Article

Asymptotic Properties of MSE Estimate for the False Discovery Rate Controlling Procedures in Multiple Hypothesis Testing

Sofia Palionnaya ^{1,2,*} and Oleg Shestakov ^{1,2,3,*}

¹ Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 119991 Moscow, Russia

² Moscow Center for Fundamental and Applied Mathematics, 119991 Moscow, Russia

³ Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 119333 Moscow, Russia

* Correspondence: sofiko-10@yandex.ru (S.P.); oshestakov@cs.msu.ru (O.S.)

Received: 14 October 2020; Accepted: 29 October 2020; Published: 1 November 2020

Abstract: Problems with analyzing and processing high-dimensional random vectors arise in a wide variety of areas. Important practical tasks are economical representation, searching for significant features, and removal of insignificant (noise) features. These tasks are fundamentally important for a wide class of practical applications, such as genetic chain analysis, encephalography, spectrography, video and audio processing, and a number of others. Current research in this area includes a wide range of papers devoted to various filtering methods based on the sparse representation of the obtained experimental data and statistical procedures for their processing. One of the most popular approaches to constructing statistical estimates of regularities in experimental data is the procedure of multiple testing of hypotheses about the significance of observations. In this paper, we consider a procedure based on the false discovery rate (FDR) measure that controls the expected percentage of false rejections of the null hypothesis. We analyze the asymptotic properties of the mean-square error estimate for this procedure and prove the statements about the asymptotic normality of this estimate. The obtained results make it possible to construct asymptotic confidence intervals for the mean-square error of the FDR method using only the observed data.

Keywords: false discovery rate; mean-square risk estimate; thresholding

1. Introduction

The problems involved in testing statistical hypotheses occupy an important place in applied statistics and are used in such areas as genetics, biology, astronomy, radar, computer graphics, etc. The classical methods for solving these problems are based on a single hypothesis test. There is a sample X of size m and the null hypothesis H_0 is tested against the general alternative H_1 . The hypothesis is tested using the statistic T , a function of the sample with a known distribution under the null hypothesis (zero distribution). For a given zero distribution, the attainable p -values are calculated, and the decision to reject the null hypothesis is made on their basis. Errors arising from the application of this one-time hypothesis testing algorithm are divided into two types, and the probability of falsely rejecting the correct null hypothesis (the probability of a type I error) is bounded by a given significance level α :

$$P(\text{type I error}) = P(T \geq t | H_0) \leq \alpha,$$

where t is the critical threshold value.

With this approach, we can often not only find the region for which the α -constraint on the probability of a type I error is satisfied, but also minimize the probability of a type II error, i.e., maximize the statistical power.

When considering the problem of multiple hypothesis testing, the task becomes more complicated: now we are dealing with n different null hypotheses $\{H_{0i}, i = 1, \dots, n\}$ and the alternatives $\{H_{1i}, i = 1, \dots, n\}$. These hypotheses are tested by statistics T_i with given zero distributions. Thus, for each hypothesis, the attainable p -values $\{p_i, i = 1, \dots, n\}$ can be calculated as well as type II error probabilities.

Let us introduce the notation: M_0 is the set of indices of true null hypotheses, R is the set of indices of rejected hypotheses. Then $V = |M_0 \cap R|$ is the number of type I errors. The task is to minimize V by changing the parameter R .

There are many statistical procedures that offer different ways to solve the multiple hypothesis testing problem. One of the first measures proposed to generalize the type I error was the family-wise error rate (FWER) [1]. This value is defined as the probability of making at least one type I error, i.e., instead of controlling the probability of a type I error at the level α for each test, the overall FWER is controlled: $\text{FWER} = P(V \geq 1) \leq \alpha$. However, such a strict criterion significantly increases the type II error for a large number of tested hypotheses.

In [2], an alternative measure called the false discovery rate (FDR) was proposed. This measure assumes to control the expected proportion of false rejections:

$$\text{FDR} = E\left(\frac{V}{\max(R, 1)}\right).$$

This approach is widely used in situations where the number of tested hypotheses is so large that it is preferable to allow a certain number of type I errors in order to increase the statistical power.

