Article

On the Convergence of the Benjamini–Hochberg Procedure

Dean Palejev 1,2,* and Mladen Savov 1,3,*

1 Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
2 Big Data for Smart Society (GATE) Institute, Sofia University “St. Kliment Ohridski”, 1113 Sofia, Bulgaria
3 Faculty of Mathematics and Informatics, Sofia University “St. Kliment Ohridski”, 1164 Sofia, Bulgaria
* Correspondence: palejev@math.bas.bg (D.P.); mladensavov@math.bas.bg (M.S.)

Abstract: The Benjamini–Hochberg procedure is one of the most used scientific methods up to date. It is widely used in the field of genetics and other areas where the problem of multiple comparison arises frequently. In this paper we show that under fairly general assumptions for the distribution of the test statistic under the alternative hypothesis, when increasing the number of tests, the power of the Benjamini–Hochberg procedure has an exponential type of asymptotic convergence to a previously shown limit of the power. We give a theoretical lower bound for the probability that for a fixed number of tests the power is within a given interval around its limit together with a software routine that calculates these values. This result is important when planning costly experiments and estimating the achieved power after performing them.

Keywords: multiple hypothesis testing; Benjamini–Hochberg procedure; convergence rate

1. Introduction

In many modern studies scientists perform the same statistical test on many objects, e.g., genes or single-nucleotide polymorphisms. The number of these objects could be very large, say millions, resulting in a multiple comparison issue. This is due to the fact that even under the null hypothesis we expect some of these tests to have $p$-values below a predetermined significance level $\alpha$. There are various techniques to overcome this problem with the Benjamini–Hochberg method [1] being the most used. Given $a$, for a set of $m$ $p$-values, the method rejects the tests corresponding to the smallest $k$ $p$-values, where $k$ is the largest index such that $p(k) \leq k\alpha/m$. The method controls the FDR (false discovery rate) in a sense that the expected proportion of false discoveries out of all discoveries (rejected hypotheses) is not more than $\alpha$ when the tests are independent as shown in the original paper [1]. In some cases, rather than finding the cutoff index for rejecting tests, the so-called Benjamini–Hochberg-adjusted $p$-values are used; these are the $p$-values transformed by the formula on the right-hand side of the above inequality and truncated at 1.

In the field of genetics it is used in many software packages, for example, DESeq2 [2], edgeR [3], cuffdiff [4], GenomeStudio (Illumina). It is also often used when studying various types of images, e.g., PET or CT [5], fMRI [6], or astrophysical [7].

Despite the fact that there are improved methods either in terms of weaker assumptions, e.g., dependency or better statistical power, e.g., [8–10], the Benjamini–Hochberg method is one of the most cited scientific articles, determined to be 59th most cited article in [11]. We should underline that in this work we are not interested in improving the Benjamini–Hochberg method, but rather than that in inferring about the statistical power of the original method as it is the most used one in practice. Moreover, article [12] shows that no correction of the standard methods is needed in certain cases even when there is some degree of dependence among the tests. See Section 5 for more information and related work.

Calculating the statistical power, that is the probability of correctly rejecting the null when the alternative hypothesis is true, is an important task when designing scientific
studies. For example it allows us to estimate the required sample size, and therefore the budget needed to perform the experiment, in order to make sure that we have a reasonable chance of detecting the objects for which there is indeed a difference between the two groups. After performing the study, it allows us to determine whether we have found a relatively large proportion of significant objects.

For practical applications we are interested not only in the limit of the statistical power as shown in previous works (see Section 5), but also in the speed of convergence to that limit. The speed of convergence would allow us to find the probability that the statistical power is within a given range around its limit. Consequently this would allow us to calculate the probability that the statistical power is above a pre-determined threshold, typically 80%. In cases when this probability is relatively low, one would have to increase the sample size in order to have more favorable parameters of the Beta distribution that models the p-values under the alternative hypothesis.

In more detail, in this work we show that under natural condition, that we call (M), the convergence is exponential and give explicit expression of the constants involved in the asymptotics. That condition allows the usage of suitable Beta distributions amongst others. The relationship between our condition and the non-criticality condition in [13] will be discussed below. Condition (M) allows us to find a lower bound for the probability that for fixed parameters of the model the power is within given interval around that asymptotic limit. In this sense our work can be understood as a tool which can be used in the planning of the experiment and which complements [14], another article by the authors of [15], which deals with a variety of questions, such as estimation of the proportion of significant tests, the effect of the sample size to the quality of results, etc.

2. Main Results and Their Proofs

2.1. Some Preliminary Notation

Let $X_1, X_2, \cdots, X_m$, $m \in \mathbb{N}^+$, with $\mathbb{N}^+ = \{1, 2, \cdots\}$, be independent random variables that are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For some $1 \leq m_0 < m$ the random variables $(X_i)_{i \leq m_0}$ are assumed to have a common cumulative distribution function (c.d.f.) $F$, whereas $(X_i)_{m_0 < i \leq m}$ are assumed to have a common c.d.f. $G$. In this work we set $F$ to be the c.d.f. of the uniform random variable on $(0,1)$ and let for the time being $G$ be any continuous c.d.f. We note that even if $G$ is discrete, one can always approximate it with a suitable continuous distribution. We also work with the condition that $\frac{m_0}{m} = \gamma \in (0,1)$ is fixed and set $m_1 := m - m_0 \in (0, m)$, which is strictly non-binding since we develop exponential bounds that are valid with different degree of accuracy for any fixed $m_0, m$ and the respective ratio $\frac{m_0}{m}$. It will be usually the case that $\gamma$ is larger than 1/2 as the first $m_0$ random variables represent the non-significant observations. If $X^{(i)}_i$, $i \leq m$, is $i$-th order statistics of $X_1, X_2, \cdots, X_m$ then for a fixed level of rejection $\alpha \in (0,1)$, the Benjamini–Hochberg procedure declares significant the first $R_m$ order statistics, where

$$R_m = \max \left\{ i : X^{(i)}_m \leq \alpha \frac{i}{m} \right\}. \quad (1)$$

The truly non-significant tests have c.d.f $F$ and the truly significant tests possess c.d.f $G$. It is well known from [15] (1.2) that for $x \in (0,1)$ we can express the event $\{R_m \geq mx\}$ in the following fashion

$$\left\{ \frac{R_m}{m} \geq x \right\} = \left\{ \max_{t \in A_k(m,x)} \frac{H_m(t) - t}{t} \geq \frac{1}{\alpha} - 1 \right\}, \quad (2)$$

where $H_m(t) = \frac{1}{m} \sum_{j=1}^{m} I\{X_j \leq t\}$ is the empirical c.d.f. of $X_1, X_2, \cdots, X_m$ and

$$A_k(m,x) = \left\{ \frac{k}{m} \alpha : \lfloor mx \rfloor \leq k \leq m \right\}, \quad (3)$$
Let us compute $K$ better to use the sharper estimate on the function $c$ in $K$

