$$

$$\ldots \times \int_0^{y_3} \left[y_3 - y_2\right]^{n_2} dy_2 \int_0^{y_2} \left[y_2 - y_1\right]^{n_1} y_1^s \, dy_1 =$$

$$= m!(n!)^2 \sum_{p_1(n)} \sum_{p_2(n)} \prod_{j=0}^{m} \binom{n_j' + n_j''}{n_j'} \sum_{s=0}^{n_0' + n_0''} c^{n_0' + n_0'' - s} b^{2n - n_0' - n_0'' + s} \times$$

$$\times \left[(n_0' + n_0'' - s)!(m + 2n - n_0' - n_0'' + s)!\right]^{-1}, \tag{13}$$

где вновь суммирование проводится по всем парам $p_1(n)$, $p_2(n)$ разбиений числа n, удовлетворяющим равенствам (11), если при этом вычисляется вероятность $P\{\mu_{i1}^+ = u, \mu_{i2}^+ = \mathrm{v} \mid H_1\}$, и равенствам (12), если вычисляется вероятность $P\{\mu_{i1}^- = u, \mu_{i2}^- = \mathrm{v} \mid H_1\}$.

Остановимся на частном случае $n = 1$. Здесь только одно из чисел n_j' (аналогично n_j'') каждого разбиения числа n равно 1, а остальные равны 0. Отсюда числа u (аналогично v) и n_j в (11) (аналогично в (12)) однозначно определены. Поэтому, согласно (13), получаем, в частности, следующие (отчасти неожиданные) свойства вероятностей: $P\{\mu_{i1}^+ = u, \mu_{i2}^+ = u \mid H_1\} = P\{\mu_{i1}^+ = \mathrm{v}, \mu_{i2}^+ = \mathrm{v} \mid H_1\} = 2P\{\mu_{i1}^+ = u, \mu_{i2}^+ = \mathrm{v} \mid H_1\}$ и $P\{\mu_{i1}^+ = m, \mu_{i2}^+ = u \mid H_1\} = P\{\mu_{i1}^+ = m, \mu_{i2}^+ = \mathrm{v} \mid H_1\}$ при $u, \mathrm{v} < m$ и $u \neq \mathrm{v}$.

Отсюда вытекает, что в случае экспоненциальных распределений наблюдаемых величин и $n = 1$ в качестве подходящего события A_i (аналогично B_i) следует взять множество, состоящее из крайних элементов вышеуказанной матрицы всех возможных событий и максимального допустимого события A_i, т. е. из событий вида $\{\mu_{i1}^+ = u_i, \mu_{i2}^+ = m\}$ и $\{\mu_{i1}^+ = m, \mu_{i2}^+ = u_i\}$, где $1 \leq u_i \leq m$.

Полученные значения мощности и соответствующие уровни значимости критериев (колонка с $a = 0$) для различных значений m, n, k и a представлены в табл. 4. Для каждого значения k первая строка принадлежит первому критерию и т. д. При $m = n = 2$ подходящее событие A_i (аналогично B_i) совпало с рекомендованным в [2] событием $\{\mu_{i1}^+ > 2, \mu_{i2}^+ > 2\}$, пополненным двумя событиями $\{\mu_{i1}^+ = 2, \mu_{i2}^+ = 4\}$ и $\{\mu_{i1}^+ = 4, \mu_{i2}^+ = 2\}$. Сравнение значений мощности в табл. 4 показывает, что из всех трех критериев нерандомизированный ν-критерий является наиболее мощным. Его превосходство может быть значительным уже при $k = 5$. Из таблицы также видно, что с увеличением k и рандомизированный ν-критерий становится более мощным, чем $\mu_1\mu_2$-критерий.