To control FDR, the Benjamini–Hochberg [2] multiple hypothesis testing algorithm is often used, which under the condition of independency of the testing statistics allows the FDR value to be bounded by the parameter α , i.e.,

$$E\left(\frac{V}{\max(R, 1)}\right) \leq \alpha.$$

In this procedure, the significance levels change linearly:

$$\alpha_i = \frac{i}{n}\alpha, \quad i = 1, \dots, n.$$

To apply the Benjamini–Hochberg method, a variational series is constructed from the attained p -values:

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}.$$

All hypotheses $H_{0_1}, \dots, H_{0_k}$ are rejected, where $k, k \in \{1, n\}$, is the maximum index for which

$$p_{(i)} \leq \alpha_i.$$

There are other measures to control the total number of type I errors. In [1], a q -value is considered that provides control of the positive false discovery rate (pFDR). Controlling the full coverage ratio (FCR) involves solving the problem of multiple hypothesis testing in terms of the confidence intervals [3]. The papers [4,5] are devoted to the harmonic mean p -value (HMP) method. However, in this paper we focus on the properties of the FDR method. It is believed that the widespread use of the FDR measure is due to the development of technologies that allow collecting and analyzing large amounts of data. Computing power makes it easy to perform hundreds or thousands of statistical tests on a given data set, and therefore the use of FWER loses its relevance.

In this paper, we study the asymptotic properties of the mean-square risk estimate for the FDR method in the problem of multiple hypothesis testing for the mathematical expectation of a Gaussian vector with independent components. The consistency of this estimate was proved in [6]. In this paper, we prove its asymptotic normality.

The paper is organized as follows. Section 2 provides some basic information about the statement of the problem and the considered vector classes. In Section 3 we define the mean-square risk of the thresholding method and describe the properties of the FDR-threshold. Section 4 considers the asymptotic properties of the mean-square risk estimate, and Section 5 contains some concluding remarks.

2. Preliminaries

Consider the problem of estimating the mathematical expectation of a Gaussian vector

$$X_i = \mu_i + W_i, \quad i = 1, \dots, n, \quad (1)$$

where W_i are independent normally distributed random variables with zero expectation and known variance σ^2 , and $\mu = (\mu_1, \dots, \mu_n)$ is an unknown vector belonging to some given set (class). The key assumption adopted in this paper is the “sparsity” of the vector μ , i.e., it is assumed that only a relatively small number of its components are significantly large. A similar problem statement arises, for example, in the analysis and processing of signals containing noise. In this case, the sparsity or “economical” representation of the signal is achieved using some special preprocessing, for example, a discrete wavelet transform of the signal vector.

In this paper, we consider the following definitions of sparsity. Let $\|\mu\|_0$ denote the number of nonzero components of μ . Fixing η_n , define the class

$$L_0(\eta_n) = \{\mu \in \mathbb{R}^n : \|\mu\|_0 \leq \eta_n n\}.$$

For small values of η_n , only a small number of vector components are nonzero.

Another possible way to define sparsity is to limit the absolute values of μ_i . To do this, consider the sorted absolute values

$$|\mu|_{(1)} \geq \dots \geq |\mu|_{(n)}$$

and for $0 < p < 2$ define the class

$$L_p(\eta_n) = \{\mu \in \mathbb{R}^n : |\mu|_{(k)} \leq \eta_n n^{1/p} k^{-1/p} \text{ for all } k = 1, \dots, n\}.$$

In addition, sparsity can be modeled using the ℓ_p -norm

$$\|\mu\|_p = \left(\sum_{i=1}^n |\mu_i|^p \right)^{1/p}.$$

In this case, the sparse class is defined as

$$M_p(\eta_n) = \{\mu \in \mathbb{R}^n : \sum_{i=1}^n |\mu_i|^p \leq \eta_n^p n\}.$$

There are important relationships between these classes. As $p \rightarrow 0$, the ℓ_p -norm approaches ℓ_0 : $\|\mu\|_p^p \rightarrow \|\mu\|_0$. The embedding $M_p(\eta_n) \subset L_p(\eta_n)$ is also valid.

3. Mean-Square Risk and Properties of the FDR Threshold

In the considered problem, one of the widespread and well-proven methods for constructing an estimate of μ is the method of (hard) thresholding of each vector component:

$$\hat{\mu}_i = \rho_H(X_i, T) = \begin{cases} X_i & |X_i| > T, \\ 0 & |X_i| \leq T, \end{cases} \quad (2)$$

i.e., the vector component is zeroed if its absolute value does not exceed the critical threshold T . This procedure is equivalent to testing the hypothesis of zero mathematical expectation for each component of the vector, and when using the FDR method, the threshold value T is selected according to the following rule. The initial sample is used to construct a variational series of decreasing absolute values

$$|X|_{(1)} \geq \dots \geq |X|_{(n)},$$

and $|X|_{(k)}$ are compared with the right tail Gaussian quantiles $t_k = \sigma z(\alpha/2 \cdot k/n)$. Let k_F be the largest index k for which $|X|_{(k)} \geq t_k$, then the threshold $T_F = t_{k_F}$ is chosen.