Remark 2. Given that $c$

General Inequalities and Information concerning the Distribution of $R_m$

We introduce the function $c(\epsilon) = (1 + \epsilon) \ln(1 + \epsilon) - \epsilon$, for $\epsilon > 0$, which is investigated in Proposition A1 in Appendix A. Then the following inequalities hold true.

**Proposition 1.** For any distribution function $T$ with $T(ax) > 0$ and empirical cumulative distribution function $T_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq \cdot\}$, where $(X_i)_{i \geq 1}$ are i.i.d. random variables with c.d.f. $T$, any $k, m \in \mathbb{N}^+$, any $a, x, \rho \in (0, 1)$ and $A_\rho(x, m)$ as in (3) above, we have, for every $\epsilon > 0$, that

$$
\mathbb{P}\left( \rho \max_{t \in A_\rho(m, x)} \frac{T_k(t) - T(t)}{t} > \epsilon \right)
\leq 2 \min\left\{ m(1 - x) e^{-k T(ax)} c\left( \frac{x}{\rho T(t)} \right), e^{-\frac{2 k}{\rho}} (\epsilon ax)^2 \right\}
:= K_T(x, m, k, \rho, \epsilon, a).
$$

Finally, if $T(t)c\left( \frac{x}{\rho T(t)} \right)$ as a function of $t$ is non-decreasing on $[ax, a]$ we get that

$$
\mathbb{P}\left( \rho \max_{t \in A_\rho(m, x)} \frac{T_k(t) - T(t)}{t} > \epsilon \right)
\leq 2 \min\left\{ m(1 - x) e^{-k T(ax)} c\left( \frac{x}{\rho T(t)} \right), e^{-\frac{2 k}{\rho}} (\epsilon ax)^2 \right\} = K_T(x, m, k, \rho, \epsilon, a).
$$

**Remark 1.** Let us compute $K_T$ when $T(t) = F(t) = t$ for $t \in (0, 1)$. Then $T(t)c\left( \frac{x}{\rho T(t)} \right) = tc\left( \frac{x}{\rho} \right)$ is non-decreasing in $t$ and substituting in the right-hand side of (6) we deduce that

$$
K_T(x, m, k, \rho, \epsilon, a) = 2 \min\left\{ m(1 - x) e^{-k T(ax)} c\left( \frac{x}{\rho T(t)} \right), e^{-\frac{2 k}{\rho}} (\epsilon ax)^2 \right\}.
$$

**Remark 2.** Given that $c(\epsilon) \geq \min\left\{ 1, \frac{\epsilon}{2} \right\}$, see Proposition A1 we may eradicate the dependence on the function $c$ in $K_T$, see (5) by providing an upper bound for it:

$$
K_T(x, m, k, \rho, \epsilon, a) \leq 2 \min\left\{ m(1 - x) e^{-k \min\{ T(ax), \inf_{t \in A_\rho(m, x)} t^2 / 2 \pi^2 \}, e^{-\frac{2 k}{\rho}} (\epsilon ax)^2 \right\}.
$$

**Remark 3.** Inspecting the proof below one easily discovers that for computational purposes it is better to use the sharper estimate

$$
\mathbb{P}\left( \rho \max_{t \in A_\rho(m, x)} \left| \frac{T_k(t) - T(t)}{t} \right| > \epsilon \right) \leq 2 \min\left\{ \sum_{t \in A_\rho(m, x)} e^{-k T(t)} c\left( \frac{x}{\rho T(t)} \right), e^{-\frac{2 k}{\rho}} (\epsilon ax)^2 \right\}.
$$
For theoretical purposes the bounds stated in the theorem are more suitable.

**Proof of Proposition 1.** The second part of the minimum (5) follows from the Dvoretzky–Kiefer–Wolfowitz inequality, see [16], which gives, for any \( x, \epsilon > 0 \) that

\[
\mathbb{P}\left( \rho \max_{t \in A_\alpha(m,x)} \left| \frac{T_k(t) - T(t)}{t} \right| > \epsilon \right) \leq \mathbb{P}\left( \rho \sup_{t \in (a,x)} \left| \frac{T_k(t) - T(t)}{t} \right| > \epsilon \right)
\]

\[
\leq \mathbb{P}\left( \sup_{t \in (a,x)} \left| T_k(t) - T(t) \right| > \frac{e \alpha x}{\rho} \right)
\]

\[
\leq 2e^{-2k \frac{(\alpha x)^2}{\rho^2}}.
\]

In Proposition A1 we use the function \( c(\epsilon) = (1 + \epsilon) \ln(1 + \epsilon) - \epsilon \). The first part of the minimum of (5) is immediate from Proposition A1 since \( kT_k(t) = \sum_{j=1}^{k} \mathbb{1}_{\{X_j \leq t\}} \) is binomially distributed with parameters \( k \) and \( T(t) = \mathbb{P}(X_1 \leq t) \), and thus

\[
\mathbb{P}\left( \rho \max_{t \in A_\alpha(m,x)} \left| \frac{T_k(t) - T(t)}{t} \right| > \epsilon \right) \leq \sum_{t \in A_\alpha(m,x)} \mathbb{P}\left( kT_k(t) - kT(t) > \frac{e \alpha x}{\rho} T(t) \right)
\]

\[
\leq 2 \sum_{t \in A_\alpha(m,x)} e^{-kT(t)c\left( \frac{\epsilon \alpha x}{\rho T(t)} \right)}
\]

