DOI 10.52209/1609-1825_2022_1_317 УДК 629.7.05.001

# Непараметрические методы теории решений при исследовании случайных процессов

***ЕСМАГАМБЕТОВ Булат-Батыр Саухымович,*** *д.т.н., профессор, bulatbatyr@mail.ru,*
*Южно-Казахстанский университет им. М.О. Ауэзова, Казахстан, 160012, Шымкент, пр. Тауке хана, 5.*

***Аннотация.*** *Традиционные способы оценки вероятностных характеристик случайных процессов требуют при обработке наличия стационарных свойств и знания априорных распределений данных измерений. На практике такие требования часто не выполняются потому, что значительная часть данных измерений относится к нестационарным случайным процессам. Это относится ко многим радиотехническим системам, например, радиотелеметрическим системам космических летательных аппаратов [1], в которых для обработки данных используются, так называемые, методы сжатия данных. Такие системы отличаются тем, что, как правило, в спектре измеряемых параметров доминирующее положение занимают широкополосные быстроменяющиеся процессы, поэтому внедрение методов квазиобратимого сжатия данных только для медленно меняющихся процессов, во-первых, не решает проблемы совокупного сокращения объема передаваемых данных и разгрузки канала передачи данных, а во-вторых, не позволяет существенно сократить сроки экспериментальных исследований из-за отсутствия возможности оперативного использования всей измеряемой информации. К тому же обрабатываемые на борту данные очень часто представляют единственную реализацию нестационарных случайных процессов. Современная математическая статистика не располагает методами для оценки вероятностных характеристик таких процессов. В условиях априорной неопределенности о функции распределения и ее параметрах для обработки данных измерений можно использовать непараметрические методы теории статистических решений. В статье рассмотрен метод выделения среднего значения нестационарного случайного процесса с помощью непараметрической статистики Кендалла.*

***Ключевые слова:*** *непараметрическая статистика, знаковая функция, порядковая статистика, ранговая статистика, непараметрическая гипотеза, случайный процесс, функция распределения, обобщенные инверсии.*

**Введение**

Известно [2], что непараметрической статистикой называют некоторую функцию случайной величины с неизвестным распределением вероятностей. Сама эта функция обладает известным распределением, свойства которого некоторым образом характеризуют свойства неизвестного распределения исходной случайной величины. Зная распределение непараметрической статистики, можно с его помощью сформулировать и проверить разные гипотезы о свойствах неизвестных распределений (например, их симметрии, нормальности, стационарности и так далее) [3].

**Теория вопроса**

Рассмотрим характер конкретных задач, решаемых с помощью непараметрических методов. Прежде всего эта задача оценивания неизвестных распределений, которая отличается от задачи аппроксимации неизвестного распределения известными функциями, рассматриваемой в обычной статистике. В непараметрической постановке эта задача может быть сформулирована как оценка отличия неизвестного распределения от заданного класса распределений. При необходимости конкретизировать эти различия формулируется задача оценивания параметров распределений. При этом оценивается не сам параметр, а параметр различия между распределениями внутри заданного непараметрического класса. Другая категория непараметрических задач – проверка непараметрических гипотез. В любой задаче проверки непараметрических гипотез, состоящей из двух конкурирующих гипотез, альтернативная всегда непараметрична, а нулевая гипотеза может быть либо простой, либо непараметричной. Различие между гипотезами не связано с конкретным видом функции распределения, поскольку в одной из гипотез присутствует класс неизвестных распределений. Суть процедуры заключается в том, что на основании исходной выборки требуется приложить алгоритм, результатом которого будет решение об истинности одной из гипотез.

Разберем кратко основные непараметрические задачи проверки гипотез.