In combination with hypothesis testing methods, the penalty method is also widely used, in which the target loss function is minimized with the addition of a penalty term [7–9]. In a particular case, this method leads to the so-called soft thresholding: the estimates of the vector components are calculated according to the rule

$$\hat{\mu}_i = \rho_S(X_i, T) = \begin{cases} X_i - T & X_i > T, \\ X_i + T & X_i < -T, \\ 0 & |X_i| \leq T. \end{cases} \quad (3)$$

This approach is in some cases more adequate than (2), since the function ρ_S in (3) is continuous in T .

The mean-square error (or risk) of the considered procedures is determined as

$$R(T) = \sum_{i=1}^n E(\hat{\mu}_i - \mu_i)^2. \quad (4)$$

Methods for selecting the threshold value T are usually focused on minimizing the risk (4) provided that the vector μ belongs to a given class. A “perfect” value of the threshold is

$$T_{min} : R(T_{min}) = \min_T R(T).$$

Note that the expression (4) contains unknown values of μ_i and it is impossible to calculate $R(T)$ and T_{min} in practice. Therefore, a minimax approach is used. The threshold T_F is calculated based on the observed values of X_i and has the property of an adaptive minimax optimality in the considered sparse classes [7]. In addition, T_F has the following important property [7], which we will use later in proving the asymptotic normality of the risk estimate.

Theorem 1. [7] Suppose that $\mu \in L_0(\eta_n)$ or $\mu \in L_p(\eta_n)$, $0 < p < 2$, where $\eta_n \in [n^{-1}(\log n)^5, n^{-\gamma}]$ for $L_0(\eta_n)$ and $\eta_n^p \in [n^{-1}(\log n)^5, n^{-\gamma}]$ for $L_p(\eta_n)$, $0 < \gamma < 1$. Then there exists $c > 0$ such that for the FDR-threshold T_F with a controlling parameter $\alpha_n \rightarrow 0$ and large n ,

$$\begin{aligned} \sup_{\mu \in L_0(\eta_n)} P(T_F < T_1) &\leq 2n \exp\{-c \alpha_n \kappa_n \gamma_n^2\}, \\ \sup_{\mu \in L_p(\eta_n)} P(T_F < T_1) &\leq 2n \exp\{-c \alpha_n \kappa_n \gamma_n^2\}, \end{aligned}$$

where

$$\gamma_n = \frac{1}{\log \log n}, \kappa_n = \frac{n \eta_n}{1 - \alpha_n - \gamma_n}, T_1 = \sigma(2 \log \eta_n^{-1})^{1/2} \quad (5)$$

for $L_0(\eta_n)$ and

$$\gamma_n = \frac{1}{\log \log n}, \tau_\eta = \sigma(2 \log \eta_n^{-p})^{1/2}, \kappa_n = \frac{n \eta_n^p \tau_\eta^{-p}}{1 - \alpha_n - \gamma_n}, T_1 = \sigma(2 \log \eta_n^{-p})^{1/2} \quad (6)$$

for $L_p(\eta_n)$.

Thus, if α_n is chosen so that $\frac{\alpha_n \kappa_n \gamma_n^2}{\log n} \rightarrow \infty$, the value of T_1 is the lower bound for the threshold T_F .

Note also that the so-called universal threshold $T_U = \sigma \sqrt{2 \log n}$ is popular as well. This threshold is, in a certain sense, the maximum (it was shown in [10,11] that $T > T_U$ can be ignored). Based on this, we will assume everywhere that $T \leq T_U$.

4. Asymptotic Properties of the Risk Estimate

As already mentioned, since the expression (4) explicitly depends on the unknown values of μ_i , it cannot be calculated in practice. However, it is possible to construct its estimate, which is calculated using only the observed data. This estimate is determined by the expression

$$\hat{R}(T) = \sum_{i=1}^n F[X_i, T], \quad (7)$$

where $F[X_i, T] = (X_i^2 - \sigma^2)\mathbf{1}(|X_i| \leq T) + \sigma^2\mathbf{1}(|X_i| > T)$ for the hard thresholding and $F[X_i, T] = (X_i^2 - \sigma^2)\mathbf{1}(|X_i| \leq T) + (\sigma^2 + T^2)\mathbf{1}(|X_i| > T)$ for the soft thresholding [12].