\[
\leq 2|A_\alpha(m,x)|e^{-k \inf_{t \in A_\alpha(m,x)} T(t)c\left( \frac{\epsilon \alpha x}{\rho T(t)} \right)}
\]

\[
\leq 2m(1-x)e^{-k \inf_{t \in A_\alpha(m,x)} T(t)c\left( \frac{\epsilon \alpha x}{\rho T(t)} \right)},
\]

where we have used the fact that the number of elements in \( A_\alpha(m,x) \) is bounded from above as follows \( |A_\alpha(m,x)| \leq m(1-x) \). Next, (6) is deduced from the assumption that \( T(t)c\left( \frac{\epsilon \alpha x}{\rho T(t)} \right) \) as a function of \( t \) is non-decreasing on \( [a,x, a) \) and the simple fact \( A_\alpha(m,x) \subseteq [a,x, a) \). \( \square \)

Applying Proposition 1 with \( T = F \) and \( T = G \) and respectively \( m_0 = m \gamma \) and \( m_1 = m(1- \gamma) \) and \( \rho = \gamma \) and \( \rho = 1- \gamma \) we can confirm from (4) the proof of Theorem 3.1 in [15] in the sense that we have the convergence in probability

\[
\lim_{m \to \infty} \max_{t \in \left\{ \frac{m}{m_0}, \ldots, \frac{m}{m_1}, a \right\}} \frac{H_m(t) - t}{t} = (1- \gamma) \sup_{[a,x, a]} \frac{G(t) - t}{t} =: u_{\alpha, \gamma}(x).
\]

Clearly, \( u_{\alpha, \gamma} : [0,1] \to \mathbb{R} \), with \( u_{\alpha, \gamma}(1) = \frac{G(a) - a}{a} \), is a non-increasing function on \( (0,1] \). Denote

\[
\underline{\Lambda}_{\alpha, \gamma} = \inf\{ x \in (0,1) : u_{\alpha, \gamma}(x) \leq \frac{1}{\beta} - 1 \}, \alpha, \beta \in (0,1),
\]

and

\[
\overline{\Lambda}_{\alpha, \gamma} = \inf\{ x \in (0,1) : u_{\alpha, \gamma}(x) < \frac{1}{\beta} - 1 \}, \alpha, \beta \in (0,1).
\]

Since \( u_{\alpha, \gamma} \) is non-increasing we have that \( \underline{\Lambda}_{\alpha, \gamma} \geq \overline{\Lambda}_{\alpha, \gamma} \). Therefore from relations (2) and (4) we can easily get that

\[
\lim_{m \to \infty} \mathbb{P}\left( \frac{R_m}{m} \geq x \right) = 1, \quad \text{for } x < \underline{\Lambda}_{\alpha, \gamma}
\]

\[
\lim_{m \to \infty} \mathbb{P}\left( \frac{R_m}{m} \geq x \right) = 0, \quad \text{for } x > \overline{\Lambda}_{\alpha, \gamma}.
\]
We can provide more precise information about probabilities of the type $P\left(\frac{R_m}{m} < x\right)$ in the general case. In the sequel we use $\mathbb{1}_{\{A\}}$ for the indicator function of the set $A$.

**Proposition 2.** We have for any $\epsilon > 0$ and $x \in (0, 1)$ that
\begin{align*}
P\left(\frac{R_m}{m} < x\right) & \geq 1 - K_F(x, m, m_0, \gamma, \epsilon, \alpha) - K_G(x, m, m_1, 1 - \gamma, \epsilon, \alpha) \\
& - \mathbb{1}_{\{(1 - \gamma) \max_{t \in A_t(x,m)} \frac{G(t) - t}{t} \geq \frac{1}{2} - 1 - 2\epsilon\}}.
\end{align*}

**Proof.** Using successively (2), (4) and (10) with $\rho = \gamma, k = m_0 = \gamma m$ for the probability involving $F_{m_0}$ and $\rho = 1 - \gamma, k = m_1 = (1 - \gamma)m$ for the expression concerning $G_{m_1}$, we arrive at the inequalities
\begin{align*}
P\left(\frac{R_m}{m} < x\right) & = 1 - P\left(\frac{R_m}{m} \geq x\right) \\
& = 1 - P\left(\left\{ \max_{t \in A_t(x,m)} \frac{H_m(t) - t}{t} \geq \frac{1}{\alpha} - 1 \right\} \right) \\
& \geq 1 - P\left(\gamma \max_{t \in A_t(x,m)} \left| \frac{F_m(t) - t}{t} \right| > \epsilon \right) - P\left(\left(1 - \gamma\right) \max_{t \in A_t(x,m)} \left| \frac{G_m(t) - G(t)}{t} \right| > \epsilon \right) \\
& - \mathbb{1}_{\{(1 - \gamma) \max_{t \in A_t(x,m)} \frac{G(t) - t}{t} \geq \frac{1}{2} - 1 - 2\epsilon\}} \times \\
& \times P\left(\max_{t \in A_t(x,m)} \left(\gamma \frac{F_m(t) - t}{t} \right) \left(1 - \gamma\right) \frac{G_m(t) - G(t)}{t} \leq \epsilon \right) \\
& \geq 1 - K_F(x, m, m_0, \gamma, \epsilon, \alpha) - K_G(x, m, m_1, 1 - \gamma, \epsilon, \alpha) \\
& - \mathbb{1}_{\{(1 - \gamma) \max_{t \in A_t(x,m)} \frac{G(t) - t}{t} \geq \frac{1}{2} - 1 - 2\epsilon\}}.
\end{align*}

We only note that in the last expression of the first inequality above we have used that once the dependence on $F_{m_0}, G_{m_1}$ is estimated away, the remaining term is deterministic and thus the indicator function appears. \(\square\)

### 2.3. Condition (M)

In view of the fact that Theorem 3.1 in [15] holds, the limit in probability, that is $\lim_{m \to \infty} \frac{R_m}{m} = x_{a, \gamma}$ exists if and only if $x^a_{a, \gamma} = x^w_{a, \gamma} = x_{a, \gamma}$. We shall therefore restrict our attention only to the case when for fixed $a$ we have that the function $\frac{G(t) - t}{t}$ is decreasing on $(0, \alpha)$. In this case we have from (11) that
\begin{equation}
\mu_{x, \gamma}(x) = (1 - \gamma) \frac{G(ax) - ax}{ax}.
\end{equation}

If in addition $\lim_{t \to 0} \frac{G(t) - t}{t} = \infty$ then
\begin{equation}
x^\beta_{a, \gamma} := x^\beta_{a, \gamma} = x^\beta_{a, \gamma}, \text{ for any } \beta < 1,
\end{equation}
and if $\beta = a$ then we set $x_{a, \gamma} := x^a_{a, \gamma}$. From now on we work under this condition which we call condition (M) and which is synthesized as follows.