Задача симметрии. Эта задача проверки симметричности распределения $F(x)$ относительно точки $x_0$. В этом случае формируются две гипотезы:

$$H_0\colon F^{-1}(0{,}5)=x_0 \text{ и } F(x) \text{ – симметрична;}$$

$$H_i\colon F^{-1}(0{,}5)\neq x_0 \text{ или } F(x) \text{ – несимметрична.}$$

Задача сдвига. Во многих случаях возникает задача обнаружения сдвига распределения. Сам факт сдвига в ту или иную сторону считается известным (причем не обязательно только сдвиг, но сдвиг – обязательно). В простейшем случае задача сводит альтернативную гипотезу только к сдвигу, то есть $F(x\,|\,\theta)=F_0(x-\theta)$. Основная гипотеза $H_0$: $\theta=0$. Альтернатива $H_i$: $\theta\neq 0$. В качестве альтернативы могут быть использованы две простые односторонние гипотезы:

$$H_I'\colon\ \theta>0 \quad H_I''\colon\ \theta<0.$$

В некоторых случаях известно, что распределение симметрично относительно медианы. В этом случае задача сдвига может быть сформулирована следующим образом:

$$H_0\colon F^{-1}(0{,}5)=x_0 \text{ и } F(x) \text{ – симметрична относительно } x_0;$$

$$H_i\colon F^{-1}(0{,}5)>x_0 \text{ и } F(x) \text{ – несимметрична, либо}$$

$$H_I'\colon\ F^{-1}(0{,}5)=x_0 \text{ и } F(x) \text{ – несимметрична.}$$

Задача случайности. Эта задача вытекает из необходимости проверить, является ли наблюдаемая выборка действительно чисто случайной, то есть состоящей из одинаково распределенных и независимых случайных величин. Гипотезы в этой задаче имеют вид:

$$H_0\colon\ F(x_1,x_2,\ldots,x_N)=\prod_{i=1}^{N}F(x_i),$$

$$H_i\colon\ F(x_1,x_2,\ldots,x_N)=\prod_{i=1}^{N}F_i(x_i)\neq\prod_{i=1}^{N}F(x_i).$$

Эта задача перекликается с задачей независимости, но в этом случае упор делается на одинаковую распределенность величин.

В задаче независимости проверяется, являются ли компоненты случайного вектора статистически независимым. Причем задача является непараметричной, если распределения считаются неизвестными. Гипотезы формируются в виде:

$$H_0\colon F(x_1,x_2,\ldots,x_k)=F_1(x_1)\,F_2(x_2)\ldots F_k(x_k),$$

$$H_i\colon F(x_1,x_2,\ldots,x_k)\neq F_1(x_1)\,F_2(x_2)\ldots F_k(x_k)$$

хотя бы для некоторых значений $x_i$.

Рассмотрим для примера процедуру формирования решающих правил для проверки гипотезы о симметрии распределения некоторой случайной переменной $y$ (рисунок 1).

Введем в рассмотрение функцию счетчика знаков

$$Z=\sum_{i=1}^{n}u_i(y_i), \tag{1}$$

называемой в литературных источниках знаковой статистикой:

$$u(y)=1 \text{ при } y\geq 0;\ u(y)=0 \text{ при } y<0.$$

Знаковая статистика $Z$ имеет биноминальное распределение

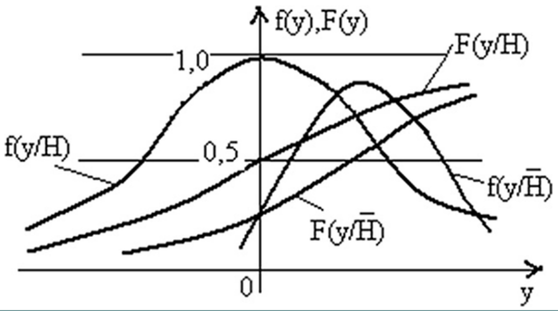


**Рисунок 1 – Проверка гипотезы о симметрии**

$$p_n(u=i)=C_i^n p^i q^{n-i}.$$

с параметром $n$, равным объему выборки.

Как видно из рисунка 1 для симметричного распределения $P=0{,}5$, а для несимметричного $P\neq 0{,}5$.