Таблица 4

m	n	k	a								
			0,0	0,1	0,2	0,3	0,5	0,7	1,0	1,5	2,0
2	1	5	0,0680	0,1145	0,1743	0,2442	0,3977	0,5453	0,7220	0,8898	0,9587
			0,0680	0,1173	0,1775	0,2452	0,3893	0,5267	0,6956	0,8682	0,9467
			0,0658	0,1189	0,1853	0,2606	0,4200	0,5685	0,7418	0,9017	0,9647
		10	0,0465	0,1034	0,1891	0,2972	0,5327	0,7284	0,8987	0,9846	0,9979
			0,0432	0,1027	0,1911	0,3005	0,5347	0,7288	0,8994	0,9858	0,9984
			0,0432	0,1071	0,2037	0,3233	0,5742	0,7712	0,9280	0,9929	$0{,}9^348$
		20	0,0226	0,0803	0,1951	0,3573	0,6899	0,8892	0,9835	$0{,}9^358$	
			0,0222	0,0874	0,2165	0,3935	0,7347	0,9192	0,9917	$0{,}9^417$	
			0,0222	0,0896	0,2243	0,4081	0,7537	0,9304	0,9939	$0{,}9^456$	
3	1	5	0,0663	0,1207	0,1910	0,2721	0,4436	0,5992	0,7720	0,9189	0,9721
			0,0663	0,1224	0,1918	0,2691	0,4285	0,5227	0,7388	0,8943	0,9593
			0,0618	0,1262	0,2058	0,2930	0,4667	0,6164	0,7786	0,9176	0,9703
		10	0,0403	0,1022	0,1989	0,3203	0,5737	0,7686	0,9222	0,9900	0,9988
			0,0403	0,1086	0,2118	0,3365	0,5870	0,7754	0,9239	0,9902	0,9989
			0,0388	0,1142	0,2315	0,3727	0,6439	0,8289	0,9539	0,9963	$0{,}9^376$
4	1	5	0,0644	0,1236	0,2003	0,2880	0,4695	0,6287	0,7981	0,9334	0,9786
			0,0644	0,1245	0,1999	0,2839	0,4537	0,6023	0,7659	0,9098	0,9664
			0,0612	0,1474	0,2557	0,3710	0,5820	0,7400	0,8816	0,9711	0,9933
2	2	5	0,0752	0,1344	0,2161	0,3158	0,5330	0,7196	0,8903	0,9823	0,9975
			0,0752	0,1273	0,1967	0,2803	0,4676	0,6433	0,8311	0,9620	0,9926
			0,0612	0,1160	0,1973	0,3011	0,5348	0,7358	0,9102	0,9899	0,9991

О мощности критериев для обнаружения гребней (хребтов). Как и в [1, 2], рассмотрим простейший вариант задачи обнаружения хребтов: с минимальным числом наблюдений на области склонов хребта, а в конце дадим критерий для общего случая.

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 ζ_i — в точке на самом проверяемом положении гребня, а ξ_{i1}, ξ_{i2} и ψ_{i1}, ψ_{i2} — по разные стороны от последнего (на склонах). Будем считать, что при отсутствии объекта (при H_0) для каждого i величины $\xi_{i1}, \xi_{i2}, \zeta_i, \psi_{i1}, \psi_{i2}$ имеют одну и ту же функцию распределения вероятностей F_i (стохастически равны), а при наличии на проверяемом положении гребня (при H_1) — стохастически $\zeta_i > \xi_{i1} > \xi_{i2}$ и $\zeta_i > \psi_{i1} > \psi_{i2}$.

Сравнение рассматриваемых здесь критериев ведется для случая, когда при наличии гребня величины ζ_i имеют распределение

$$1 - \exp(-(x - 2a)), \qquad a > 0, \ \ x \geq 2a,$$

величины ξ_{i1} и ψ_{i1} — распределение

$$1 - \exp(-(x - a)), \qquad a > 0, \ \ x \geq a,$$

а величины ξ_{i2} и ψ_{i2} — распределение $1 - \exp(-x)$ (см. формулы (1)—(3)). При H_0 $a = 0$.