In [6] it is proved that the estimate (7) is consistent.

Theorem 2. [6] Let the conditions of Theorem 1 be satisfied and $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ so that $\frac{\alpha_n \kappa_n \gamma_n^2}{\log n} \rightarrow \infty$, then

$$\frac{\hat{R}(T_F) - R(T_{min})}{n} \xrightarrow{P} 0.$$

Let us prove a statement about the asymptotic normality of the estimate (7), which, in particular, allows constructing asymptotic confidence intervals for the mean-square risk (4). In the proof, we will use the same notation C for different positive constants that may depend on the parameters of the classes and methods under consideration, but do not depend on n .

First, consider the class $L_0(\eta_n)$.

Theorem 3. Let $\mu \in L_0(\eta_n)$, $\eta_n \in [n^{-1}(\log n)^5, n^{-\gamma}]$, $1/2 < \gamma < 1$. Let T_F be the FDR-threshold with a controlling parameter $\alpha_n \rightarrow 0$ and $\frac{\alpha_n \kappa_n \gamma_n^2}{\log n} \rightarrow \infty$ as $n \rightarrow \infty$, where κ_n and γ_n are defined in (5). Then

$$\frac{\hat{R}(T_F) - R(T_{min})}{\sigma^2 \sqrt{2n}} \Rightarrow N(0, 1)$$

Proof. Let us prove the theorem for the soft thresholding method. In the case of hard thresholding, the proof is similar.

Denote

$$U(T) = \hat{R}(T) - \hat{R}(T_{min}) = \sum_{i=1}^n H_i(T, T_{min}),$$

where

$$H_i(T, T_{min}) = F[X_i, T] - F[X_i, T_{min}],$$

and write

$$\hat{R}(T_F) - R(T_{min}) + \hat{R}(T_{min}) - \hat{R}(T_{min}) = \hat{R}(T_{min}) - R(T_{min}) + U(T_F).$$

Let us show that

$$\frac{\hat{R}(T_{min}) - R(T_{min})}{\sigma^2 \sqrt{2n}} \Rightarrow N(0, 1) \quad (8)$$

With soft thresholding, $\hat{R}(T_{min})$ is an unbiased estimate of $R(T_{min})$, and with hard thresholding, under the conditions of the theorem the bias tends to zero when divided by $\sqrt{n}$ [12]. For the variance of the numerator [13]

$$\lim_{n \rightarrow \infty} \frac{D \sum_{i=1}^n (F[X_i, T_{min}] - EF[X_i, T_{min}])}{D \sum_{i=1}^n X_i^2} = 1. \quad (9)$$

Moreover, since X_i are independent, $DX_i^2 = 2\sigma^4 + 4\sigma^2\mu_i^2$ and the number of nonzero μ_i does not exceed $\eta_n n$, we obtain

$$\lim_{n \rightarrow \infty} \frac{D \sum_{i=1}^n X_i^2}{2\sigma^4 n} = 1. \quad (10)$$

Finally, the Lindeberg condition is met: for any $\epsilon > 0$ as $n \rightarrow \infty$

$$\frac{1}{V_n^2} \sum_{i=1}^n E[(F[X_i, T_{min}] - EF[X_i, T_{min}])^2 \mathbf{1}(|F[X_i, T_{min}] - EF[X_i, T_{min}]| > \epsilon V_n)] \rightarrow 0, \quad (11)$$

where $V_n^2 = D \sum_{i=1}^n (F[X_i, T_{min}] - EF[X_i, T_{min}])$. Indeed, due to (9) and (10) and since the summands in $\hat{R}(T_{min})$ are modulo bounded by the value $T_U^2 + \sigma^2$, starting from some n all indicators in (11) vanish. Therefore, (8) holds, and to prove the theorem it remains to show that

$$\frac{U(T_F)}{\sqrt{n}} \xrightarrow{P} 0.$$

Repeating the reasoning from [14–16] it can be shown that $T_{min} \geq T_1 - \alpha_n$, where $|\alpha_n| \leq C \frac{\log \log n}{\sqrt{\log n}}$. To shorten the notation without compromising the proof, we can omit α_n and assume that $T_{min} \geq T_1$. For any $\epsilon > 0$

$$\begin{aligned} P\left(\frac{|U(T_F)|}{\sqrt{n}} > \epsilon\right) &\leq P(T_F \leq T_1) + P\left(\frac{\sup_{T \in [T_1, T_U]} |U(T)|}{\sqrt{n}} > \epsilon\right) \leq \\ &P(T_F \leq T_1) + P\left(\frac{\sup_{T \in [T_1, T_U]} |U(T) - EU(T)| + \sup_{T \in [T_1, T_U]} |EU(T)|}{\sqrt{n}} > \epsilon\right). \end{aligned}$$

