

УДК 621.396

ББК 22.172

## ПРИМЕНЕНИЕ ПОРЯДКОВЫХ СТАТИСТИК В ЗАДАЧАХ ОБНАРУЖЕНИЯ

Рудько И. М.<sup>1</sup>

(Учреждение Российской академии наук  
Институт проблем управления РАН, Москва)

*Введено понятие усеченной порядковой статистики. Выведены аналитические выражения для математических ожиданий и дисперсий усеченных порядковых статистик для порога отсечения, фиксированного по заданному номеру члена вариационного ряда. Для систем обнаружения, реализующих алгоритм проверки двух статистических гипотез и использующих энергетический критерий обнаружения, разработан алгоритм обнаружения сигналов на фоне шума, основанный на статистических свойствах усеченных порядковых статистик.*

Ключевые слова: порядковая статистика, проверка статистических гипотез, системы обнаружения, вероятности обнаружения и ложной тревоги, математическое моделирование.

### 1. Введение

Во многих системах обработки локационной информации, работающих в пассивном режиме, системах спектрального анализа и анализа вибраций решаются задачи обнаружения сигнала на фоне помехи, причем статистические свойства сигнала и помехи одинаковы и единственным их отличием являются энергии (дисперсии). В простейшей форме операция обнаружения – это задача проверки двух статистических гипотез: нулевой гипотезы  $H_0$ , когда данные относятся только к шуму, и

---

<sup>1</sup> Игорь Михайлович Рудько, кандидат технических наук, с.н.с.  
(igor-rudko@mail.ru).

альтернативной гипотезы  $H_1$ , когда данные относятся к совместному воздействию сигнала и шума. При решении задачи обнаружения приемник вычисляет отношение правдоподобия, которое представляет собой отношение плотностей распределения вероятностей для гипотез  $H_1$  и  $H_0$ .

При этом, например, в пассивных локационных системах [1] сигнал появляется в случайные моменты времени, его продолжительность также случайна, известными параметрами являются только статистические свойства помехи, и даже энергетические спектры шума и сигнала близки по форме. Модель обнаружения в этих задачах обычно представляется как энергетический порог, установленный над средним значением помехи.

В системах спектрального анализа бывает необходимо решать задачу поиска «звукорядов» [6], т.е. узкополосных компонент энергетического спектра, кратных основной частоте. Так как очень часто «звукоряд» является неполным, т.е. в нем отсутствуют многие составляющие (но шумовая составляющая присутствует), обычно производится поиск отдельных узкополосных компонент, из которых затем формируют «звукоряд».

Таким образом, и в первой, и во второй из рассмотренных выше задач представляют интерес лишь случаи превышения порога, а шумовой фон лишь «забывает» тракт обработки, особенно при принятии решения оператором. Поэтому представляет интерес создание алгоритмов, осуществляющих поиск сигнала непосредственно, с отбрасыванием шумовых составляющих.

## **2. Усеченные порядковые статистики**

### **2.1. МОМЕНТЫ УСЕЧЕННОЙ ПОРЯДКОВОЙ СТАТИСТИКИ**

Рассмотрим выборку, состоящую из  $m$  случайных величин  $X_i$ :  $\{X_1, \dots, X_i, \dots, X_m\}$ . Пусть случайная величина  $X_i$  описывается плотностью распределения:

$$(1) \quad f(x) = \frac{1}{\sigma^2} k_n \left( \frac{x}{\sigma^2} \right), \quad x \geq 0,$$

где  $k_n(\cdot)$  – плотность центрального  $\chi^2$ -распределения с  $n$  степенями свободы;  $\sigma^2$  – дисперсия. Функция распределения случайной величины  $X_i$  описывается формулой

$$(2) \quad K_n(x) = \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2\sigma^2} \int_0^x t^{n/2-1} e^{-t/2\sigma^2} dt.$$

Сравним статистические свойства случайных величин  $Z$  и  $W$ , сформированных из случайной выборки  $X_i$ ,  $1 \leq i \leq m$ , двумя различными способами:

- $Z = \sum_{i=1}^m X_i$  – имеет центральное  $\chi^2$ -распределение с  $nm$  степенями свободы и математическим ожиданием

(3)  $\mu_Z = nm\sigma^2$

и дисперсией

$$(4) \quad \sigma_Z^2 = 2nm\sigma^4,$$

и в силу центральной предельной теоремы для достаточно больших значений  $m$  ее функция плотности распределения нормализуется

$$Z \sim N(\mu_Z, \sigma_Z^2).$$

- $W = \sum_{i=1}^m X_{(i)}$ , где  $X_{(i)}$ ,  $1 \leq i \leq m$ , – упорядоченные величины

(порядковые статистики) статистики  $X_i$ , такие, что  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(i)} \leq \dots \leq X_{(m)}$ . Если случайные величины  $X_i$  статистически независимы и одинаково распределены, то случайные величины  $X_{(i)}$  зависимы из-за неравенств между ними. В дальнейшем будем называть статистику  $W$  усеченной порядковой статистикой (УПС), а параметр  $k$  – порогом отсечения.

В работе [3] приведены выражения для вычисления моментов порядковых статистик, которые в общем случае имеют следующий вид:

– математическое ожидание  $\mu_j$  величины  $X_{(j)}$  определяется по формуле

$$\mu_j = \int_{-\infty}^{\infty} x f_j(x) dx,$$

– дисперсия

$$\sigma_j^2 = \int_{-\infty}^{\infty} (x - \mu_j)^2 f_j(x) dx,$$

– ковариация

$$\sigma_{jk} = \int_{-\infty}^{\infty} \int_{-\infty}^y (x - \mu_j)(y - \mu_k) f_{jk}(x, y) dx dy, \quad 1 \leq j < k \leq m,$$

где плотность распределения величины  $X_{(j)}$ :

$$f_j(x) = \frac{m!}{(m-j)!(j-1)!} P(x)^{j-1} [1 - P(x)]^{m-j} p(x),$$

совместная плотность распределения величин  $X_{(j)}, X_{(k)}$ :

$$f_{jk}(x, y) = \frac{m! p(x) p(y)}{(m-k)!(k-j-1)!(j-1)!} \times \\ \times P(x)^{j-1} [P(y) - P(x)]^{k-j-1} [1 - P(y)]^{m-k},$$

а  $p(\cdot)$  и  $P(\cdot)$  – плотность и функция распределения случайной величины соответственно.

