

УДК 621.396

И.М. Рудько

СТАТИСТИЧЕСКИЕ СВОЙСТВА СУММЫ ЧЛЕНОВ УСЕЧЕННОГО ВАРИАЦИОННОГО РЯДА

Введено понятие усеченной порядковой статистики. Выведены аналитические выражения для математических ожиданий и дисперсий усеченных порядковых статистик для двух вариантов задания порога отсеечения: порога, определяемого квантилем функции распределения, и порога, определяемого номером члена вариационного ряда.

Ключевые слова: *порядковая статистика, моменты случайной величины, системы обнаружения, математическое моделирование.*

Рассматривается задача определения математического ожидания и дисперсии суммы членов вариационного ряда (порядковой статистики) в случае, когда отбрасывается фиксированное число меньших членов вариационного ряда (усеченная порядковая статистика (УПС)) для двух вариантов задания порога отсеечения: порога, фиксированного по заданному квантилю функции распределения, и порога, фиксированного по заданному номеру члена вариационного ряда, при условии, что статистические свойства членов вариационного ряда могут быть описаны центральным χ^2 -распределением с n степенями свободы.

Задачи, которые могут быть описаны такой моделью, возникают при обработке результатов спектрального анализа, вибрационных испытаний, обработке гидроакустической информации и в других задачах, в которых используется энергетический критерий обнаружения [1]. Ранее в работе [2] методами математического моделирования была исследована модель эвристической двухпороговой системы обнаружения сигнала, в которой применение усеченной порядковой статистики в случае, когда известны дисперсии сигнала и шума, позволяет при фиксированной вероятности ложных тревог получить существенный выигрыш в вероятности обнаружения по сравнению с традиционной системой обнаружения. В настоящей работе приводится теоретическое обоснование алгоритма работы этой эвристической двухпороговой системы обнаружения сигнала.

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 функцией плотности распределения

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

где $k_n(\cdot)$ – плотность центрального χ^2 -распределения с n степенями свободы, σ^2 – дисперсия.

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

- 1) $Z = \sum_{i=1}^m X_i$ имеет центральное χ^2 -распределение с nm степенями свободы,

математическим ожиданием

$$\mu_Z = nm\sigma^2, \quad (2)$$

дисперсией

$$\sigma_Z^2 = 2nm\sigma^4,$$

и в силу центральной предельной теоремы при достаточно больших значениях m ее функция плотности распределения нормализуется, т.е. $Z \sim N(\mu_Z, \sigma_Z^2)$.

$$2) \quad V = \sum_{i=l}^m X_{(i)}, \quad (4)$$

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 $X_{(i)}$ зависят из-за неравенств между ними, l выбирается из условия $X_{(l-1)} < h \leq X_{(l)}$, а $h = \text{const}$ – фиксированный порог. Таким образом, l – случайная величина, зависящая от порядковых статистик выборки $X_{(i)}$.

Рассмотрим случайную величину Y_i , которая связана со случайной величиной X_i нелинейным преобразованием

$$Y_i = \begin{cases} X_i, & \text{если } X_i \geq h, \\ 0, & \text{если } X_i < h. \end{cases}$$

Функция плотности $g(x)$ случайной величины Y_i определяется следующим образом (рис. 1):

$$g(x) = \delta(0)\alpha + f(x)I(x),$$

где $\alpha = \int_0^h f(x)dx$, индикатор $I(x) = \begin{cases} 0 & \text{при } 0 \leq x < h, \\ 1 & \text{при } x \geq h, \end{cases}$ h – порог, δ – дельта-

функция, введенная для сохранения условий нормировки $\int_0^\infty g(x)dx = 1$.

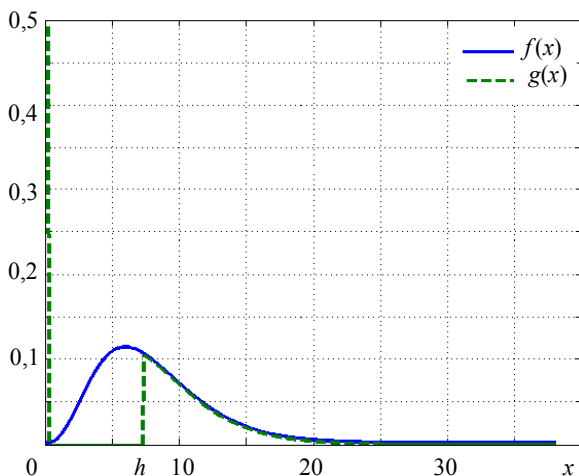


Рис. 1. Функции плотности случайных величин $f(x)$ и $g(x)$

Математическое ожидание случайной величины Y определяется следующим образом:

$$\begin{aligned}\mu_Y &= \int_0^{\infty} xg(x)dx = \int_h^{\infty} xf(x)dx = \frac{1}{2^{n/2}\sigma^n\Gamma(n/2)} \int_h^{\infty} x^{n/2} e^{-x/2\sigma^2} dx = \\ &= \frac{2\sigma^2\left(\frac{(n+2)}{2}-1\right)}{2^{(n+2)/2}\sigma^{(n+2)}\Gamma\left(\frac{(n+2)}{2}\right)} \int_h^{\infty} x^{(n+2)/2-1} e^{-x/2\sigma^2} dx = n\sigma^2 \left[1 - K_{n+2}\left(\frac{h}{\sigma^2}\right)\right];\end{aligned}\quad (5)$$

$$K_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2\sigma^2} \int_0^x t^{n/2-1} e^{-t/2\sigma^2} dt. \quad (6)$$

В (6) $K_n(x)$ – функция распределения центрального χ^2 -распределения с n степенями свободы.

Аналогично, 2-й начальный момент Y определяется по формуле

$$M_2 = \int_h^{\infty} x^2 f(x)dx = 2n\left(\frac{n}{2}+1\right)\sigma^4 \left[1 - K_{n+4}\left(\frac{h}{\sigma^2}\right)\right],$$

а дисперсия равна $\sigma_Y^2 = M_2 - \mu_Y^2$.

Таким образом, случайную величину V можно представить в следующем виде:

$$V = \sum_{i=1}^m X_{(i)} = \sum_{i=1}^m Y_i,$$

при условии, что порог $h = \text{const}$ для всей выборки X_i ($1 \leq i \leq m$) (например, h – медиана плотности $f(x)$).

Для случайной величины V математическое ожидание

$$\mu_V = m\mu_Y; \quad (7)$$

дисперсия

$$\sigma_V^2 = m\sigma_Y^2, \quad (8)$$

и в силу центральной предельной теоремы при достаточно больших значениях m ее функция плотности распределения нормализуется ($V \sim N(\mu_V, \sigma_V^2)$).

В случае, если число степеней свободы n – четное, для определения моментов Y можно использовать представление интегралов для вычисления μ_Y и M_2 в виде конечного ряда ([3], формулы (3.381(3)) и (8.352(2))):

- математическое ожидание Y

$$\mu_Y = n\sigma^2 e^{-(h/2\sigma^2)} \left[1 + \sum_{k=1}^{n/2} \frac{(h/2\sigma^2)^k}{k!}\right],$$

- 2-й начальный момент Y

$$M_2 = 2n\left(\frac{n}{2}+1\right)\sigma^4 e^{-(h/2\sigma^2)} \left[1 + \sum_{k=1}^{n/2+1} \frac{(h/2\sigma^2)^k}{k!}\right],$$

- дисперсия Y

$$\sigma_Y^2 = M_2 - \mu_Y^2.$$

3) $W = \sum_{i=k}^m X_{(i)}$, при условии, что порог $k = \text{const}$. В дальнейшем будем называть статистику W – усеченная порядковая статистика (УПС), а параметр k – порог отсеечения.

В работе [4] приведены выражения для вычисления моментов порядковых статистик. Для случайной величины, описываемой формулами (1) и (6), т.е. плотностью распределения $f(x)$ и функцией распределения $K_n(x)$, формулы для определения μ_j , σ_j^2 и σ_{jk} приобретают следующий вид:

$$\begin{aligned}\mu_j &= \frac{m!}{(j-1)!(m-j)!} \int_0^\infty K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[1 - K_n\left(\frac{x}{\sigma^2}\right)\right]^{m-j} x f(x) dx, \\ \sigma_j^2 &= \frac{m!}{(j-1)!(m-j)!} \int_0^\infty K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[1 - K_n\left(\frac{x}{\sigma^2}\right)\right]^{m-j} (x - \mu_j)^2 f(x) dx, \quad (9) \\ \sigma_{jk} &= E[X_{(j)}X_{(k)}] = \frac{m!}{(m-k)!(k-j-1)!(j-1)!} \int_0^\infty \int_0^y C(x, y) x y f(x) f(y) dx dy, \\ \text{где } C(x, y) &= K_n\left(\frac{x}{\sigma^2}\right)^{j-1} \left[K_n\left(\frac{y}{\sigma^2}\right) - K_n\left(\frac{x}{\sigma^2}\right) \right]^{k-j-1} \left[1 - K_n\left(\frac{y}{\sigma^2}\right)\right]^{m-k}.\end{aligned}$$

Для случайной величины W математическое ожидание

$$\mu_W(l) = \sum_{j=l}^m \mu_j \quad \text{для } 1 \leq l \leq m, \quad (10)$$

а дисперсия с учетом зависимости случайных величин $X_{(i)}$ [5]

$$\sigma_W^2(l) = \sum_{k=l}^m \sigma_k^2 + 2 \sum_{l \leq j < k \leq m} \sigma_{jk} \quad \text{для } 1 \leq l \leq m, \quad (11)$$

и в силу центральной предельной теоремы при достаточно больших значениях m ее функция плотности распределения также нормализуется – $W \sim N(\mu_W, \sigma_W^2)$.