Введем основную гипотезу о симметрии распределения

$$H\colon f(y)=f(-y), \text{ или } P=0{,}5$$

и альтернативную гипотезу – о его несимметрии

$$\overline{H}\colon\ f(y)\neq f(-y), \text{ или } P\neq 0{,}5.$$

Учитывая, что при объемах выборки $n>20$ биномиальное распределение хорошо аппроксимируется распределением Гаусса, решающее правило по критерию Неймана-Пирсона можно записать в следующем виде: если $Z>C_1$ – верна альтернативная гипотеза; если $Z>C_1$ – верна основная гипотеза. При этом $C_1$ – порог решающего правила:

$$C_1=\frac{Z_\alpha}{2}\sqrt{n}+\frac{n}{2}-1. \tag{2}$$

Значение порога решающего правила выбирается из нижеприведенного условия, которое можно найти в книге [3]

$$\alpha=1-F\left[\frac{C_1+1-\frac{n}{2}}{\frac{\sqrt{n}}{2}}\right]. \tag{3}$$

Здесь $\alpha$ – параметр распределения Гаусса, называемый уровнем значимости (в литературных источниках этот параметр часто называют вероятностью ошибки первого рода, или «вероятностью ложной тревоги»), а $n$ – как уже было отмечено, является объемом выборки.

Такое решающее правило является несмещенным лишь для значения $P>0{,}5$. При $P<0{,}5$ несмещенным оказывается решающее правило $Z>C_2$, где порог

$$C_2=\frac{Z_{1-\alpha}}{2}\sqrt{n}+\frac{n}{2}-1. \tag{4}$$

При этом вероятности ошибки второго рода (пропуска сигнала) определяются из соотношений:

$$\beta_1=F\left[\frac{Z_{\alpha/2}-\sqrt{n}\,(p-0{,}5)}{\sqrt{pq}}\right]; \tag{5}$$

$$\beta_2 = F\left[\frac{Z_{1-\alpha/2} - \sqrt{n}\,(p-0,5)}{\sqrt{pq}}\right], \tag{6}$$

которые описаны в книге [4].

При незначительных объемах наблюдений ($n<20$) величину уровня значимости $\alpha$ можно определить в соответствии с распределением Бернулли

$$\alpha = P(y > C_1 \,|\, H) = \sum_{m=C_1}^{n} C_n^m 0,5^m 0,5^{n-m}, \tag{7}$$

откуда определяют значение порога $C_1$. Величина ошибки второго рода в этом случае определится из соотношения

$$\beta_1 = P(y > C_1 \,|\, \overline{H}) = \sum_{m=0}^{m=C_1} C_n^m p^m (1-p)^{n-m},$$

если параметр распределения $P>0,5$. В том же случае, когда $P<0,5$ величина ошибки второго рода должна определяться как

$$\beta_2 = P(y > C_2 \,|\, \overline{H}) = \sum_{m=C_2}^{n} C_n^m p^m (1-p)^{n-m}.$$

Процентные точки, а также распределения различных модификаций переменной (7) можно найти в работах [4, 5].

**Описание методов анализа**

Покажем, как с помощью непараметрических статистик можно осуществить обработку нестационарных случайных процессов.

По выборочным отсчетам измеряемого случайного процесса можно сформировать обобщенные инверсии первого порядка

$$T = \sum_{K=i+1}^{n} u(y_i, y_k), \tag{8}$$

где

$$u(y_i, y_k) = \begin{cases} 1, & y_i \geq y_k, \\ 0, & y_i < y_k, \end{cases}$$

и обобщенные инверсии второго порядка

$$T^2 = \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} u(y_i, y_k). \tag{9}$$

Переменная $T^2$ известна в литературе как статистика Кендалла.

Можно также сформировать обобщенные инверсии более высокого порядка

$$T^3 = \sum\sum\sum u(y_i, y_j, y_k), \tag{10}$$

и так далее.

Результаты исследований различных непараметрических статистик [6] показали, что мощность решающего правила при формировании гипотез растет с ростом порядка инверсий. Однако при этом растет и объем вычислений. Так, при использовании суммы инверсий третьего порядка объем затрат на вычисление статистики (10) более чем в три раза превышает последний для статистики Кендалла.