Рассмотрим вначале следующие критерии:

1) рандомизированный ν-критерий, основанный на числах ν_A, ν_B, ν_C событий

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

$$B_i = \{\zeta_i < \xi_{i1} < \xi_{i2},\ \zeta_i < \psi_{i1} < \psi_{i2}\},$$

$$C_i = \overline{A_i \cup B_i};$$

2) нерандомизированный ν-критерий, основанный на тех же числах ν_A, ν_B, ν_C, что и рандомизированный;

3) составной $\mu_1^*\mu_2^*$-критерий [1], основанный на статистиках

$$\mu_j = \sum_{i=1}^{k} \mu_{ij} \qquad (j = 1, ..., 4).$$

Здесь $\mu_{i1} = I\{\zeta_i > \xi_{i1}\}$, $\mu_{i2} = I\{\zeta_i > \psi_{i1}\}$, $\mu_{i3} = I\{\xi_{i1} > \xi_{i2}\}$, $\mu_{i4} = I\{\psi_{i1} > \psi_{i2}\}$. Он отклоняет гипотезу H_0, когда одновременно $\mu_1^* = \min(\mu_1, \mu_2) > \lambda_1$ и $\mu_2^* = \min(\mu_3, \mu_4) > \lambda_2$, где λ_1 и λ_2 разумно выбирать так, чтобы обеспечить заданный уровень значимости критерия и равенство $P(\mu_1^* > \lambda_1 \mid H_0) \cong P(\mu_2^* > \lambda_2 \mid H_0)$.

Результаты численных исследований этих и других критериев, рассматриваемых ниже, представлены в табл. 5. Для $k = 9$ в первой строке даны значения мощности первого критерия, во второй — второго и т. д. При $k = 15$ отсутствуют значения мощности $\mu_1^*\mu_2^*$-критерия, вычисление которых на ЭВМ требует чрезмерно много времени. Значения при $a = 0$ являются фактическими уровнями значимости критериев. Из таблицы видно, что рандомизированный ν-критерий наименее мощный. Недостаток мощности обоих ν-критериев, особенно при малых значениях a, связан здесь с тем фактом, что в данном случае вероятность «мешающих» событий C_i слишком велика. В частности, из формулы (16) работы [1] следует, что $P\{C_i \mid H_0\} = 0{,}9$.

Результаты предыдущих разделов побудили изучить еще следующие составные критерии:

4) $\nu_1\mu_2^*$-критерий, основанный на числах $\nu_{A_1}, \nu_{B_1}, \nu_{C_1}$ событий

$$A_{i1} = \{\zeta_i > \xi_{i1}, \zeta_i > \psi_{i1}\},\quad B_{i1} = \{\zeta_i < \xi_{i1}, \zeta_i < \psi_{i1}\},\quad C_{i1} = \overline{A_{i1} \cup B_{i1}} \qquad (14)$$

Таблица 5

k	a								
	0,0	0,1	0,2	0,3	0,5	0,7	1,0	1,5	2,0
9	0,0600	0,0890	0,1237	0,1672	0,2894	0,4545	0,7098	0,9423	0,9929
	0,0493	0,1098	0,2001	0,3151	0,5715	0,7823	0,9436	0,9969	$0{,}9^389$
	0,0517	0,1269	0,2407	0,3777	0,6409	0,8208	0,9461	0,9941	$0{,}9^343$
	0,0415	0,1125	0,2284	0,3742	0,6585	0,8444	0,9602	0,9965	$0{,}9^370$
	0,0370	0,1094	0,2342	0,3942	0,6988	0,8809	0,9775	0,9991	$0{,}9^474$
	0,0386	0,1335	0,2794	0,4452	0,7292	0,8912	0,9780	0,9989	$0{,}9^448$
15	0,0250	0,0482	0,0829	0,1346	0,3054	0,5442	0,8454	0,9911	$0{,}9^379$
	0,0210	0,0677	0,1581	0,2925	0,6146	0,8533	0,9809	$0{,}9^379$	$0{,}9^607$
	0,0241	0,0980	0,2460	0,4410	0,7770	0,9342	0,9922	$0{,}9^383$	$0{,}9^562$
	0,0192	0,0883	0,2411	0,4515	0,8064	0,9532	0,9965	$0{,}9^478$	$0{,}9^727$
	0,0182	0,1057	0,2849	0,5022	0,8263	0,9548	0,9958	$0{,}9^461$	$0{,}9^679$