**Definition 1.** We say that condition (M) holds for fixed $a \in (0, 1)$ if and only if $H(t) := \frac{G(t) - t}{t}$ is decreasing on $(0, \alpha]$ with $\lim_{t \to 0} H(t) = \infty$.

**Remark 4.** As mentioned in [17], "[u]nder the alternative hypothesis, the $p$-values will have a distribution that has high density for small $p$-values and the density will decrease as the $p$-values"
Theorems 1 and 3 strengthen this by showing that the speed of convergence is of an exponential order. As discussed earlier, this allows a direct estimate on the probability that $S_n$ is in a variety of other examples. Moreover, their results go beyond the scope of the Benjamini–Hochberg procedure. In the particular case of the Benjamini–Hochberg procedure, however, we know from [15] that the probabilities of the events $\left\{ \left| \frac{R_n}{n} - x_{\alpha,\gamma} \right| \leq \beta x_{\alpha,\gamma} \gamma \right\}$ and of $\left\{ \left| \frac{S_n}{n} - x_{\alpha,\gamma} \gamma \right| \leq \beta x_{\alpha,\gamma} \gamma \right\}$, where $S_n/n$ is the power of that procedure, converge to 1. Our Theorems 1 and 3 strengthen this by showing that the speed of convergence is of an exponential order. As discussed earlier, this allows a direct estimate on the probability that $S_n/n$ is in a prescribed interval.

Remark 6. We note that upon the validity of condition (M) it is impossible that $G$ has atoms on $(0, \alpha)$ as otherwise if $a \in (0, \alpha)$ is an atom then

$$\lim_{t \downarrow a} H(t) = H(a) > \lim_{t \uparrow a} H(t) = H(a^-),$$

which leads to a contradiction with the assumption that $H$ is decreasing. Thus, our initial requirement that $G$ is continuous is included in condition (M). This also means that $H$ is continuous on $(0, \alpha)$ and thus $\frac{\Delta^\beta}{\alpha,\gamma} = x_{\alpha,\gamma}^\beta = x_{\alpha,\gamma}^\beta$.

We then have the following elementary lemma.

Lemma 1. Condition (M) is valid for fixed $\alpha \in (0, 1)$ if $G$ has on $(0, \alpha)$ a density $g = G'$ satisfying

$$\lim_{x \to 0} g(x) = g(0) = \infty \text{ and } t g(t) - G(t) < 0 \text{ on } (0, \alpha).$$

The latter is satisfied if, for example, $g'(t) < 0$ on $(0, \alpha)$.

Proof. From the assumptions it is clear that on $(0, \alpha)$

$$\frac{dH(t)}{dt} = \frac{(g(t) - 1)t - \int_0^t (g(s) - 1)ds}{t^2} < 0$$

$$\iff t g(t) - G(t) < 0.$$

Furthermore, since $\lim_{t \to 0} \frac{G(t)}{t} = g(0) = \infty$ we conclude that $\lim H(t) = \infty$. Thus the first part of the proof is settled. Clearly, $t g(t) - G(t) < 0$ if $g'(t) < 0$ on $(0, \alpha)$ and the overall proof is thus completed. \(\square\)
The next lemma ensures an easier expression for $K_G$, see (5) and (6), provided condition (M) holds.

**Lemma 2.** Let condition (M) hold for $G$ and $m_1 = m(1 - \gamma)$. Then we have that

$$K^M_G(x, m, 1 - \gamma, \epsilon, a) = K_G(x, m, m_1, 1 - \gamma, \epsilon, a) = 2 \min \left\{ m(1 - x) e^{-m(1 - \gamma) G(\alpha)c(\frac{x}{\gamma})}, e^{-2\frac{m}{\epsilon}(\alpha a x)^2} \right\}. \tag{19}$$

The same holds for the c.d.f. of the uniform distribution $F$. In particular if $\gamma > 1/2$ then

$$K^M_F(x, m, 1 - \gamma, \epsilon, a) = K_F(x, m, m_0, 1 - \gamma, \epsilon, a) = 2 \min \left\{ m(1 - x) e^{-m\gamma a x c(\frac{\epsilon}{\alpha})}, e^{-2\frac{m}{\epsilon}(\alpha a x)^2} \right\}. \tag{20}$$

**Remark 7.** Under condition (M) we have that $G(t) c \left( \frac{d}{G(t)} \right)$ is non-decreasing on $(0, a)$, for any $d > 0$, since $\frac{d}{G(t)}$ is non-decreasing on $(0, a)$ and $c$ is increasing on $(0, \infty)$, see Proposition A1. Hence the function $K^M_G(x, m, 1 - \gamma, \epsilon, a)$ is decreasing in $x$ as long as the other parameters stay fixed. The same is valid for $K^M_F$.