Для выборки, описываемой формулами (1) и (2), т.е. плотностью распределения  $f(x)$  и функцией распределения  $K_n(x)$ , формулы для определения  $\mu_j, \sigma_j^2$  и  $\sigma_{jk}$  приобретают следующий вид:

$$(5) \quad \mu_j = \frac{m!}{(j-1)!(m-j)!} \int_0^{\infty} K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[1 - K_n\left(\frac{x}{\sigma^2}\right)\right]^{m-j} x f(x) dx$$

$$\sigma_j^2 = \frac{m!}{(j-1)!(m-j)!} \times,$$

$$\times \int_0^{\infty} K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[1 - K_n\left(\frac{x}{\sigma^2}\right)\right]^{m-j} (x - \mu_j)^2 f(x) dx,$$

$$\sigma_{jk} = E[X_{(j)} X_{(k)}] =$$

$$= \frac{m!}{(m-k)!(k-j-1)!(j-1)!} \int_0^{\infty} \left[ \int_0^y C(x, y) x y f(x) f(y) dx \right] dy,$$

где

$$C(x, y) = K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[ K_n\left(\frac{y}{\sigma^2}\right) - K_n\left(\frac{x}{\sigma^2}\right) \right]^{k-j-1} \left[ 1 - K_n\left(\frac{y}{\sigma^2}\right) \right]^{m-k}.$$

Для случайной величины  $W$  математическое ожидание определяется по формуле

$$(6) \quad \mu_W(k) = \sum_{j=k}^m \mu_j, \quad 1 \leq k \leq m,$$

а дисперсия с учетом зависимости случайных величин  $X_{(i)}$  [4]:

$$(7) \quad \sigma_W^2(k) = \sum_{l=k}^m \sigma_l^2 + 2 \sum_{k \leq j < l \leq m} \sigma_{jl}, \quad 1 \leq k \leq m,$$

и в силу центральной предельной теоремы при достаточно больших значениях  $m$  ее функция плотности распределения также нормализуется:

$$W \sim N(\mu_W, \sigma_W^2).$$

Аналитические исследования статистических свойств случайной величины  $W$  очень сложны, и поэтому большинство последующих результатов получено на ПК с использованием символьного программирования в среде *MATLAB*.

На рис. 1 приведены рассчитанные по формулам (6) и (7) зависимости математических ожиданий  $\mu_W(k)$  и с.к.о.  $\sigma_W(k)$  от порога при следующих значениях параметров:  $n = 8$ ,  $m = 12$ ,  $\sigma = 1$ . Для случайной величины  $W$  порог  $k$  дискретный,  $1 \leq k \leq m$ . Из графиков видно, что  $\mu_W(k)$  и  $\sigma_W(k)$  монотонно убывающие функции, причем  $\mu_W(1) = \mu_Z = nm$  и  $\sigma_W(1) = \sigma_Z = (2nm)^{1/2}$ .

На рис. 2 приведены рассчитанные по формулам (3), (4), (6) и (7) плотности распределения случайных величин  $Z$  и  $W$  и гистограммы этих плотностей, полученные на моделях с параметрами  $n = 8$ ,  $m = 100$ ,  $\sigma = 1$ . (Размеры массивов для построения гистограмм – 5000). Наблюдается достаточно близкое совпадение между теоретическими и модельными результатами.

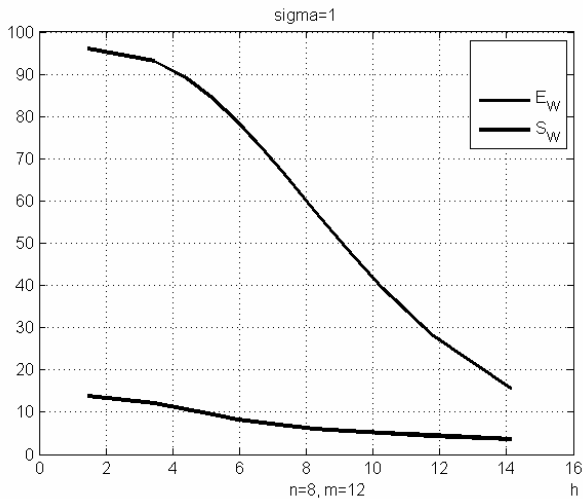


Рис. 1. Зависимости математических ожиданий  $\mu_W(k)$  (на рис. –  $E_W$ ) и с.к.о.  $\sigma_W(k)$  (на рис. –  $S_W$ ) от порога  $k$

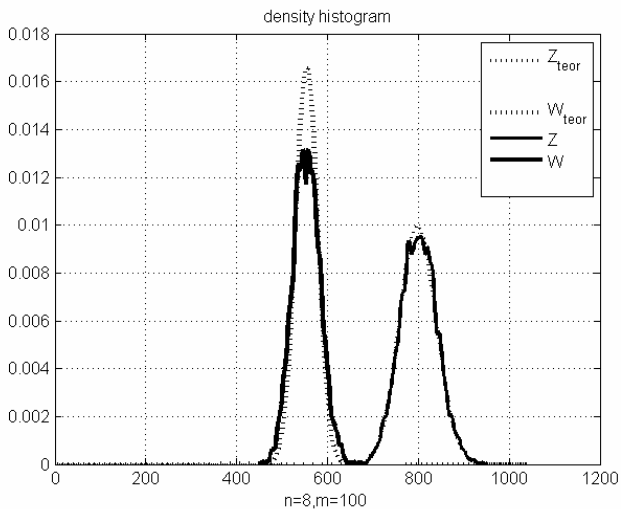


Рис. 2. Плотности распределения (на рис. –  $Z_{teor}$  и  $W_{teor}$ ) и гистограммы случайных величин  $Z$  и  $W$

## 2.2. ФОРМУЛЫ АППРОКСИМАЦИИ ДЛЯ УПС

Точный расчет математического ожидания  $\mu_w$  и дисперсии  $\sigma_w$  требует очень большого машинного времени (например, уже при  $n = 8$  и  $m = 12$  время вычисления  $\mu_w(k)$  и  $\sigma_w(k)$  составляет около 2 часов). Поэтому при больших значениях параметров  $n$  и  $m$  интерес представляют формулы аппроксимации для вычисления  $\mu_w(k)$  и  $\sigma_w(k)$ .