2. Моделирование

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

На рис. 2 приведены рассчитанные по формулам (7), (8), (10) и (11) зависимости математических ожиданий $\mu_V(h)$, $\mu_W(l)$ и с.к.о. $\sigma_V(h)$, $\sigma_W(l)$ от порога при следующих значениях параметров: $n = 8$, $m = 12$, $\sigma = 1$.

Для случайной величины V порог $h \geq 0$, для W – порог l дискретный, $1 \leq l \leq m$, и для удобства сравнения моментов случайных величин V и W на оси абсцисс для

УПС W отложены дискретные значения порога

$$h_1 = \frac{\mu_1}{2}, \quad h_l = \mu_{l-1} + \left(\frac{\mu_l - \mu_{l-1}}{2} \right),$$

где μ_l – рассчитанные по формуле (9) значения математических ожиданий $2 \leq l \leq m$. Из графиков видно, что $\mu_V \approx \mu_W$, $\sigma_V(h)$ имеет максимум, приблизительно совпадающий с максимумом плотности распределения $f(x)$ величины X , а $\sigma_W(l)$ монотонно убывающая функция, причем $\mu_V(0) = \mu_W(1) = \mu_Z = nm$ и $\sigma_V(0) = \sigma_W(1) = \sigma_Z = \sqrt{2nm}$.

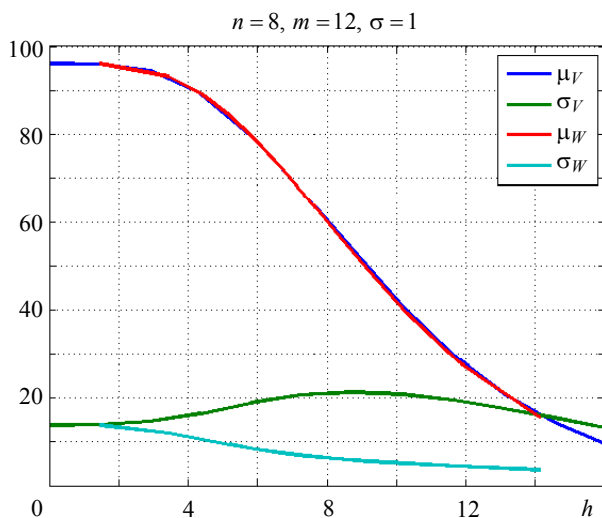


Рис. 2. Зависимости математических ожиданий $\mu_V(h)$, $\mu_W(l)$ и с.к.о. $\sigma_V(h)$, $\sigma_W(l)$ от порога для случайных величин V и W

Сравним между собой статистические свойства случайных величин V и W .
Случайная величина

$$W = \sum_{i=k}^m X_{(i)}, \quad (12)$$

где $X_{(i)}$ ($1 \leq i \leq m$), а $k = \text{const}$.

Случайная величина

$$V = \sum_{i=l}^m X_{(i)}, \quad (13)$$

где l такая, что $X_{(l-1)} \leq h < X_{(l)}$.

Еще раз отмечаем, что l – случайная величина, зависящая от порядковых статистик выборки, и единственным отличием УПС V и W является то, что число членов в сумме (12) $k = \text{const}$, а в сумме (13) l определяется из условия $X_{(l-1)} < h \leq X_{(l)}$, где $h = \text{const}$.

На рис. 3 приведены рассчитанные по формулам (2), (3), (7), (8), (10) и (11) плотности распределения случайных величин Z , V и W (рассчитанные при следующих значениях параметров: $n = 8$, $m = 100$, порог отсеечения $k = 51$; медиана порядковой статистики $X_{(i)}$ ($1 \leq i \leq m$), $h = 7,344$ – медиана χ^2 -распределения с 8 степенями свободы) и гистограммы этих плотностей, полученные на моделях. (Размеры массивов для построения гистограмм – 5000). Наблюдается достаточно близкое совпадение между теоретическими и модельными результатами.