**Proof of Lemma 2.** The proof is immediate from (6) since under condition (M) we have that the function $G(t) c \left( \frac{d}{G(t)} \right)$ is non-decreasing on $(0, a)$, see Remark 7. We just note that by definition $m_1 = m(1 - \gamma)$ and we have employed this for $k$ in (6) to derive (19). Finally, (20) follows from (7). \(\square\)

2.4. **Theoretical Bounds on the Distribution of $R_m$ under Condition (M)**

If condition (M) holds we have the identity equivalent to (17)

$$\max_{t \in A_\alpha(x, m)} H(t) = \frac{G\left( a \frac{[\alpha x]}{m} \right) - a \frac{[\alpha x]}{m}}{a \frac{[\alpha x]}{m}} = H\left( a \frac{[\alpha x]}{m} \right) \tag{21}$$

and with $\beta = \alpha$ relation (18) together with (12) and (13) gives that

$$\frac{1}{\alpha} - 1 = (1 - \gamma) H(a x_{a, \gamma}) = (1 - \gamma) \left( \frac{G(a x_{a, \gamma}) - a x_{a, \gamma}}{a x_{a, \gamma}} \right). \tag{22}$$

Let $\beta > 0$ and denote using (21)

$$m_G(x, \beta, \alpha, \gamma) : = \inf \{ k \in \mathbb{N}^+ : \max_{m \geq k} \mathbb{I}\{ (1 - \gamma) \max_{x \in A_\alpha(x, m)} H(t) \geq \beta \} = 0 \} = \inf \{ k \in \mathbb{N}^+ : \max_{m \geq k} \mathbb{I}\{ (1 - \gamma) H(a \frac{[\alpha x]}{m}) \geq \beta \} = 0 \}. \tag{23}$$

We then have the following key result:

**Theorem 1.** Let condition (M) hold for given $G$ and fixed $\alpha \in (0, 1)$ with $G(\alpha) < 1$. Fix $\gamma \in (0, 1)$. Recall from (18) that $x_{a, \gamma} = \frac{a^\gamma}{\alpha^\gamma} = x_{x_{a, \gamma}}$. Then $\lim_{m \to \infty} \frac{R_m}{m} = x_{a, \gamma} \in (0, 1)$. Moreover, for any $x \in (x_{a, \gamma}, 1), \exists \varepsilon(x)$ such that for any $0 \leq \varepsilon \leq \varepsilon(x)$ the inequality $(1 - \gamma) H(ax) < \frac{1}{\alpha} - 1 - 2\varepsilon$ holds and then for any $m \geq m_G(x, \frac{1}{\alpha} - 1 - 2\varepsilon, \alpha, \gamma)$ we have that

$$\mathbb{P}\left( \frac{R_m}{m} \leq x \right) \geq 1 - K^M_F(x, m, \gamma, \epsilon, a) - K^M_G(x, m, 1 - \gamma, \epsilon, a). \tag{24}$$
Finally, for any \( x \in (0, x_{a, \gamma}) \), \( \exists \varepsilon(x) \) such that for any \( 0 \leq \varepsilon \leq \varepsilon(x) \) the inequality 
\[(1 - \gamma)H(ax) > \frac{1}{\alpha} - 1 + 2\varepsilon \]
holds and then for any \( m \geq m_G(x, \frac{1}{\alpha} - 1 + 2\varepsilon, a, \gamma) \)
\[P\left(\frac{R_m}{m} < x\right) \leq K_M^G(x, m, \gamma, \varepsilon, a) + K_M^G(x, m, 1 - \gamma, \varepsilon, a). \]  
(25)

**Remark 8.** It seems important to evaluate or approximate \( m_G(x, \frac{1}{\alpha} - 1 + 2\varepsilon, a, \gamma) \) with say \( m' \) so that (24) is valid beyond this estimate, that is for \( m \geq m' \). If \( H_d := \inf_{t \in (0, a)} |H'(t)| > 0 \), since \( ax < a \left\lfloor \frac{mx}{m} \right\rfloor 1 > x > x_{a, \gamma} \) and \( H \) is decreasing on \((0, a)\) we get using 22 in the first relation
\[\frac{1}{\alpha} - 1 = (1 - \gamma)H(ax_{a, \gamma}) = (1 - \gamma)(H(ax_{a, \gamma}) - H(ax)) + (1 - \gamma)H(ax) \geq (1 - \gamma)H_d a(x - x_{a, \gamma}) + (1 - \gamma)H(ax). \]

However, if \((1 - \gamma)H_d a(x - x_{a, \gamma}) > 2\varepsilon\) we get from the last observation that
\[ (1 - \gamma)H\left(a \left\lfloor \frac{mx}{m} \right\rfloor \right) \leq (1 - \gamma)H(ax) \leq \frac{1}{\alpha} - 1 - (1 - \gamma)H_d a(x - x_{a, \gamma}) < \frac{1}{\alpha} - 1 - 2\varepsilon = \frac{1}{\alpha} \varepsilon \]
for any \( m \geq 1 \) and therefore if \((1 - \gamma)H_d a(x - x_{a, \gamma}) > 2\varepsilon\) then
\[ m_G(x, \frac{1}{\alpha} - 1 - 2\varepsilon, a, \gamma) = \inf \{k \in \mathbb{N} : \max_{m \geq k} \mathbb{I}_{\{(1 - \gamma)H\left(a \left\lfloor \frac{mx}{m} \right\rfloor \geq \frac{1}{\alpha} - 1 - 2\varepsilon\} = 0\} = 1. \]

and (24) is always valid. Of course various similar and more precise estimates can be achieved, but these are beyond the scope of this work.

**Remark 9.** Considering Remark 3 the inequalities (24) and (25) can be improved for computational purposes by using as an upper bound the expression
\[ 2 \min \left\{ \sum_{t \in A_0(m, \varepsilon)} e^{-kT(t)c\left(\frac{t}{m}\right)}e^{-2\frac{1}{\rho}(ax)^2} \right\} \]
for \( K_M^G, T = G, F \) with \( k = m_1, m_0, \rho = 1 - \gamma, \gamma \) respectively.