Известно [3], что при больших значениях  $m$  достаточно точным приближением к  $\mu_j$  является значение, удовлетворяющее уравнению

$$P(x) = \frac{j}{m+1}.$$

Если ввести в рассмотрение функцию  $Q$ , такую, что  $Q[P(x)] = x$ , то получим асимптотическую формулу

$$(8) \quad \mu_j \sim Q\left(\frac{j}{m+1}\right),$$

где  $1 \leq j \leq m$ , т.е. математические ожидания  $\mu_j$  приближенно равны квантилям функции распределения  $P(x)$ , взятым с равномерным шагом. Математические ожидания  $\mu_w(k)$  определяются по формуле (6). На рис. 3 приведены значения  $\mu_j$ , рассчитанные по формулам (5) и (8), из которых видно, что расхождения между значениями  $\mu_j$ , рассчитанными по аппроксимирующей формуле (8), и точными значениями (5) наблюдаются только на «хвостах».

Как уже указывалось выше, для малых значений  $m$  ( $m \leq 12$ ) можно рассчитать точные значения  $\mu_w(k)$  и  $\sigma_w^2(k)$  для  $1 \leq k \leq m$ . Как следует из формулы аппроксимации (8), при увеличении  $m$  достаточно рассчитать большее число квантилей  $\mu_j$ , по которым, используя формулу (6), можно подсчитать приблизительные значения для  $\mu_w(k)$ .

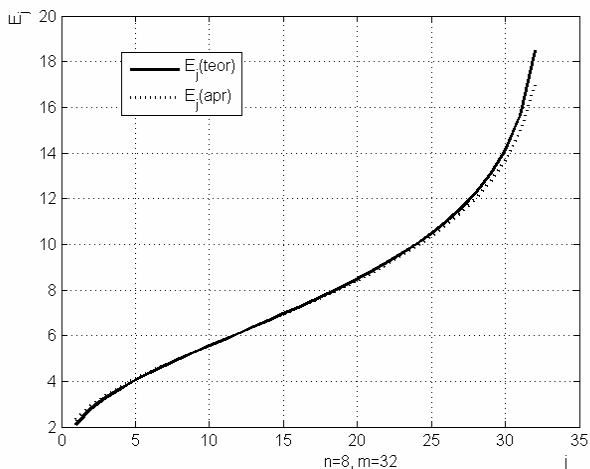


Рис. 3. Значения  $\mu_j$ , рассчитанные по точной (на рис. –  $E_{(teor)}$ ) и по аппроксимирующей (на рис. –  $E_{(apr)}$ ) формулам

При аппроксимации дисперсии  $\sigma_w^2(j)$  для больших значений  $m$  практически возможно рассчитать только крайние точки функции дисперсии  $\sigma_w^2(j)$ , а именно,

$$\sigma_w^2(1) = 2nm ,$$

$$\sigma_w^2(m) = \sigma_m^2 = m \int_0^{\infty} K_n \left( \frac{x}{\sigma^2} \right)^{m-1} (x - \mu_m)^2 f(x) dx ,$$

где  $\mu_m$  определяется по формуле (5), и несколько промежуточных точек  $\sigma_w^2(k)$  для малого числа  $m_k$  ( $m_k \ll m$ ). Аппроксимацию проще всего проводить, используя значения в этих точках и учитывая, что  $\sigma_w^2(k)$  монотонно убывающая функция.

Рассмотрим аппроксимацию экспоненциальной функцией  $y = Y_0 e^{-\alpha x}$ .

Пусть  $y_1 = Y_0 \exp(-\alpha x_1)$  и  $y_2 = Y_0 \exp(-\alpha x_2)$ . Тогда  $y_1 \exp(-\alpha x_2) = y_2 \exp(-\alpha x_1)$  и

$$\alpha = \frac{\ln y_1 - \ln y_2}{x_2 - x_1} ,$$

а формула аппроксимации приобретает вид:



$$\tilde{\sigma}_W^2(j) = \sigma_W^2(k) \exp(-\alpha_k \mu_j),$$

где  $\alpha_k = \frac{\ln \sigma_W^2(k) - \ln \sigma_W^2(k+1)}{\mu_{k+1} - \mu_k},$

а  $k$  выбирается из условия  $k/m_k < j/m < (k+1)/m_k.$

В случае простейшей аппроксимации (аппроксимации по двум крайним точкам)

$$(9) \quad \tilde{\sigma}_W^2(j) = nm \exp(-\alpha \mu_j),$$

где  $\alpha = \frac{\ln \sigma_W^2(1) - \ln \sigma_W^2(m)}{\mu_m - \mu_1}.$

На рис. 4 приведены значения  $\sigma_W^2(j)$ , рассчитанные по формулам (7) и (9), из которых видно, что значения  $\tilde{\sigma}_W^2(j)$ , рассчитанные по формуле аппроксимации (9) могут являться оценками сверху теоретических значений  $\sigma_W^2(j)$ .