Для увеличения точности обработки измеряемого нестационарного случайного процесса необходимо разделять временной ряд наблюдений на оптимальные с точки зрения погрешности оценивания интервалы стационарности, длина которых в основном определяется вероятностными характеристиками нестационарной составляющей и является произвольной. Задача определения числа и длительности однородных интервалов, на которые необходимо разделить интервал наблюдений, особенно важна при отсутствии априорной информации о виде функции распределения случайного процесса. К тому же сведения о длительности интервалов квазистационарности представляют самостоятельный интерес, так как они одновременно являются интервалами сжатия данных. Учитывая вышесказанное, рассмотрим на примере статистики Кендалла различные процедуры разделения данных наблюдений на интервалы квазистационарности.

Использование статистики Кендалла (или любой другой непараметрической статистики) позволяет довольно легко разделить временной ряд $y(t)$ на конечное множество квазистационарных с наперед заданной вероятностью $P=I-\alpha$ интервалов времени по таким параметрам, как среднее $m[y(t)]$ и дисперсия $D[y(t)]$. Процедура деления заключается в вычислении текущих значений $T^2$ и допустимых границ $T^2_{\min}$ $[i;\ 1-\alpha/2]$ и $T^2_{\max}$ $[i;\ \alpha/2]$, и в проверке условия о стационарности по критерию Неймана-Пирсона [7]:

$$T^2_{\min} < T_i^2 \leq T^2_{\max}. \tag{11}$$

Более подробно с процедурой формирования статистической гипотезы о стационарности можно ознакомиться в работах [8, 9, 10].

**Обсуждение результатов**

В результате применения такой процедуры получается случайным образом сформированное некоторое количество участков квазистационарности, на которых среднее значение процесса с вероятностью $P$ считается постоянным.

Если использовать порядковые статистики, то легко можно определить среднее значение процесса на участке квазистационарности. Например, использование центральной порядковой статистики $x_{(c)}$ позволяет произвести следующее вычисление:

$$m = x_{(c)}.$$

Вычисление среднего значения по формуле $m=x_{(c)}$ позволяет по сути аппроксимировать нестационарную составляющую измеряемого процесса (тренда) ступенчатым полиномом (полиномом нулевой степени). На рисунке 2 приведены результаты компьютерного моделирования по определению нестационарного среднего случайного процесса вида $Y(t)=X(t)+F(t)$, где $F(t)=A(1-\exp(-a_1 t))$. Параметры $A$ и $a_1$ варьировались в различных пределах для целей моделиро-