Let $U(T) = S_1(T) + S_2(T)$, $T \in [T_1, T_U]$, where the sum $S_2(T)$ contains terms with $\mu_i = 0$, and $S_1(T)$ contains all other terms. By the definition of the class $L_0(\eta_n)$, the number of terms in

$S_1(T)$ does not exceed $n_1 \approx \eta_n n$. Moreover, the absolute value of each term is bounded by $T_U^2 + \sigma^2$. For convenience, we will assume that $S_1(T)$ contains terms with indices from 1 to n_1 , i.e.,

$$S_1(T) = \sum_{i=1}^{n_1} H_i(T, T_{\min}), \quad S_2(T) = \sum_{i=n_1+1}^n H_i(T, T_{\min}).$$

Next, $EU(T) = ES_1(T) + ES_2(T)$ and $\sup_{T \in [T_1, T_U]} |ES_1(T)| \leq n_1(T_U^2 + \sigma^2)$. Given the definition of the class $L_0(\eta_n)$ and the form of T_1 , it can be shown that for the terms of S_2 the estimate $|EH_i(T, T_{\min})| \leq C(\log n)^{1/2} n^{-\gamma}$ is valid when $T \in [T_1, T_U]$. So

$$\sup_{T \in [T_1, T_U]} |EU(T)| \leq C \log n n^{1-\gamma}$$

and for $\gamma > 1/2$

$$\frac{\sup_{T \in [T_1, T_U]} |EU(T)|}{\sqrt{n}} \rightarrow 0 \quad (12)$$

as $n \rightarrow \infty$.

Next, take $T < T'$ and denote

$$Z_1(T) = S_1(T) - ES_1(T), \quad N_1(T, T') = \sum_{i=1}^{n_1} \mathbf{1}(T < |x_i| \leq T').$$

Then [10]

$$|Z_1(T) - Z_1(T')| \leq 4\sigma^2 N_1(T, T') + 2n_1(T'^2 - T^2) \text{ a.s.} \quad (13)$$

Divide the segment $[T_1, T_U]$ into equal parts: $T_j = T_1 + j\delta_{n_1} \in [T_1, T_U]$, $j = 1, \dots, n_1 - 1$, $\delta_{n_1} = (T_U - T_1)/n_1$. Then

$$A_n = \left\{ \sup_{T \in [T_1, T_U]} |S_1(T) - ES_1(T)| \geq 5\varepsilon\sqrt{n} \right\} \subset D_n \cup E_n,$$

where

$$D_n = \left\{ \sup_j |Z_1(T_j)| > \varepsilon\sqrt{n} \right\}, \quad E_n = \left\{ \sup_j \sup_{T \in [T_j, T_j + \delta_{n_1}]} |Z_1(T) - Z_1(T_j)| \geq 4\varepsilon\sqrt{n} \right\}.$$

Applying the Hoeffding inequality [17] for D_n , we obtain the estimate

$$P(D_n) \leq \sum_j P(|Z_1(T_j)| > \varepsilon\sqrt{n}) \leq 2n_1 \exp \left\{ -\frac{\varepsilon^2 n}{2(T_U^2 + 2\sigma^2)n_1} \right\} \leq 2n^{1-\gamma} \exp \left\{ -\frac{C\varepsilon^2 n^{1-\gamma}}{\log n} \right\}.$$

Then, given (13),

$$\begin{aligned} E_n &\subset \left\{ \sup_j \sup_{T \in [T_j, T_j + \delta_{n_1}]} 4\sigma^2 N_1(T_j, T) + 4n_1 \cdot T_U \delta_{n_1} \geq 4\varepsilon\sqrt{n} \right\} \subset \\ &\subset \left\{ \sup_j \sigma^2 N_1(T_j, T_j + \delta_{n_1}) + n_1 T_U \delta_{n_1} \geq \varepsilon\sqrt{n} \right\} = \left\{ \sup_j \sigma^2 N_1(T_j, T_j + \delta_{n_1}) \geq \varepsilon\sqrt{n} - n_1 T_U \delta_{n_1} \right\}. \end{aligned}$$