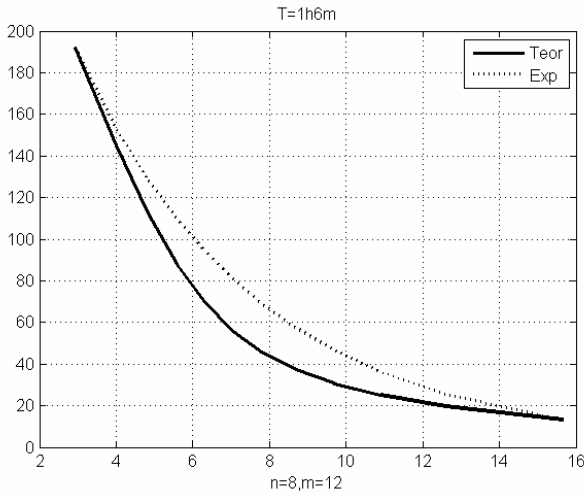


Рис. 4. Значения  $\sigma_W^2(j)$ , рассчитанные по точной (Теор) и по аппроксимирующей (Exp) формулам

### 3. Применение УПС в задачах обнаружения

#### 3.1. «КЛАССИЧЕСКАЯ» ЗАДАЧА ПРОВЕРКИ ДВУХ ГИПОТЕЗ

Пусть на входе системы обнаружения принимается сигнал  $S(t) = \eta(t) + \zeta(t)$ , состоящий из аддитивной смеси сигнала  $\eta(t)$  и помехи  $\zeta(t)$ . Математической моделью сигнала и помехи служит случайный процесс, так что задача обнаружения сигнала на фоне помех решается на основе статистической теории оптимальных решений. Чаще всего в расчетах распределение сигнала на входе системы обнаружения предполагается гауссовским с нулевым математическим ожиданием как в случае одной помехи (гипотеза  $H_0: N(0, \sigma_{\text{ш}}^2)$ ), так и в случае смеси сигнала с помехой (гипотеза  $H_1: N(0, \sigma_{\text{ш}}^2 + \sigma_c^2)$ ). Распределения для помехи (шума) и смеси сигнала с шумом различаются только дисперсией (мощностью) наблюдаемых сигналов  $S(t)$ .

Достаточно часто в качестве статистики от наблюдений выбирают энергию сигнала, наблюдаемого на интервале  $[0, T_0]$ , – энергетический критерий обнаружения [1, 5]:

$$(10) \quad X = \sum_{i=1}^n S^2(i\Delta t),$$

где  $T_0 = n\Delta t$ ;  $\Delta t = 1/2\Delta F$ ;  $\Delta t$  – интервал дискретизации по времени;  $\Delta F$  – полоса пропускания входного фильтра системы обнаружения.

Закон распределения энергии принимаемого сигнала, заданный плотностью вероятности  $f(x)$  случайной величины  $X$ , позволяет математически описать вероятностные характеристики обнаружения: вероятность правильного обнаружения и вероятность ложной тревоги:

$$(11) \quad P_{\text{пт}} = \int_h^{\infty} f_{\text{ш}}(x)dx, \quad P_{\text{обн}} = \int_h^{\infty} f_{\text{с+ш}}(x)dx,$$

где  $h$  – порог обнаружения;  $f_{\text{ш}}(x)$  – плотность вероятности случайной величины  $X$  при гипотезе  $H_0$ ;  $f_{\text{с+ш}}(x)$  – плотность вероятности  $X$  при гипотезе  $H_1$ .

Плотность распределения статистики  $X$  для гауссовского процесса  $S(t)$  описывается плотностью распределения (1). По-

сколькx распределение для гипотезы  $H_0$  отличается от распределения для гипотезы  $H_1$  только величиной дисперсии, то дисперсия  $\sigma^2$  статистики  $X$  с плотностью распределения (1) равна:

– в случае гипотезы  $H_0$ :

$$(12) \sigma_0^2 = 2n\sigma_{\text{ш}}^4,$$

– в случае гипотезы  $H_1$ :

$$\sigma_1^2 = 2n(\sigma_{\text{ш}}^2 + \sigma_c^2)^2.$$

Здесь  $n$  – число степеней свободы, которое в случае цифровой обработки равно числу слагаемых в сумме (10), а в случае аналоговой обработки определяется как  $n = 2T_0\Delta F$ .

Определим отношение сигнал/помеха в полосе  $\Delta F$  как

$$(13) \rho = \frac{\sigma_c^2}{\sigma_{\text{ш}}^2}.$$

Тогда дисперсия в случае гипотезы  $H_1$  запишется в виде

$$\sigma_1^2 = 2n[\sigma_{\text{ш}}^2(1+\rho)]^2 = 2n\sigma_{\text{ш}}^4(1+\rho)^2.$$

Таким образом, для гипотезы  $H_0$  вероятность ложной тревоги  $P_{\text{лт}}$  является функцией от  $\sigma_{\text{ш}}^2$  и  $n$ , а для гипотезы  $H_1$  вероятность правильного обнаружения  $P_{\text{обн}}$  является функцией от  $\sigma_{\text{ш}}^2$ ,  $n$  и  $\rho$ .

Определим вероятность обнаружения как  $P_{\text{обн}} = \beta$ , тогда  $P_{\text{необн}} = 1 - \beta$ , и в соответствии с (2) и (4)

$$(14) P_{\text{необн}} = 1 - \beta = \int_0^h \frac{1}{\sigma_1^2} k_n \left( \frac{x}{\sigma_1} \right) dx = \int_0^{h/\sigma_1} k_n(u) du = K_n(h/\sigma_1^2).$$

Здесь порог  $h$  определяется из первого уравнения (11) путем задания вероятности ложной тревоги  $P_{\text{лт}} = \alpha$ . Тогда аналогично формуле (14)

$$1 - \alpha = \int_0^{h/\sigma_0} k_n(u) du = K_n(h/\sigma_0^2).$$

В работе [2] приведены соотношения, связывающие между собой величины  $P_{\text{лт}}$ ,  $P_{\text{обн}}$  и  $\rho$ . Если

$$P_{\text{лт}} = \alpha = 1 - \int_0^{h/\sigma_0^2} k_n(u) du = K_n(h/\sigma_0^2) = 1 - K_n(h_0),$$