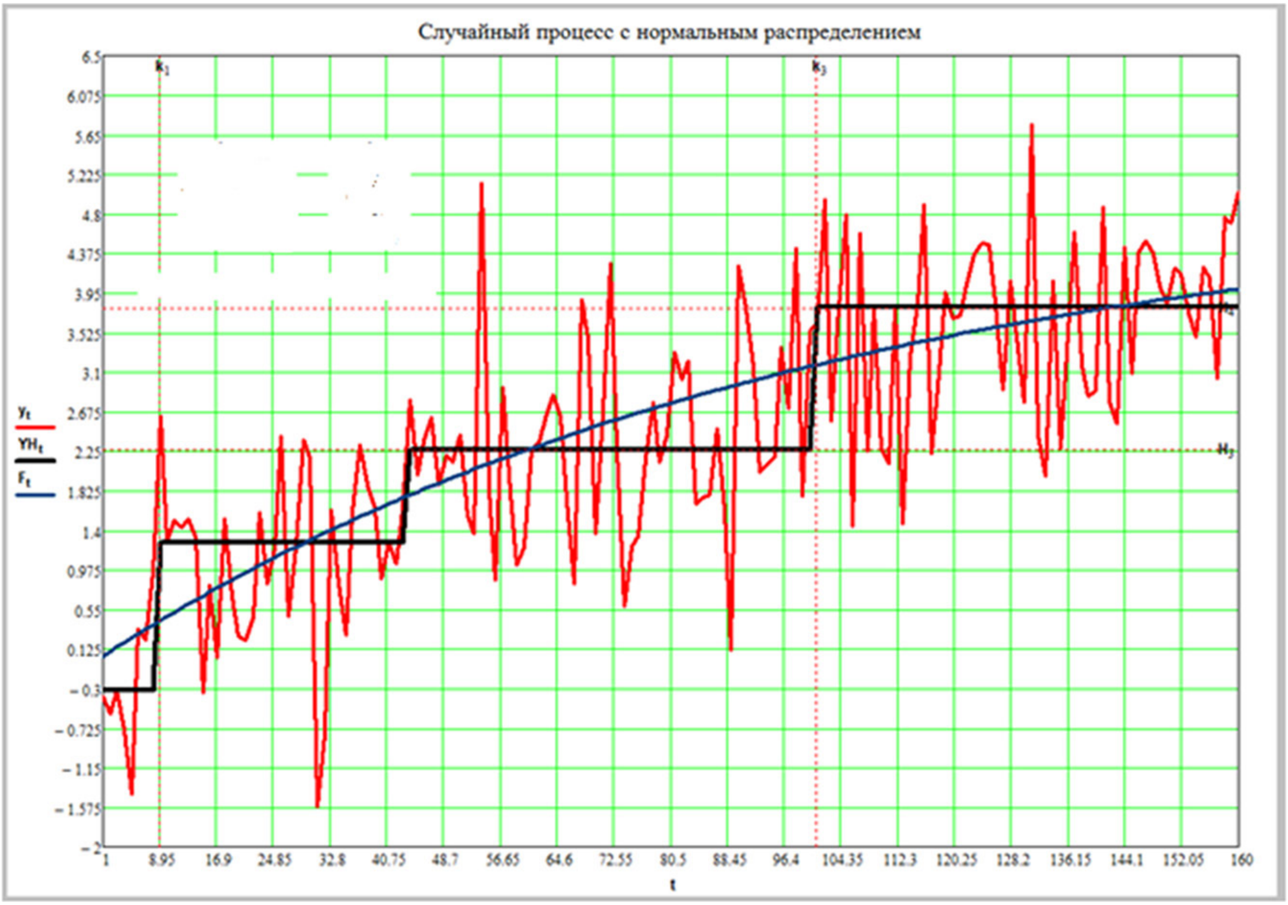


**Рисунок 2 – Определение нестационарной составляющей случайного процесса**

вания. На рисунке мы имеем четыре участка квазистационарности с длительностью в 9 отсчетов, 35 отсчетов, 57 и 59 отсчетов соответственно.

Моделирование проводилось в среде MathCad. Представленный рисунок является фотографией результатов моделирования. На рисунке моделируемый случайный процесс $Y(t)$ показан красным цветом, его среднее значение $F(t)$ – синим цветом. Черным цветом изображена функция $m=x_{(c)}$, которая является по сути аппроксимацией среднего значения полиномом нулевой степени.

**Заключение**

В статье были рассмотрены вопросы использования непараметрических методов теории решений при обработке информации в радиотелеметрических системах. В таких системах применение непараметрических методов является единственно возможным, если учесть, что обрабатываемые данные, как правило, представляют собой нестационарные широкополосные случайные процессы в условиях априорной неопределенности о статистических свойствах измеряемых процессов. Приведены результаты компьютерного моделирования по определению нестационарного среднего случайного процесса. Отметим, также, что применение рассмотренной процедуры определения нестационарного среднего значения измеряемого процесса позволяет получить очень большие значения коэффициента сжатия, что очень важно при проектировании информационно-измерительных систем. Однако следует иметь в виду, что использование служебной информации при формировании канального пакета данных [11], приводит к уменьшению коэффициента сжатия.

## СПИСОК ЛИТЕРАТУРЫ

1. Назаров А.В., Козырев Г.И. и др. Современная телеметрия в теории и на практике: Учебный курс. – СПб.: Наука и техника, 2007. – 667 с.
2. Тарасенко Ф.П. Непараметрическая статистика. – Томск: Издательство Томского университета, 1976. – 293 с.
3. Лившиц Н.А., Пугачев В.Н. Вероятностный анализ систем автоматического управления. – М.: Советское радио, 1968. – 504 с.
4. Виленкин С.Я. Статистическая обработка результатов исследования случайных функций. – М.: Энергия, 1979. – 320 с.