It is easy to show that $EN_1(T_j, T_j + \delta_{n_1}) \leq C n_1 \delta_{n_1}$. Hence,

$$\left\{ \sup_j \sigma^2 N_1(T_j, T_j + \delta_{n_1}) \geq \varepsilon \sqrt{n} - n_1 T_U \delta_{n_1} \right\} \subset \left\{ \sup_j \frac{1}{n_1} |N_1(T_j, T_j + \delta_{n_1}) - EN_1(T_j, T_j + \delta_{n_1})| \geq \frac{\varepsilon \sqrt{n}}{\sigma^2 n_1} - \frac{T_U \delta_{n_1}}{\sigma^2} - \frac{C}{\sigma^2} \delta_{n_1} \right\} = E'_n.$$

Applying the Hoeffding inequality, we obtain

$$\begin{aligned} P(E''_j) &= P\left(\frac{1}{n_1} |N_1(t_j, t_j + \delta_{n_1}) - EN_1(t_j, t_j + \delta_{n_1})| \geq \frac{\varepsilon \sqrt{n}}{\sigma^2 n_1} - \frac{T_U \delta_{n_1}}{\sigma^2} - \frac{C}{\sigma^2} \delta_{n_1}\right) \leq \\ &\leq 2 \exp \left\{ -2 \left(\frac{\varepsilon \sqrt{n}}{\sigma^2 n_1} - \frac{T_U \delta_{n_1}}{\sigma^2} - \frac{C}{\sigma^2} \delta_{n_1} \right)^2 n_1 \right\} \leq 2 \exp \left\{ -C \varepsilon^2 n^{1-\gamma} \right\}. \end{aligned}$$

Hence,

$$\mathbb{P}(E'_n) \leq \sum_{j=1}^{n_1} \mathbb{P}(E''_j) \leq C n^{1-\gamma} \exp \left\{ -C \varepsilon^2 n^{1-\gamma} \right\}.$$

Thus, for an arbitrary $\varepsilon > 0$

$$P\left(\sup_{T \in [T_1, T_U]} \frac{|S_1(T) - ES_1(T)|}{\sqrt{n}} > \varepsilon\right) \rightarrow 0 \quad (14)$$

as $n \rightarrow \infty$.

Let us now consider the sum $S_2(T)$. For large n , the number of terms in this sum is $n - n_1 \approx n$. Repeating the above reasoning, we divide the segment $[T_1, T_U]$ into equal parts: $T_j = T_1 + j\delta_{n_1} \in [T_1, T_U], j = 1, \dots, n-1, \delta_n = (T_U - T_1)/n$. Then

$$Z_2(T) = S_2(T) - ES_2(T), \quad N_2(T, T') = \sum_{i=n_1+1}^n \mathbf{1}(T < |x_i| \leq T'),$$

$$A_n = \left\{ \sup_{T \in [T_1, T_U]} |S_2(T) - ES_2(T)| \geq 5\varepsilon \sqrt{n} \right\} \subset D_n \cup E_n,$$

where

$$D_n = \left\{ \sup_j |Z_2(T_j)| > \varepsilon \sqrt{n} \right\}, \quad E_n = \left\{ \sup_j \sup_{T \in [T_j, T_j + \delta_n]} |Z_2(T) - Z_2(T_j)| \geq 4\varepsilon \sqrt{n} \right\}.$$

Taking into account the definition of the class $L_0(\eta_n)$ and the form of T_1 , we can bound the variance of the terms in S_2 (and hence Z_2): $DH_i(T, T_{min}) \leq C \left(\log \frac{n}{(\log n)^5} \right)^{3/2} n^{-\gamma}$. Then, applying Bernstein's inequality [18] for D_n , we obtain

$$P(D_n) \leq \sum_j P(|Z_2(T_j)| > \varepsilon \sqrt{n}) \leq 2n \exp \left\{ - \frac{C \varepsilon^2 n}{2 \left(\log \frac{n}{(\log n)^5} \right)^{3/2} n^{1-\gamma} + 2(T_U^2 + \sigma^2) \sqrt{n}} \right\}.$$

Next,

$$E_n \subset \left\{ \sup_j |N_2(T_j, T_j + \delta_n) - \mathbb{E}N_2(T_j, T_j + \delta_n)| \geq \frac{\varepsilon\sqrt{n}}{\sigma^2} - \frac{nT_U\delta_n}{\sigma^2} - \frac{Cn}{\sigma^2}\delta_n \right\} = E'_n.$$

The variance of the terms in $N_2(T_j, T_j + \delta_n)$ is bounded by $C \left(\log \frac{n}{(\log n)^5} \right)^{1/2} n^{-\gamma}$.