и статистике μ_2^* $\mu_1^*\mu_2^*$-критерия. Он отклоняет гипотезу H_0, когда одновременно $\nu_{A_1} > \lambda_1(k - \nu_{C_1})$ и $\mu_2^* > \lambda_2$, где $\lambda_1(z)$ — наименьшее целое число, такое что

$$\sum_{i=\lambda_1+1}^{z} \binom{z}{i} 2^{-z} \le \alpha_1.$$

Здесь α_1 и λ_2 выбирались так, чтобы получить нужный уровень значимости критерия и равенство

$$P(\nu_{A_1} > \lambda_1(k - \nu_{C_1}) \mid H_0) \simeq P(\mu_2^* > \lambda_2 \mid H_0).$$

5) $\nu_1\nu_2$-критерий, основанный на числах $\nu_{A_1}, \nu_{B_1}, \nu_{C_1}$ событий (14) и числах $\nu_{A_2}, \nu_{B_2}, \nu_{C_2}$ событий

$$A_{i2} = \{\xi_{i1} > \xi_{i2},\ \psi_{i1} > \psi_{i2}\},\quad B_{i2} = \{\xi_{i1} < \xi_{i2},\ \psi_{i1} < \psi_{i2}\},\quad C_{i2} = \overline{A_{i2} \cup B_{i2}}. \quad (15)$$

Он отклоняет гипотезу H_0, когда одновременно $\nu_{A_1} > \lambda_1(k - \nu_{C_1})$ и $\nu_{A_2} > \lambda_2(k - \nu_{C_2})$, где $\lambda_j(z)$ $(j = 1, 2)$ — наименьшее целое число, такое что

$$\sum_{i=\lambda_j+1}^{z} \binom{z}{i} 2^{-z} \le \alpha_j \qquad (j = 1, 2). \quad (16)$$

И здесь α_1 и α_2 выбирались аналогично так, чтобы получить подходящий уровень значимости критерия и равенство

$$P(\nu_{A_1} > \lambda_1(k - \nu_{C_1}) \mid H_0) \simeq P(\nu_{A_2} > \lambda_2(k - \nu_{C_2}) \mid H_0).$$

6) $\nu_1^*\nu_2$-критерий, основанный на числах $\nu_{A_1}^*, \nu_{B_1}^*, \nu_{C_1}^*$ событий

$$A_{i1} = \{\zeta_i > \max(\xi_{i1}, \xi_{i2}, \psi_{i1}, \psi_{i2})\},$$

$$B_{i1} = \{\zeta_i < \min(\xi_{i1}, \xi_{i2}, \psi_{i1}, \psi_{i2})\}, \qquad C_{i1} = \overline{A_{i1} \cup B_{i1}}$$

и числах $\nu_{A_2}, \nu_{B_2}, \nu_{C_2}$ событий (15). Он отклоняет гипотезу H_0, когда одновременно $\nu_{A_1}^* > \lambda_1(k - \nu_{C_1})$ и $\nu_{A_2} > \lambda_2(k - \nu_{C_2})$, где $\lambda_1(z)$ и $\lambda_2(z)$ определялись аналогично тому, как и в $\nu_1\nu_2$-критерии.

Уровень значимости и мощность каждого из этих критериев можно найти, исходя из распределения пары векторов $r_i^1 = (r_{i1}^1, r_{i2}^1, r_{i3}^1)$ и $r_i^2 = (r_{i1}^2, r_{i2}^2, r_{i3}^2)$ рангов элементов в совокупности $(\zeta_i, \xi_{i1}, \xi_{i2})$ и $(\zeta_i, \psi_{i1}, \psi_{i2})$ соответственно. Каждый вектор такой пары может быть равен любой из 3! перестановок трех чисел: 1, 2, 3. Вероятность каждого из $(3!)^2$ значений пары можно получить по схеме вывода формулы (16) в [1]. Далее, вероятность события $E_{i1} \cap E_{i2}$, где E_{ij}, например, в случае $\nu_1\nu_2$-критерия — любое из событий A_{ij}, B_{ij} и C_{ij}, равна сумме вероятностей значений пар векторов рангов (r_i^1, r_i^2), дающих событие $E_{i1} \cap E_{i2}$. Теперь для вычисления уровня значимости или соответственно мощности критерия достаточно все события $E_{i1} \cap E_{i2}$ занумеровать и перейти к полиномиальному распределению.