**Remark 10.** We point out that in our setting \( \lim_{m \to \infty} \frac{R_m}{m} > 0 \) and there have been a number of works that attempt to offer FDR even under weak dependence structure between the random variables following the uniform distribution or \((X_i)_{i \leq m}\). We point out, for example, to [19]. There is an interesting contribution, see [20], which discusses difficulties in controlling FDR under weak dependence in the case when \( \lim_{m \to \infty} \frac{R_m}{m} = 0. \)

**Proof of Theorem 1.** Under condition (M) we have that \( \sum_{a, \gamma} x = x_{a, \gamma} := x_{a, \gamma}, \) see (18). Clearly, \( x_{a, \gamma} \in (0, 1) \) since \( u_{a, \gamma}(0) = \infty \) and \( u_{a, \gamma}(1) = \frac{G(a) - a}{\alpha} < \frac{1}{\alpha} - 1, \) and by assumption \( G(a) < 1. \) The fact that \( \lim_{m \to \infty} \frac{R_m}{m} = x_{a, \gamma} \in (0, 1) \) follows from Theorem 3.1 in [15]. Henceforth, it remains to show (24) and (25). The former thanks to (16) and (19) follows for any \( m \geq m_G(x, \frac{1}{\alpha} - 1 - 2\varepsilon, a, \gamma) \), if \( m_G(x, \frac{1}{\alpha} - 1 - 2\varepsilon, a, \gamma) < \infty, \) since then
\[ \mathbb{I}_{\{(1 - \gamma)\max_{t \in A_0(x, m)} \frac{G(t) - 1}{\alpha} \geq \frac{1}{\alpha} - 1 - 2\varepsilon\}} = \mathbb{I}_{\{(1 - \gamma)H\left(a \left\lfloor \frac{mx}{m} \right\rfloor \geq \frac{1}{\alpha} - 1 - 2\varepsilon\} = 0. \]
However, since \( \lim_{m \to \infty} a \left\lfloor \frac{mx}{m} \right\rfloor = ax, x > x_{a, \gamma} \) and \( H \) decreasing on \((0, a)\), then
\[
\lim_{m \to \infty} (1 - \gamma) H \left( a \frac{mx}{m} \right) = (1 - \gamma) H(ax) < (1 - \gamma)H(ax_{a, \gamma}) = \frac{1}{a} - 1.
\]

Thus, for all \( \epsilon \) small enough, we conclude that \((1 - \gamma)H(ax) < \frac{1}{a} - 1 - 2\epsilon \) and thus \( m \log G(\frac{1}{a} - 1 - 2\epsilon, a, \gamma) < \infty \), see (23). Hence, (24) follows.

To prove (25) for any \( x \in (0, x_{a, \gamma}) \) we work with the complement event, that is \( \{ \frac{R_m}{m} \geq x \} \). Similarly as in (16) and utilizing (4) and (21) we get that on the event
\[
\left\{ \max_{t \in A_{\infty}(x, m)} \left| \frac{F_{m_0}(t) - t}{t} \right| = \frac{\epsilon}{\gamma} ; \max_{t \in A_{\infty}(x, m)} \left| G_{m_1}(t) - G(t) \right| \leq \frac{\epsilon}{1 - \gamma} \right\}
\]
we have that \( \max_{t \in A_{\infty}(x, m)} \frac{H_{m_0}(t) - t}{t} \geq (1 - \gamma)H \left( a \frac{mx}{m} \right) - 2\epsilon \) and therefore
\[
\mathbb{P}\left( \frac{R_m}{m} \geq x \right) \geq \mathbb{P}\left( \max_{t \in A_{\infty}(x, m)} \left| \frac{F_{m_0}(t) - t}{t} \right| \leq \frac{\epsilon}{\gamma} ; \max_{t \in A_{\infty}(x, m)} \left| G_{m_1}(t) - G(t) \right| \leq \frac{\epsilon}{1 - \gamma} \right) \times \mathbb{I}\left\{ (1-\gamma)H \left( a \frac{mx}{m} \right) \geq \frac{1}{a} - 1 + 2\epsilon \right\}.
\]

Again as \( \frac{[mx]}{m} \to x < x_{a, \gamma} \) and \( H(ax) > H(ax_{a, \gamma}) \) we conclude that for all \( \epsilon \) small enough the inequality \((1 - \gamma)H(ax) \geq \frac{1}{a} - 1 + 2\epsilon \) holds and we have that
\[
\lim_{m \to \infty} \mathbb{I}\left\{ (1-\gamma)H \left( a \frac{mx}{m} \right) \geq \frac{1}{a} - 1 + 2\epsilon \right\} = 1.
\]

From Proposition 1, the expressions for \( K_{F}^{m}, K_{G}^{m} \) in Lemma 2 and the independence of \((X_j)_{j \leq m_0}\) and \((X_j)_{m \geq j > m_0}\) we have that
\[
\mathbb{P}\left( \max_{t \in A_{\infty}(x, m)} \left| \frac{F_{m_0}(t) - t}{t} \right| \leq \frac{\epsilon}{\gamma} ; \max_{t \in A_{\infty}(x, m)} \left| G_{m_1}(t) - G(t) \right| \leq \frac{\epsilon}{1 - \gamma} \right) = \mathbb{P}\left( \max_{t \in A_{\infty}(x, m)} \left| \frac{F_{m_0}(t) - t}{t} \right| \leq \frac{\epsilon}{\gamma} \right) \mathbb{P}\left( \max_{t \in A_{\infty}(x, m)} \left| G_{m_1}(t) - G(t) \right| \leq \frac{\epsilon}{1 - \gamma} \right) \geq \left( 1 - K_{F}^{m}(x, m, \gamma, \epsilon, a) \right) \left( 1 - K_{G}^{m}(x, m, 1 - \gamma, \epsilon, a) \right),
\]
and we conclude the proof of the theorem via the now trivial relations
\[
\mathbb{P}\left( \frac{R_m}{m} < x \right) = 1 - \mathbb{P}\left( \frac{R_m}{m} \geq x \right) \leq 1 - \mathbb{P}\left( \frac{R_m}{m} \geq x ; \max_{t \in A_{\infty}(x, m)} \left| \frac{F_{m_0}(t) - t}{t} \right| \leq \frac{\epsilon}{\gamma} ; \max_{t \in A_{\infty}(x, m)} \left| G_{m_1}(t) - G(t) \right| \leq \frac{\epsilon}{1 - \gamma} \right) \leq 1 - \left( 1 - K_{F}^{m}(x, m, \gamma, \epsilon, a) \right) \left( 1 - K_{G}^{m}(x, m, 1 - \gamma, \epsilon, a) \right) \times \mathbb{I}\left\{ (1-\gamma)H \left( a \frac{mx}{m} \right) \geq \frac{1}{a} - 1 + 2\epsilon \right\}
\]
and an appeal to (26). \( \square \)