Applying Bernstein's inequality, we obtain

$$\begin{aligned} P(E''_j) &= P \left(\left| N_1(t_j, t_j + \delta_n) - \mathbb{E}N_1(t_j, t_j + \delta_n) \right| \geq \frac{\varepsilon\sqrt{n}}{\sigma^2} - \frac{nT_U\delta_n}{\sigma^2} - \frac{Cn}{\sigma^2}\delta_n \right) \leq \\ &\leq 2 \exp \left\{ - \frac{C\varepsilon^2 n}{2 \left(\log \frac{n}{(\log n)^5} \right)^{1/2} n^{1-\gamma} + 2\sqrt{n}} \right\}. \end{aligned}$$

Hence,

$$\mathbb{P}(E'_n) \leq \sum_{j=1}^n \mathbb{P}(E''_j) \leq 2n \exp \left\{ - \frac{C\varepsilon^2 n}{2 \left(\log \frac{n}{(\log n)^5} \right)^{1/2} n^{1-\gamma} + 2\sqrt{n}} \right\}.$$

Thus, for an arbitrary $\varepsilon > 0$

$$P \left(\sup_{T \in [T_1, T_U]} \frac{|S_2(T) - \mathbb{E}S_2(T)|}{\sqrt{n}} > \varepsilon \right) \rightarrow 0 \quad (15)$$

as $n \rightarrow \infty$.

Combining (8), (12), (14) and (15), we obtain the statement of the theorem. $\square$

A similar statement is true for the class $L_p(\eta_n)$.

Theorem 4. Let $\mu \in L_p(\eta_n)$, $0 < p < 2$, $\eta_n^p \in [n^{-1}(\log n)^5, n^{-\gamma}]$, $1/2 < \gamma < 1$. Let T_F be the FDR-threshold with a controlling parameter $\alpha_n \rightarrow 0$ and $\frac{\alpha_n \kappa_n \gamma_n^2}{\log n} \rightarrow \infty$ as $n \rightarrow \infty$, where κ_n and γ_n are defined in (6). Then

$$\frac{\hat{R}(T_F) - R(T_{min})}{\sigma^2 \sqrt{2n}} \Rightarrow N(0, 1).$$

Proof. The main steps in the proof of this theorem repeat the proof of Theorem 3. We also write

$$\hat{R}(T_F) - R(T_{min}) + \hat{R}(T_{min}) - \hat{R}(T_{min}) = \hat{R}(T_{min}) - R(T_{min}) + U(T_F).$$

The statement

$$\frac{\hat{R}(T_{min}) - R(T_{min})}{\sigma^2 \sqrt{2n}} \Rightarrow N(0, 1)$$

is proved exactly the same as the statement (8). Let $U(T) = S_1(T) + S_2(T)$, $T \in [T_1, T_U]$, where the sum $S_2(T)$ contains terms with $|\mu_i| \leq C/T_1$, and $S_1(T)$ contains all other terms. By the definition of the class $L_p(\eta_n)$, the number of terms in $S_1(T)$ does not exceed $n_1 \approx C\eta_n^p n$ and each term is modulo bounded by $T_U^2 + \sigma^2$. Considering the form of T_1 , it can be shown that the mathematical expectations of the terms in S_2 do not exceed $C(\log n)^{1/2} n^{-\gamma}$, and their variances do not exceed $C \left(\log \frac{n}{(\log n)^5} \right)^{3/2} n^{-\gamma}$. Next, arguing as in Theorem 3, we see that

$$\frac{\sup_{T \in [T_1, T_U]} |EU(T)|}{\sqrt{n}} \rightarrow 0$$

and

$$P \left(\sup_{T \in [T_1, T_U]} \frac{|U(T) - EU(T)|}{\sqrt{n}} > \varepsilon \right) \rightarrow 0$$

for an arbitrary $\varepsilon > 0$ as $n \rightarrow \infty$. Thus, since $P(T_F \leq T_1) \rightarrow 0$,

$$\frac{U(T_F)}{\sqrt{n}} \xrightarrow{P} 0,$$

as $n \rightarrow \infty$. $\square$

The above statements demonstrate that the considered method for constructing estimates in the model (1) has very similar properties to the method based on minimizing the estimate (7) in the parameter T (see [19]).