Полученные числовые результаты, представленные в табл. 5, показывают, что среди рассмотренных нет критерия, который был бы наиболее мощным при всех значениях a. Однако при малых значениях a последний критерий имеет большую мощность, чем остальные, и этим нельзя пренебречь, а при больших значениях a его мощность незначимо меньше только мощности $\nu_1\nu_2$-критерия.

Другое чрезвычайно важное преимущество $\nu_1^*\nu_2$-критерия обусловлено тем, что при гипотезе H_0 его тройки событий (A_{i1}, B_{i1}, C_{i1}) и (A_{i2}, B_{i2}, C_{i2}) статистически независимы. Действительно, пусть r_ξ — ранг ξ_{i1} в совокупности

(ξ_{i1}, ξ_{i2}), r_ψ — ранг ψ_{i1} в совокупности (ψ_{i1}, ψ_{i2}), наконец, r'_ζ и r''_ζ — ранги ζ_i в совокупности $(\zeta_i, \xi_{i1}, \xi_{i2})$ и $(\zeta_i, \psi_{i1}, \psi_{i2})$ соответственно. В силу предположения независимости наблюдаемых величин и теоремы Б из [9, с. 170] пары (r'_ζ, r''_ζ) и (r_ξ, r_ψ) независимы при H_0, а потому независимы пары событий $(A_{i1} = \{r'_\zeta = 3, r''_\zeta = 3\}$, $B_{i1} = \{r'_\zeta = 1, r''_\zeta = 1\})$ и $(A_{i2} = \{r_\xi = 2, r_\psi = 2\}$, $B_{i2} = \{r_\xi = 1, r_\psi = 1\})$ и, значит, также независимы события C_{i1} и C_{i2}, порожденные независимыми парами событий. Независимость троек событий доказана.

Отсюда, во-первых, уровень значимости $\nu_1^*\nu_2$-критерия равен просто произведению уровней значимости ν_1^*-критерия и ν_2-критерия

$$P(\nu^*_{A_1} > \lambda_1(k - \nu_{C_1}),\ \nu_{A_2} > \lambda_2(k - \nu_{C_2}) \mid H_0) =$$

$$= P(\nu^*_{A_1} > \lambda_1(k - \nu_{C_1}) \mid H_0)P(\nu_{A_2} > \lambda_2(k - \nu_{C_2}) \mid H_0),$$

и поэтому его существенно легче вычислить, чем уровень значимости $\nu_1\nu_2$-критерия, точное вычисление которого с ростом k быстро становится довольно трудным. Во-вторых, что особенно важно, $\nu_1^*\nu_2$-критерий можно обобщить на случай большего числа наблюдений на склонах хребта.

Например, если для $i = 1, ..., k$ наблюдаются величины ξ_{i1}, ξ_{i2}, ξ_{i3}, ζ_i, ψ_{i1}, ψ_{i2}, ψ_{i3}, то подходящим будет составной критерий, основанный на числах $\nu^*_{A_1}$, $\nu^*_{B_1}$, $\nu^*_{C_1}$ событий

$$A_{i1} = \{\zeta_i > \max(\xi_{i1}, ..., \xi_{i3},\ \psi_{i1}, ..., \psi_{i3})\},$$

$$B_{i1} = \{\zeta_i < \min(\xi_{i1}, ..., \xi_{i3},\ \psi_{i1}, ..., \psi_{i3})\}, \qquad C_{i1} = \overline{A_{i1} \cup B_{i1}},$$