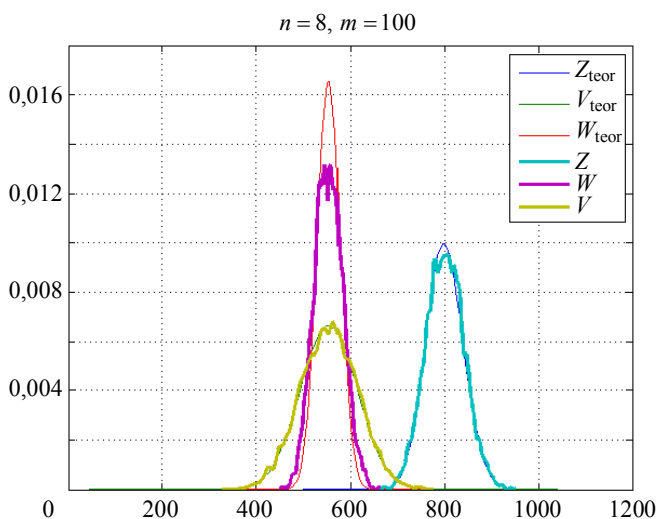


Рис. 3. Плотности распределения и гистограммы случайных величин Z , V и W

В системах обнаружения, использующих энергетический критерий обнаружения [1], применение статистик W может позволить увеличить вероятность обнаружения слабых сигналов. Ниже в качестве иллюстрации приведены заимствованные из этой работы результаты математического моделирования двух систем обнаружения:

- однопороговой системы обнаружения сигнала (традиционной), реализующей алгоритм обнаружения, основанный на статистике Z (алгоритм-1),
- двухпороговой системы обнаружения сигнала, реализующей алгоритм обнаружения, основанный на статистике W (алгоритм-2).

На рис. 4 приведены графики оценок вероятности обнаружения $P_{\text{обн}} = f(\sigma_{\text{с+ш}})$ как функции от с.к.о. смеси сигнала и шума $\sigma_{\text{с+ш}}$ (при условии, что с.к.о. шума $\sigma_{\text{ш}} = 1$) для

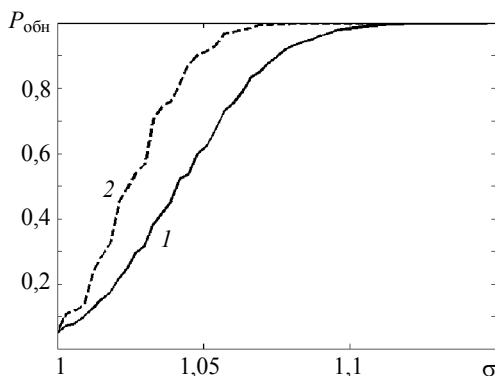


Рис. 4. Вероятности обнаружения $P_{\text{обн}} = f(\sigma_{\text{с+ш}})$ для алгоритма-1 (кр. 1) и алгоритма-2 (кр. 2)

статистики $Z - P_{\text{обн}}^Z$ (кр. 1) (алгоритм-1) и для статистики $W - P_{\text{обн}}^W$ (кр. 2) (алгоритм-2), рассчитанные при следующих параметрах модели: $n = 8$, $m = 100$, порог отсечения $k = 51$ и $P_{\text{пт}} = 0,05$ по выборке размером – 10000.

В работе [6] результаты работы [2] обобщены на случай неизвестной дисперсии сигнала и дано обоснование выбора порога отсечения для этого случая.

Заключение

В данной работе получены аналитические выражения математических ожиданий и дисперсий для усеченных порядковых статистик. Показано, что фиксированный порог отсечения в УПС обеспечивает меньшую дисперсию, чем порог отсечения, определенный фиксированным квантилем исходной статистики, и меньшую, чем дисперсия исходной статистики. Приводятся результаты моделирования.

ЛИТЕРАТУРА

1. Левин Б.Р. Теоретические основы статистической радиотехники. Т. 2. М.: Сов. радио, 1968. 504 с.
2. Рудько И.М. Исследование методами математического моделирования двухпороговой системы обнаружения сигналов // Информационные технологии и математическое моделирование. Материалы VIII Всероссийской научно-практической конференции с международным участием. Томск: Изд-во Том. ун-та, 2009. Ч. 1. С. 82–86.
3. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. М.: Наука, 1971. 1108 с.
4. Дэйвид Г. Порядковые статистики. М.: Наука, 1979. 336 с.
5. Левин Б.Р. Теоретические основы статистической радиотехники. Т. 1. М.: Сов. радио, 1969. 752 с.
6. Рудько И.М. Применение порядковых статистик в задачах обнаружения // IX Международная конференция «Идентификация систем и задачи управления» SICPRO'12, Москва, 30 января – 2 февраля 2012, С. 1101–1116.

Рудько Игорь Михайлович

Институт проблем управления им. В.А. Трапезникова РАН

E-mail: igor-rudko@mail.ru

Поступила в редакцию 2 мая 2012 г.

Rudko Igor M. (Institute of Control Sciences of RAS, Moscow). Statistical characteristics to sum of terms of truncated variational series.

Keywords: order statistics, moments of random quantity, detection systems, mathematical modeling.

The concept of truncated order statistic is introduced. Analytical expressions for the mean value and the variance of truncated order statistic are derived for the two version of a threshold: for a threshold determined by the distribution function quantile and for a threshold determined by the number of member of the variational series.