то вероятность обнаружения согласно (14) определяется выражением:

$$P_{\text{обн}} = \beta = 1 - K_n\left(\frac{h}{\sigma_1}\right) = 1 - K_n(h_1) = 1 - K_n\left(\frac{h}{(1+\rho)}\right).$$

Часто в литературе заменяют  $\chi^2$ -распределение нормальным распределением, так как уже при  $n \geq 30$   $\chi^2$ -распределение с большой точностью аппроксимируется гауссовским распределением со следующими параметрами:

– гипотеза  $H_0$ :

$$(15) \mu_0 = n\sigma_{\text{ш}}^2, \sigma_0^2 = 2n\sigma_{\text{ш}}^4,$$

– гипотеза  $H_1$ :

$$\mu_1 = n(\sigma_{\text{ш}}^2 + \sigma_c^2), \sigma_1^2 = 2n(\sigma_{\text{ш}}^2 + \sigma_c^2)^2.$$

При использовании гауссовского приближения для гипотезы  $H_0$  при заданной вероятности ложной тревоги  $P_{\text{лт}} = \alpha$  квантиль  $\tilde{h}_0$  уровня  $(1 - \alpha)$  определяется из уравнения

$$(1 - \alpha) = \Phi\left(\frac{h - m_0}{\sigma_0}\right) = \Phi(\tilde{h}_0),$$

а для гипотезы  $H_1$  при заданной вероятности обнаружения  $P_{\text{обн}} = \beta$  квантиль  $\tilde{h}_1$  уровня  $(1 - \beta)$  – из уравнения

$$(1 - \beta) = \Phi\left(\frac{h - m_1}{\sigma_1}\right) = \Phi(\tilde{h}_1),$$

где  $\Phi(\cdot)$  – функция нормального распределения.

В одиночном акте наблюдения (испытании) в нашем распоряжении имеется выборка из  $m$  результатов наблюдений – выборка (вектор)  $X$ . Выписав отношение правдоподобия для гипотез  $H_1$  и  $H_0$ , получаем достаточную статистику

$$\Lambda(X) = \sum_{i=1}^m X_i.$$

Приведенный выше «классический» алгоритм базируется на существовании равномерно наиболее мощного критерия для

проверки гипотезы  $H_0$  против альтернативы  $H_1$ . При этом используется вся информация, содержащаяся в оцениваемой выборке  $X$ .

### 3.2. ИСПОЛЬЗОВАНИЕ УПС В ЗАДАЧЕ ПРОВЕРКИ ДВУХ ГИПОТЕЗ

Вернемся опять к задаче проверки двух гипотез. Рассмотрим, как изменяется  $P_{лт}$ ,  $P_{обн}$  и  $\rho$  в зависимости от порога отсеечения  $k$  УПС  $W$ . Математическое ожидание  $\mu_W(k)$  и дисперсия  $\sigma^2_W(k)$  определяются формулами (6) и (7) и так же как для случайной величины  $X$  зависят только от дисперсии  $\sigma^2$ . Таким образом, распределения случайной величины  $W$  для шума ( $H_0$ ) и смеси сигнала с шумом ( $H_1$ ) различаются только дисперсией (мощностью)  $\sigma^2$  наблюдаемого сигнала  $S(t)$ .

Обычно для исследования зависимостей  $P_{лт}$ ,  $P_{обн}$  и  $\rho$  используют рабочие характеристики приемника (систему номограмм) [4], что требует много достаточно трудоемких расчетов. В настоящей работе предлагается более простой способ сравнения статистик для гипотез  $H_0$  и  $H_1$ , если статистики могут быть описаны нормальными распределениями:  $N_0(\mu_0, \sigma_0^2)$  и  $N_1(\mu_1, \sigma_1^2)$ .

Обозначим – порог  $h$ , тогда после замены переменных в плотности нормального распределения получаем

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^h e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(h-\mu)/\sigma} e^{-\frac{1}{2}t^2} dt = \Phi\left(\frac{h-\mu}{\sigma}\right),$$

где  $t = (x - \mu)/\sigma$ , а  $\Phi(\cdot)$  – функция нормального распределения. При условии равенства ошибок первого (ложная тревога) и второго (пропуск цели) рода получаем уравнение

$$P_{ош} = P_{пр} = P_{лт} = \Phi\left(\frac{h-\mu_1}{\sigma_1}\right) = 1 - \Phi\left(\frac{h-\mu_0}{\sigma_0}\right) = \Phi\left(-\frac{h-\mu_0}{\sigma_0}\right).$$

Решая это уравнение, определяем порог

$$h = \frac{\mu_0\sigma_1 + \mu_1\sigma_0}{\sigma_1 + \sigma_0},$$

и, подставляя  $h$  в выражение для  $P_{пр}$ , получаем

$$P_{ош} = \Phi\left(\frac{h-\mu_1}{\sigma_1}\right) = \Phi\left(\frac{\mu_0-\mu_1}{\sigma_0+\sigma_1}\right) = 1 - \Phi\left(\frac{\mu_1-\mu_0}{\sigma_0+\sigma_1}\right) = 1 - \Phi(\gamma),$$

где  $\gamma$  – коэффициент разделимости. Очевидно, что чем больше  $\gamma$ , тем меньше  $P_{\text{ош}}$ . Таким образом, определяем коэффициент разделимости как

$$(16) \gamma = \frac{\mu_1 - \mu_0}{\sigma_0 + \sigma_1}.$$

На рис. 5 приведены рассчитанные по формулам (6) и (7) значения  $\mu_W(k)$  и  $\sigma_W(k)$ , и соответствующие значения  $\gamma$ , рассчитанные по формуле (16) для гипотез  $H_0$  и  $H_1$ , для следующих значений параметров:  $n = 8, m = 12, \rho = 0,1$ . При  $k = 1$   $\mu_W(1) = \mu_Z$ ,  $\sigma_W(1) = \sigma_Z$  и, следовательно,  $\gamma_W(1) = \gamma_Z$ . Существует также оптимальное значение порога отсечения  $k - k_{\text{опт}}$ , обеспечивающее максимальную разделимость (параметр  $\gamma$ ) статистик  $W$  для гипотез  $H_0$  и  $H_1$ , что объясняется тем, что числитель в формуле (16) – монотонно убывающая выпуклая функция, а знаменатель – монотонно убывающая вогнутая функция.