5. Левин Б.Р. Теоретические основы статистической радиотехники. – М.: Советское радио, 1968. – 504 с.
6. Есмагамбетов Б.-Б.С., Утепбергенов И.Т Использование непараметрических критериев при необратимом сжатии данных // Вестник Казахского Национального технического университета им. К. Сатпаева. – Алматы, 2006. – № 4 (54). – С. 87-91.
7. Гаек Я., Шидак З. Теория ранговых критериев. – М.: Наука, 1971. – 284 с.
8. Yesmagambetov B.S., Inkov A.M. Fast changing processes in radiotelemetry systems of space vehicles. Journal of Systems Engineering and Electronics. Vol. 26, no. 5. Beijing, October 2015, pp. 941-945.
9. Yesmagambetov B.S., Ajmenov Zh., Inkov A.Ismailov S., Saribaev A. Statistical data processing in rocket-space technology. Modern Applied Science, Vol. 9, no. 8. Canadian Center of Science and Education, 2015. pp. 317-334.
10. Есмагамбетов Б.-Б.С. Статистическая обработка данных в радиотелеметрических системах // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. № 1. – М., 2015. С. 13-21.
11. Есмагамбетов Б.-Б.С. Оптимизация служебной информации и помехоустойчивость в информационно-измерительных системах со сжатием данных // Университет еңбектері. Труды университета; Карагандинский технический университет. – Караганда, 2021. № 3 (84). С. 221-226.

***Кездейсоқ процестерді зерттеу кезіндегі шешімдер теориясының параметрлік емес әдістері***

***ЕСМАҒАМБЕТОВ Болатбатыр Сауқымұлы,*** *т.ғ.д., профессор, bulatbatyr@mail.ru,*
*М.О. Әуезов атындағы Оңтүстік Қазақстан университеті, Қазақстан, 160012, Шымкент, Тәуке хан даңғылы, 5.*

***Аңдатпа.*** *Кездейсоқ процестердің ықтимал сипаттамаларын бағалаудың дәстүрлі тәсілдері өңдеу кезінде стационарлық қасиеттердің болуын және осы өлшемдердің априорлық бөлінуін білуді талап етеді. Практикада мұндай талаптар жиі орындалмайды, себебі осы өлшеулердің едәуір бөлігі стационарлық емес кездейсоқ процестерге жатады. Бұл көптеген радиотехникалық жүйелерге жатады, мысалы, деректерді өңдеу үшін деректерді сығу әдістері деп аталатын ғарыштық ұшу аппараттарының радиотелеметриялық жүйелері [1]. Мұндай жүйелер, әдетте, өлшенетін параметрлер спектрінде кең жолақты жылдам өзгеретін процестердің үстем жағдайға ие болуымен ерекшеленеді, сондықтан баяу өзгеретін процестер үшін ғана деректерді квазимемлекеттік сығу әдістерін енгізу, біріншіден, берілетін деректер көлемін жиынтық қысқарту және деректерді беру арнасын түсіру проблемасын шешпейді, ал екіншіден, айтарлықтай қысқартуға мүмкіндік бермейді. Оның үстіне бортта өңделетін деректер тұрақты емес кездейсоқ процестердің жалғыз іске асырылуын білдіреді. Қазіргі математикалық статистиканың мұндай процестердің ықтимал сипаттамаларын бағалау үшін әдістері жоқ. Таратылу функциясы және оның параметрлері туралы априорлы белгісіздік жағдайында өлшеу деректерін өңдеу үшін статистикалық шешімдер теориясының параметрлік емес әдістерін пайдалануға болады. Мақалада Кендалл параметрлік емес статистикасының көмегімен стационарлық емес кездейсоқ процестің орташа мәнін бөлу әдісі қарастырылған.*

***Кілт сөздер:*** *параметрлік емес статистика, белгілік функция, реттік статистика, дәрежелік статистика, параметрлік емес гипотеза, кездейсоқ процесс, таратылу функциясы, жалпылама инверсиялар.*

***Non-Parametric Methods of Decision Theory in the Study of Random Processes***

***YESMAGAMBETOV Bulat-Batyr,*** *Dr. Tech. Sci., Professor, bulatbatyr@mail.ru,*
*M.O. Auezov South Kazakhstan University, Kazakhstan, 160012, Shymkent, Tauke Khan Avenue, 5.*