5. Conclusions

In this paper, we considered a method of estimating the mean of a Gaussian vector based on the procedure of multiple hypothesis testing. The estimation is based on the false discovery rate measure, which controls the expected percentage of false rejections of the null hypothesis. It is common to use the mean-square risk for evaluating the performance of this approach. Its value cannot be calculated in practice, so its estimate must be considered instead. We analyzed the asymptotic properties of this estimate and proved that it is asymptotically normal for the classes of sparse vectors. This result justifies the use of the mean-square risk estimate for practical purposes and allows constructing asymptotic confidence intervals for a theoretical mean-square risk. For more accurate analysis it is desirable to have guaranteed confidence intervals. These intervals could be constructed based on the estimates of the convergence rate in Theorems 3 and 4. Guaranteed confidence intervals would help to understand how the results of Theorems 3 and 4 affect the risk estimation for a finite sample size. We therefore leave the problem of estimating the rate of convergence and numerical simulation for future work.

Author Contributions: Conceptualization, O.S.; methodology, S.P. and O.S.; formal analysis, S.P. and O.S.; investigation, S.P. and O.S.; writing—original draft preparation, S.P. and O.S.; writing—review and editing, S.P. and O.S.; supervision, O.S.; funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2020-799.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Storey, J.D. A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **2002**, *64*, 479–498. [\[CrossRef\]](#)
2. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [\[CrossRef\]](#)
3. Benjamini, Y.; Yekutieli, D. False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Am. Stat. Assoc.* **2005**, *100*, 71–93. [\[CrossRef\]](#)
4. Wilson, D.J. The harmonic mean p -value for combining dependent tests. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1195–1200. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Wilson, D.J. Reply to Held: When is a harmonic mean p -value a Bayes factor? *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5857–5858. [\[CrossRef\]](#) [\[PubMed\]](#)

6. Zaspas, A.Y.; Shestakov, O.V. Consistency of the risk estimate of the multiple hypothesis testing with the FDR threshold. *Her. Tver State Univ. Ser. Appl. Math.* **2017**, *1*, 5–16. [[CrossRef](#)]
7. Abramovich, F.; Benjamini, Y.; Donoho, D.; Johnstone, I. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **2006**, *34*, 584–653. [[CrossRef](#)]
8. Donoho, D.; Jin, J. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* **2006**, *34*, 2980–3018. [[CrossRef](#)]
9. Neuvial, P.; Roquain, E. On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **2012**, *40*, 2572–2600. [[CrossRef](#)]
10. Donoho, D.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [[CrossRef](#)]
11. Marron, J.S.; Adak, S.; Johnstone, I.M.; Neumann, M.H.; Patil, P. Exact risk analysis of wavelet regression. *J. Comput. Graph. Stat.* **1998**, *7*, 278–309.
12. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: New York, NY, USA, 1999.
13. Markin, A.V. limit distribution of risk estimate of wavelet coefficient thresholding. *Inform. Appl.* **2009**, *3*, 57–63.
14. Jansen, M. *Noise Reduction by Wavelet Thresholding, Volume 161 of Lecture Notes in Statistics*; Springer: New York, NY, USA, 2001.
15. Kudryavtsev, A.A.; Shestakov, O.V. Asymptotic behavior of the threshold minimizing the average probability of error in calculation of wavelet coefficients. *Dokl. Math.* **2016**, *93*, 295–299. [[CrossRef](#)]
16. Kudryavtsev, A.A.; Shestakov, O.V. Asymptotically optimal wavelet thresholding in models with non-gaussian noise distributions. *Dokl. Math.* **2016**, *94*, 615–619. [[CrossRef](#)]
17. Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **1963**, *58*, 13–30. [[CrossRef](#)]
18. Bennett, G. Probability inequalities for the sum of independent random variables. *J. Am. Stat. Assoc.* **1962**, *57*, 33–45. [[CrossRef](#)]
19. Shestakov, O.V. Asymptotic normality of adaptive wavelet thresholding risk estimation. *Dokl. Math.* **2012**, *86*, 556–558. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).