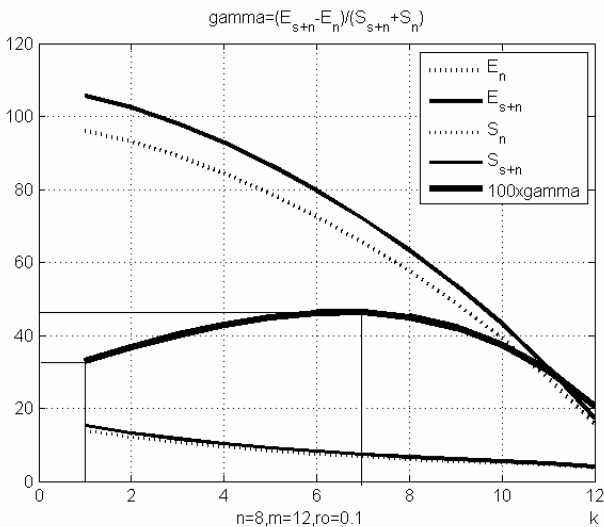


Рис. 5. Значения  $\mu_W(k)$  и  $\sigma_W(k)$ , и соответствующие значения  $\gamma(k)$ , для гипотез  $H_0$  ( $E_n$  и  $S_n$ ) и  $H_1$  ( $E_{s+n}$  и  $S_{s+n}$ )

Таким образом, процесс принятия решения при использовании УПС  $W$  является двухшаговым:

– на первом шаге по случайной выборке  $X_i$ ,  $1 \leq i \leq m$ , строится порядковая статистика  $X_{(i)}$ , по которой для заданного порога отсечения  $k$  (первый порог) формируется УПС  $W$ ;

– на втором шаге процесс принятия решения тот же, что и в «классическом» алгоритме: а именно, по заданной  $P_{лт}$  выставляется порог обнаружения (второй порог), в случае превышения которого принимается решение о справедливости гипотезы  $H_1$ .

Как уже указывалось выше, известными параметрами являются только статистические свойства помехи, а именно, математическое ожидание  $\mu_0$  и дисперсия  $\sigma_0^2$  (см. (12) или (15)). Поэтому порог  $h$  определяется согласно формуле (11) из следующего уравнения:

$$(17) \alpha = \int_0^h f_{ш}(x) dx = \frac{1}{2^{n/2} \sigma_0^n \Gamma(n/2)} \int_0^h x^{n/2-1} e^{-x^2/2\sigma_0^2} dx,$$

где  $\alpha$  – заданный квантиль, который, как следует из рис. 5, должен быть близок к медиане шума.

Рассмотрим вопрос о выборе порога отсечения  $k$  (первый порог). Для того чтобы сформировать УПС  $W$ , необходимо обеспечить условие  $k = const$  для нескольких реализаций. Чтобы выполнить это условие необходимо иметь несколько выборок  $\{X_1, \dots, X_i, \dots, X_m\}_j$ , где  $1 \leq j \leq r$ , из которых строится матрица  $X_{ij}$  размерностью  $m$  строк на  $r$  столбцов. В каждом столбце матрицы  $X_{ij}$  строится порядковая статистика  $X_{(i)j}$ , затем в каждой строке полученной матрицы  $X_{(i)j}$  определяются оценки математических ожиданий (вектор  $\hat{m}$ )

$$\hat{m}_i = \frac{1}{r} \sum_{j=1}^r X_{(i)j}, \text{ где } 1 \leq i \leq m.$$

Порог отсечения  $k$  (первый порог) определяется из условия:

$$k = \arg \min_i |h - \hat{m}_i|, \text{ где } 1 \leq i \leq m.$$

И уже по этому порогу отсечения  $k$  по матрице порядковых статистик  $X_{(i)j}$  вычисляются УПС  $W_j$ , где  $1 \leq j \leq r$ .

Таким образом, для реализации предлагаемого алгоритма обнаружения необходимо предварительное накопление выборок  $\{X_1, \dots, X_i, \dots, X_m\}_j$ , где  $1 \leq j \leq r$  (в отличие от «классического»

алгоритма), что приводит к задержке в принятии решения на время  $T = jmT_0$ , где  $1 \leq j \leq r$ . Такая задержка во многих задачах не является существенной.

Еще раз подчеркиваем, что если в «классическом» алгоритме для принятия решения используется только вектор  $\mathbf{X}$ , то в предлагаемом алгоритме – матрица  $X_{(ij)}$ , в которой текущий вектор  $\mathbf{X}$  является одним из столбцов.

На рис. 6 приведены результаты математического моделирования с.к.о. УПС  $W$  в зависимости от «глубины» памяти  $r$ , содержащей матрицу порядковых статистик  $X_{(ij)}$  для следующих значений параметров:  $n = 8$ ,  $m = 100$ . При  $r = 1$   $\sigma_W > \sigma_Z$ , затем  $\sigma_W$  спадает и уже при  $r > 4$   $\sigma_W < \sigma_Z$  и быстро стремится к своему теоретическому значению, т.е. к значению, когда «глубина» памяти велика (стремится к бесконечности).