числах $\nu^*_{A_2}$, $\nu^*_{B_2}$, $\nu^*_{C_2}$ событий

$$A_{i2} = \{\xi_{i1} > \max(\xi_{i2}, \xi_{i3}),\ \psi_{i1} > \max(\psi_{i2}, \psi_{i3})\},$$

$$B_{i2} = \{\xi_{i1} < \min(\xi_{i2}, \xi_{i3}),\ \psi_{i1} < \min(\psi_{i2}, \psi_{i3})\}, \qquad C_{i2} = \overline{A_{i2} \cup B_{i2}}$$

и числах ν_{A_3}, ν_{B_3}, ν_{C_3} событий

$$A_{i3} = \{\xi_{i2} > \xi_{i3},\ \psi_{i2} > \psi_{i3}\},\ \ B_{i3} = \{\xi_{i2} < \xi_{i3},\ \psi_{i2} < \psi_{i3}\},\ \ C_{i3} = \overline{A_{i3} \cup B_{i3}}.$$

Он отклоняет гипотезу H_0, когда одновременно $\nu^*_{A_1} > \lambda_1(k - \nu_{C_1})$, $\nu^*_{A_2} > \lambda_2(k - \nu_{C_2})$ и $\nu_{A_3} > \lambda_3(k - \nu_{C_3})$, где $\lambda_1(z)$, $\lambda_2(z)$ и $\lambda_3(z)$, как и раньше, определяются по (16), а уровень значимости критерия равен просто произведению уровней значимости составляющих его критериев. Доказательство последнего является очевидным обобщением приведенного доказательства для простого случая.

Заключение. Выведенные в настоящей работе явные формулы для совместных распределений двух статистик Манна — Уитни позволяют вычислять фактические уровни значимости, а в случае экспоненциальных распределений наблюдаемых величин — и мощность критериев.

При сравнении результатов вычислений найдено, что лишь в одной задаче обнаружения контура объекта критерий, использующий просто сумму статистик Манна — Уитни в блоках, является более мощным, чем ν-критерий, основанный на числах реализовавшихся событий A_i, B_i, C_i, определяемых значениями этих статистик.

Для других же задач обнаружения, особенно когда число k блоков велико, ν-критерии могут быть значительно мощнее, имея решительное преимущество и с точки зрения вычислений фактического уровня значимости и мощности. Полученные результаты могут оказаться полезными также при решении следующих вопросов:

1) Какое число блоков должно быть использовано и сколько наблюдаемых величин следует взять в каждом отдельном блоке?

2) Как выбирать события A_i, B_i, C_i, от которых сильно зависит мощность ν-критерия, в частности, когда следует урезать максимальные допустимые события A_i и B_i?

3) Стоит ли вводить рандомизацию в ν-критерий для «полного использования» заданного уровня значимости критерия?

СПИСОК ЛИТЕРАТУРЫ

1. **Салов Г. И.** Непараметрические критерии для обнаружения контуров, линий, полос и хребтов заданной формы на случайном фоне // Автометрия. 1994. № 1. С. 48.
2. **Салов Г. И.** Метод получения равномерно наиболее мощных критериев для обнаружения протяженных объектов на случайном фоне // Автометрия. 1995. № 1. С. 34.
3. **Кендалл М. Дж., Стьюарт А.** Статистические выводы и связи. М.: Наука, 1973.
4. **Леман Э.** Проверка статистических гипотез. М.: Наука, 1964.
5. **Красильников Н. Н.** Теория передачи и восприятия изображений. М.: Радио и связь, 1986.
6. **Прэтт У.** Цифровая обработка изображений. Кн. 1, 2. М.: Мир, 1982.
7. **Кендалл М. Дж., Стьюарт А.** Теория распределений. М.: Наука, 1966.
8. **Уилкс С.** Математическая статистика. М.: Наука, 1967.
9. **Гаек Я., Шидак З.** Теория ранговых критериев. М.: Наука, 1971.

Поступила в редакцию 10 сентября 1996 г.