***Abstract.*** *Traditional methods of estimating the probabilistic characteristics of random processes require in processing the presence of stationary properties and knowledge of a priori distributions of measurement data. In practice, such requirements are often not met because much of the measurement data relates to transient random processes. This applies to many radio systems, for example, radio telemetry systems of spacecraft [1], in which so-called data compression techniques are used for data processing. Such systems differ in that, as a rule, broadband fast-changing processes dominate the spectrum of measured parameters, therefore, the introduction of quasi-reversible data compression methods only for slowly changing processes, firstly, does not solve the problem of a combined reduction in the volume of transmitted data and the unloading of the data channel, and secondly, it does not significantly reduce the time of experimental studies due to the inability to quickly use all the measured information. In addition, data processed on board very often represent the only implementation of non-stationary random processes. Modern mathematical statistics do not have methods for assessing the probabilistic characteristics of such processes. In conditions of a priori uncertainty about the distribution function and its parameters, non-parametric methods of statistical solution theory can be used to process measurement data. The paper discusses the method of selecting the average value of a non-stationary random process using non-parametric Kendall statistics.*

***Keywords:*** *non-parametric statistics, sign function, ordinal statistics, rank statistics, non-parametric hypothesis, random process, distribution function, generalized inversions.*

## REFERENCES


1. Nazarov A.V., Kozyrev G.I. i dr. Sovremennaya telemetriya v teorii i na praktike: Uchebnyj kurs [Modern telemetry in theory and practice. Training course]. Saint Petersburg: Nauka i tehnika, 2007. 667 p.
2. Tarasenko F.P. Neparametricheskaya statistika [Non-parametric statistics]. Tomsk: Publ. Tomskogo universiteta, 1976. 293 p.
3. Livshits N.A., Pugachev V.N. Veroiatnostnyi analiz system avtomaticheskogo upravleniya [Probabilistic analysis of automatic control systems]. Moscow: Sovetskoe radio, 1968. 504 p.
4. Vilenkin S.Ia. Statisticheskaya obrabotka rezultatov issledovaniya sluchainyh funktsij [Statistical Processing of Random Function Research Results]. Moscow: Energiya, 1979. 320 p.
5. Levin B.R. Teoreticheskie osnovy statisticheskoi radiotehniki [Theoretical Foundations of Statistical Radio Engineering]. Moscow: Sovetskoe radio, 1968. 504 p.
6. Esmagambetov B.-B.S., Utepbergenov I.T. Ispolzovanie neparametricheskih kriteriev pri neobratimom szhatii dannyh [Use non-parametric criteria for irreversible data compression] // Vestnik Kazahskogo Natsionalnogo tehnicheskogo universiteta im. K. Satpaeva. Almaty, 2006. – No. 4 (54), pp. 87-91.
7. Gaek Ia., Shidak Z. Teoriya rangovyh kriteriev [Theory of Rank Criteria]. Moscow: Nauka, 1971. 284 p.
8. Yesmagambetov B.S., Inkov A.M. Fast changing processes in radiotelemetry systems of space vehicles. Journal of Systems Engineering and Electronics. Vol. 26, no. 5. Beijing, October 2015, pp. 941-945.
9. Yesmagambetov B.S., Ajmenov Zh., Inkov A.Ismailov S., Saribaev A. Statistical data processing in rocket-space technology. Modern Applied Science, Vol. 9, no. 8. Canadian Center of Science and Education, 2015. pp. 317-334.
10. Esmagambetov B.-B.S. Statisticheskaya obrabotka dannyh v radiotelemetricheskih sistemah [Statistical processing of data in radiotelemetry systems] // Vestnik MGTU im. N.E. Baumana. Ser. Priborostroenie. Moscow, 2015. – no. 1, pp. 13-21.
11. Esmagambetov B.-B.S. Optimizatsiya sluzhebnoi informatsii i pomehoustoichivost v infrmatsionno-izmeritelnyh sistemah so szhatiem dannyh [Optimization of service information and noise immunity in information-measuring systems with data compression] // Universitet enbekteri. Trudy universiteta; Karagandinskij tekhnicheskij universitet. – Karaganda, 2021. No. 3 (84). – pp. 221-226.