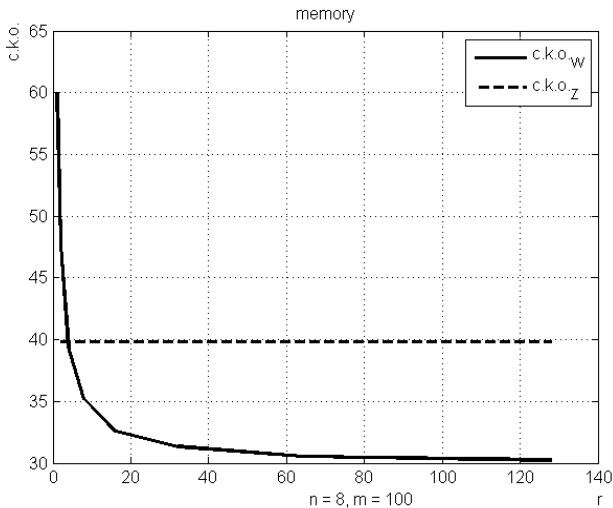


Рис. 6. Зависимость  $\sigma_W$  от «глубины» памяти  $r$

На рис. 7 приведены результаты математического моделирования систем обнаружения, реализующих «однопороговый» и «двухпороговый» алгоритмы, а именно: теоретические значения плотностей  $f_Z(\sigma_n)$ ,  $f_Z(\sigma_{n+s})$  и  $f_W(\sigma_n)$ ,  $f_W(\sigma_{n+s})$  и их гистограммы,



построенные по математической модели. Параметры модели имеют следующие значения:  $n = 8$ ,  $m = 32$ ,  $\rho = 0,1$ , первый порог равен медиане порядковой статистики  $X_{(i)}$ ,  $k = 17$ . (Размеры массивов для построения гистограмм – 5000). Коэффициенты разделимости  $\gamma$ , полученные при моделировании:  $\gamma_z = 0,27$  и  $\gamma_w = 0,48$ , что близко к теоретическим значениям.

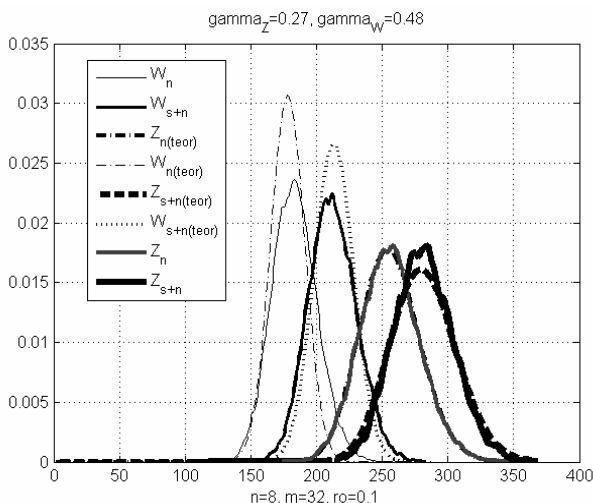


Рис. 7. Теоретические значения (на рис. –  $Z_{n(teor)}$ ,  $Z_{n+s(teor)}$  и  $W_{n(teor)}$ ,  $W_{n+s(teor)}$ ) и гистограммы (на рис. –  $Z_n$ ,  $Z_{s+n}$  и  $W_n$ ,  $W_{s+n}$ ) плотностей  $f_z(\sigma_n)$ ,  $f_z(\sigma_{n+s})$  и  $f_w(\sigma_n)$ ,  $f_w(\sigma_{n+s})$

Далее приведены результаты моделирования системы обнаружения, реализующей «классический» (однопороговый) алгоритм и системы обнаружения, реализующей алгоритм на основе УПС (двухпороговый). На рис. 8 приведены зависимости вероятности обнаружения  $P_{обн}$  от отношения сигнал/помеха  $\rho$  (13) для «однопорогового» и «двухпорогового» алгоритмов для разных размеров «глубины» памяти  $r$  (*memory*). Параметры модели имеют следующие значения:  $n = 10$ ,  $m = 32$ ,  $P_{лт} = 0,05$ , первый порог равен медиане статистики  $X_i$  (в формуле(17)  $\alpha = 0,5$ ). Размеры массивов для набора статистик – 10 000. Из приведенных графиков видно, что, например, при  $\rho = 0,075$  для

«классического» алгоритма  $P_{\text{обн}} = 0,2$ , а для «двухпорогового» алгоритма  $P_{\text{обн}} = 0,58$  при  $r = 16$ , т.е. «двухпороговый» алгоритм обеспечивает существенный выигрыш по сравнению с «классическим» (однопороговым) алгоритмом.

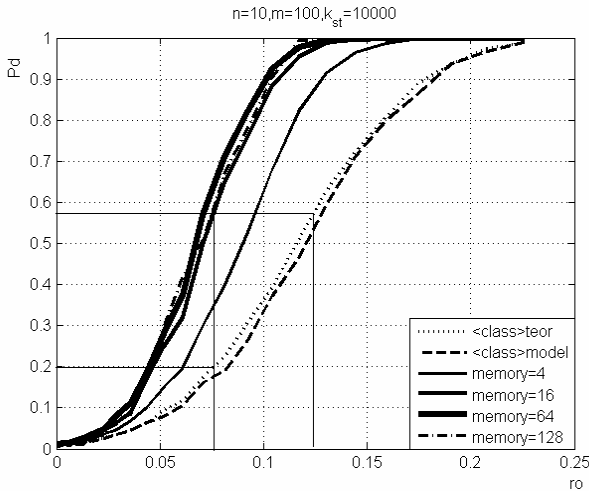


Рис. 8. Сравнение  $P_{\text{обн}}$  для «двухпорогового» и «однопорогового» алгоритмов

Рассмотрим, как влияет на вероятность обнаружения задержка в принятии решения на  $j$  шагов. Пусть  $r$  – «глубина» памяти, содержащей выборки  $\{X_1, \dots, X_i, \dots, X_m\}_j$ , где  $1 \leq j \leq r$ , соответствующие гипотезе  $H_0$ .