For practical reasons and especially for the accompanying software routine it is optimal to choose \( \epsilon \) in Theorem 1 as large as possible. The latter is due to fact that the bounds for \( K_{F}^{m}, K_{G}^{m} \), see (19) and (20), are decreasing in \( \epsilon \). Since in practical computations \( m \) is usually
available, it is then beneficial to choose the largest $\epsilon$ as a function in $x, m$ in a way that the previous theorem holds. In this direction we have the following result:

**Theorem 2.** Under the conditions of Theorem 1 and $m \in \mathbb{N}^+$, $x \in (x_{a,\gamma}, 1)$ fixed we have that

$$\mathbb{P}\left( \frac{R_m}{m} \leq x \right) \geq 1 - K^M_f(x, m, \gamma, \epsilon_G, a) - K^M_G(x, m, 1 - \gamma, \epsilon_G, a),$$  

(28)

where

$$\epsilon_G = \epsilon(x, m) = \frac{1}{2} \left( \frac{1}{\alpha} - 1 \right) - \frac{1 - \gamma}{2} \frac{H\left( \frac{m x}{m} \right)}{\frac{m x}{m} \left( \frac{m x}{m} \right)} = \frac{1 - \gamma}{2} \left( H(ax_{a,\gamma}) - H\left( \frac{m x}{m} \right) \right).$$  

(29)

Similarly, if $m \in \mathbb{N}, x \in (0, x_{a,\gamma})$ are fixed and $\frac{m x}{m} < x_{a,\gamma}$ then

$$\mathbb{P}\left( \frac{R_m}{m} < x \right) \leq K^M_f(x, m, \gamma, \epsilon_G, a) + K^M_G(x, m, 1 - \gamma, \epsilon_G, a),$$  

(30)

where

$$\epsilon_G = \epsilon(x, m) = \frac{1 - \gamma}{2} \frac{H\left( \frac{m x}{m} \right)}{\frac{m x}{m} \left( \frac{m x}{m} \right)} = \frac{1 - \gamma}{2} \left( H\left( \frac{m x}{m} \right) - H(ax_{a,\gamma}) \right).$$  

(31)

**Proof.** To show (28) with $\epsilon_G$ as in (29) it suffices to show that for the chosen $\epsilon_G$ and the given $m, x$ we have that

$$\mathbb{I}\left\{ (1 - \gamma) \max_{x \in A_{x}(x,m)} \frac{G(t) - t}{t} \geq \frac{1}{\alpha} - 1 - 2\epsilon \right\} = 0$$

in (16). However, as condition (M) holds, then

$$\max_{t \in A_{x}(x,m)} \frac{G(t) - t}{t} = \max_{t \in A_{x}(x,m)} H(t) = H\left( \frac{m x}{m} \right)$$

and we see that $\epsilon_G$ is the maximal possible choice such that

$$\mathbb{I}\left\{ (1 - \gamma) H\left( \frac{m x}{m} \right) \geq \frac{1}{\alpha} - 1 - 2\epsilon \right\} = 0.$$

For the second part of (29) we only recall that $(1 - \gamma) H(ax_{a,\gamma}) = \frac{1}{\alpha} - 1$, see (22). The rest of the theorem, that is (30) and (31), follows in a similar fashion from (27) wherein the expression of (31) is the maximal $\epsilon$ such that for the fixed $m, x \in (0, x_{a,\gamma})$

$$\mathbb{I}\left\{ (1 - \gamma) H\left( \frac{m x}{m} \right) \geq \frac{1}{\alpha} - 1 + 2\epsilon \right\} = 1.$$

We only note that in this instance we need to have that $x < \frac{m x}{m} < x_{a,\gamma}$, which is part of the assumptions of the theorem. □

2.5. **Exponential Convergence to the Theoretical Limit of the Power of the Benjamini–Hochberg Procedure**

We proceed by studying the question about the power of the test. It is clear that if we denote by

$$S_m = \sum_{j=1}^{m_1} \mathbb{I}\{ X_j \leq \alpha \frac{j m}{m} \}.$$  

(32)
where \( (X_i)_{1 \leq i \leq m_1} \) are i.i.d. with c.d.f \( G \) then

\[
S^*_m = \frac{S_m}{m_1}
\]  

(33)

is the proportion of tested elements that we have correctly identified as significant or the power of the test. If one wishes to understand the behavior of \( S^*_m \) with the increase of \( m \) one ought to decouple its dependence on \( R_m/m \) in (32). Theorem 1 aids in this matter. We note that if we choose \( x_1 < \lim_{m \to \infty} \frac{R_m}{m} = x_{a,\gamma} < x_2 \), see (18) for the definition of \( x_{a,\gamma} \), then from Theorem 1 we have for every

\[
m \geq \max \left\{ m_G(x_1, \frac{1}{\alpha} - 1 - 2\epsilon, a, \gamma); m_G(x_2, \frac{1}{\alpha} - 1 + 2\epsilon, a, \gamma) \right\} := m^*(x_1, x_2, \epsilon, a, \gamma)
\]

with \( \epsilon \leq \min\{\epsilon(x_1), \epsilon(x_2)\} \) that

\[
\mathbb{P} \left( \frac{R_m}{m} \in [x_1, x_2] \right) = \mathbb{P} \left( \frac{R_m}{m} \leq x_2 \right) - \mathbb{P} \left( \frac{R_m}{m} < x_1 \right)
\geq 1 - \left( K_M^G(x_2, m, \gamma, \epsilon, a) + K_M^G(x_2, m, 1 - \gamma, \epsilon, a) \right)
\]

(34)

and any \( \eta \).

Then we can formulate the following fundamental result that offers exponential bounds on the convergence of \( S^*_m \).