Пусть в момент времени  $t_0$  появляется сигнал (гипотеза  $H_1$ ), который постепенно заполняет матрицу памяти  $\{X_1, \dots, X_i, \dots, X_m\}_j$ , где  $1 \leq j \leq r$ . Для  $j$  независимых испытаний вероятность того, что произойдет хотя бы одно обнаружение (суммарная вероятность обнаружения) определяется по формуле

$$(18) P_{\Sigma \text{обн}} = 1 - \prod_{i=1}^j (1 - P_{\text{обн}i}),$$

где  $P_{\text{обн}i}$  – вероятность обнаружения в единичном ( $i$ -м) испытании. Для «однопорогового» алгоритма обнаружения вероятность

обнаружения в единичном испытании постоянна и равна  $\beta$ , и суммарная вероятность обнаружения определяется по формуле (19)  $P_{\Sigma\text{обн}} = 1 - (1 - \beta)^j$ .

Для «двухпорогового» алгоритма по мере заполнения матрицы памяти вероятность обнаружения в единичном испытании  $P_{\text{обн } i}$  будет увеличиваться, стремясь к своему теоретическому значению, а суммарная вероятность обнаружения определяется по формуле (18).

На рис. 9 приведены результаты моделирования.

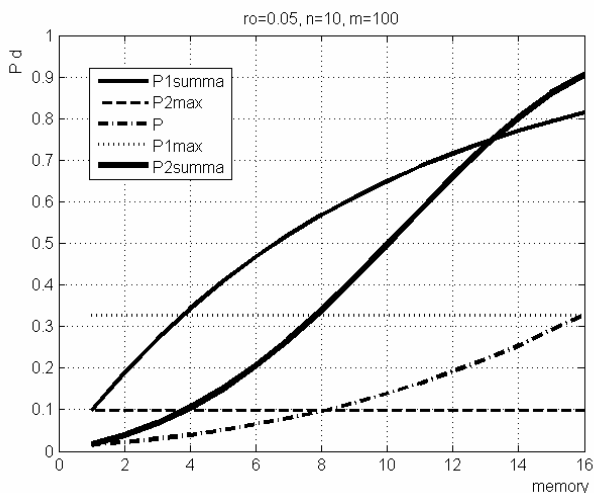


Рис. 9. Сравнение  $P_{\Sigma\text{обн}}$  для «двухпорогового» ( $P2summa$ ) и «однопорогового» ( $P1summa$ ) алгоритмов

Из приведенных графиков видно, что для «однопорогового» алгоритма обнаружения суммарная вероятность обнаружения, как следует из формулы (19), экспоненциально увеличивается, а для «двухпорогового» алгоритма суммарная вероятность обнаружения на первых шагах проигрывает «однопороговому» алгоритму, но затем из-за быстрого роста  $P_{\text{обн } i}$  (на рис. –  $P$ ) превышает его, обеспечивая выигрыш в суммарной вероятности обнаружения  $P_{\Sigma\text{обн}}$  для «двухпорогового» алгоритма.

#### 4. Заключение

Получены аналитические выражения для математических ожиданий и дисперсий для усеченных порядковых статистик. Показано, что фиксированный порог отсеечения в УПС обеспечивает меньшую дисперсию, чем дисперсия исходной статистики.

Разработан алгоритм обнаружения сигналов на фоне шума, основанный на свойствах усеченных порядковых статистик, который позволяет обеспечить большую вероятность обнаружения  $P_{\text{обн}}$  при заданной вероятности ложной тревоги  $P_{\text{лт}}$  по сравнению с «классическим» алгоритмом проверки двух гипотез. Выигрыш достигается за счет введения дополнительного порога, отсекающего малые значения обрабатываемого сигнала, и использования информации, содержащейся в предыдущих реализациях сигнала, которая в «классическом» алгоритме не используется.

Приводятся результаты математического моделирования.

Предварительные результаты настоящей работы докладывались на конференциях [7] и [8].

Выражаю благодарность с.н.с. Л.П. Сысоеву за плодотворные дискуссии и помощь в получении некоторых математических результатов.

#### Литература

1. БУРДИК В.С. *Анализ гидроакустических систем*. – Л.: Судостроение, 1988. – 392 с.
2. ДОБРОВИДОВ А.В., КУЛИДА Е.Л., РУДЬКО И.М. *Выбор траектории движения объекта в конфликтной среде* // Системы управления – 2011. – №3. – С. 64–75.
3. ДЭЙВИД Г. *Порядковые статистики*. – М.: Наука, 1979. – 336 с.
4. ЛЕВИН Б.Р. *Теоретические основы статистической радиотехники*. – Т. 1. – М.: Советское радио, 1969. – 752 с.
5. ЛЕВИН Б.Р. *Теоретические основы статистической радиотехники*. – Т. 2. – М.: Советское радио, 1968. – 504 с.

6. НАЙТ У.С., ПРИДЭМ Р.Г., КЭЙ С.М. *Цифровая обработка сигналов в гидролокационных системах* // ТИИЭР – 1981. – Т. 69, №11. – С. 84–154.
7. РУДЬКО И.М. *Исследование методами математического моделирования двухпороговой системы обнаружения сигналов* // Материалы VIII Всероссийской научно-практической конференции с международным участием «Информационные технологии и математическое моделирование». Ч. 1. – Томск: Изд.-во Томского университета, 2009. – С. 82–86.
8. РУДЬКО И.М. *Применение порядковых статистик в задачах обнаружения* // IX Международная конференция «Идентификация систем и задачи управления» SICPRO'12. – Москва, 2012. – С. 1101–1116.

## APPLYING ORDER STATISTIC TO DETECTION PROBLEMS

**Igor Rudko**, Institute of Control Sciences of RAS, Moscow, Cand.Sc. (igor-rudko@mail.ru).

*Abstract: The concept of truncated order statistic is introduced. Analytical expressions are derived for the mean value and the variance of truncated order statistic for a threshold determined by the number of the variation series member. The algorithm of noisy signal detection is developed for detection systems, which employ the algorithm of statistical hypotheses testing and energy criterion of detection. The algorithm uses statistical characteristics of the order statistic.*

Keywords: order statistics, statistical hypotheses testing, detection systems, probabilities of detection and false emergence, mathematical modeling.

*Статья представлена к публикации членом редакционной коллегии О. И. Орловым*