**Theorem 3.** Let condition (M) be valid and assume the notation of Theorem 1. Let next \( x_{a,\gamma} \in (x_1, x_2) \subset (0, 1) \). Then there exists \( m^* = m^*(x_1, x_2, \epsilon, a, \gamma) \) and \( \epsilon^* = \epsilon(x_1, x_2) \) such that for any \( m \geq m^* \) and any \( \epsilon \leq \epsilon^* \)

\[
\mathbb{P} \left( \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbb{I}\{X_j \leq \alpha x_1\} \leq S^*_m \leq \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbb{I}\{X_j \leq \alpha x_2\} \right)
\geq 1 - \left( K_M^G(x_3, m, \gamma, \epsilon, a) + K_M^G(x_3, m, 1 - \gamma, \epsilon, a) \right)
\]

(35)

Clearly, the random variables \( \sum_{j=1}^{m_1} \mathbb{I}\{X_j \leq \alpha x\} \) are Binomial with parameters \( m_1 \) and \( G(ax) = \mathbb{P}(X_1 \leq ax) \). Henceforth,

\[
\lim_{m \to \infty} S^*_m = G(ax_{a,\gamma}) = x_{a,\gamma} \frac{1 - \alpha \gamma}{1 - \gamma}.
\]  

(36)

Moreover, for any \( \eta \in (0, 1) \) with \( m, \epsilon \) as above, we have that

\[
\mathbb{P}(1 - \eta)G(ax_1) \leq S^*_m \leq (1 + \eta)G(ax_2))
\geq 1 - \left( K_M^G(x_3, m, \gamma, \epsilon, a) + K_M^G(x_3, m, 1 - \gamma, \epsilon, a) \right)
\]

(37)

\[
- c((1 + \eta) \ln(1 + \eta) - \eta, \eta > 0, \text{ is considered in Proposition A1.})
\]

**Remark 11.** We note that (36) has been proved in Corollary 3.3 in [15], but here we offer some exponential bounds on this convergence.
Proof of Theorem 3. The proof of (35) follows immediately from (32), (34) and the considerations thereafter about
\[
P\left(\frac{1}{m_1} \sum_{j=1}^{m_1} I\{x_j \leq \alpha x_1\} \leq S^*_m \leq \frac{1}{m_1} \sum_{j=1}^{m_1} I\{x_j \leq \alpha x_2\}\right) \geq P\left(\frac{R_m}{m} \in [x_1, x_2]\right). \tag{38}
\]

The fact that \( \sum_{j=1}^{m_1} I\{x_j \leq \alpha x\} \) is Binomial with parameters \( m_1 \) and \( G(\alpha x) = P(X_1 \leq \alpha x) \) is straightforward which with the help of (35) and
\[
\lim_{m \to \infty} \left( K^M_{\gamma}(x_1, m, \gamma, \epsilon, \alpha) + K^M_{\gamma}(x_1, m, 1 - \gamma, \epsilon, \alpha) \right) + \lim_{m \to \infty} \left( K^M_{\epsilon}(x_2, m, \gamma, \epsilon, \alpha) + K^M_{\epsilon}(x_2, m, 1 - \gamma, \epsilon, \alpha) \right) = 0
\]
prove (36). Indeed, the strong law of large numbers gives almost surely that
\[
\lim_{m \to \infty} \frac{1}{m_1} \sum_{j=1}^{m_1} I\{x_j \leq \alpha x\} = G(\alpha x)
\]
and from (35) \( \lim_{m \to \infty} S^*_m = G(\alpha x, \gamma) \). Next, from (22) we get that
\[
H(\alpha x, \gamma) = \frac{G(\alpha x, \gamma) - \alpha x, \gamma}{\alpha x, \gamma} = 1 - \frac{\alpha}{\alpha}
\]
and henceforth \( G(\alpha x, \gamma) = x, \gamma \left( 1 - \frac{\alpha}{\alpha} \right) \). The latter is the limit on the right-hand side of (36). The very last relation, that is (37), follows again from (35) and an application of Proposition A1 which yields in our case with \( m_1 = m(1 - \gamma) \) the bounds
\[
P\left(\frac{1}{m_1} \sum_{j=1}^{m_1} I\{x_j \leq \alpha x_2\} \geq (1 + \eta)G(\alpha x_2)\right) \leq e^{-c(\eta)(1-\gamma)mG(\alpha x_2)}
\]
and
\[
P\left(\frac{1}{m_1} \sum_{j=1}^{m_1} I\{x_j \leq \alpha x_1\} \leq (1 - \eta)G(\alpha x_1)\right) \leq e^{-c(\eta)(1-\gamma)mG(\alpha x_1)}.
\]

This concludes the proof. \( \Box \)

As in the case of \( \frac{R_m}{m} \) we state a more practical version of Theorem 3 which is aimed to compute the largest \( \epsilon \) for fixed \( m, x_1, x_2 \). It is in the spirit of Theorem 2.

Theorem 4. Let condition (M) be valid and assume the notation of Theorem 2. Let \( m \in \mathbb{N}^+ \) be fixed. Let next \( x, \gamma \in (x_1, x_2) \subset (0, 1) \) and \( x_1 \) is such that \( \frac{m x_1}{m} \leq x, \gamma \). Finally, let \( \eta \in (0, 1) \) be fixed. Then with
\[
eg\epsilon_1 = \epsilon(x_1, m) = \frac{1}{2} \left( 1 - \frac{\gamma}{\alpha} \right) - \frac{1 - \gamma}{2} H\left( \alpha \left( \frac{m x_1}{m} \right) \right) = \frac{1 - \gamma}{2} \left( H(\alpha x, \gamma) - H\left( \alpha \left( \frac{m x_1}{m} \right) \right) \right), \tag{39}
\]
and
\[
eg\epsilon_2 = \epsilon(x_2, m) = \frac{1 - \gamma}{2} H\left( \alpha \left( \frac{m x_2}{m} \right) \right) - \frac{1}{2} \left( 1 - \frac{\gamma}{\alpha} \right) = \frac{1 - \gamma}{2} \left( H\left( \alpha \left( \frac{m x_2}{m} \right) \right) - H(\alpha x, \gamma) \right). \tag{40}
\]
we have that
\[
\Pr((1-\eta)G(ax_1) \leq S_m \leq (1+\eta)G(ax_2)) \\
\geq \max\{1 - \left(K_M(x_1, m, \gamma, \epsilon_1, a) + K_M^*(x_1, m, 1-\gamma, \epsilon_1, a)\right) \\
- \left(K_M(x_2, m, \gamma, \epsilon_2, a) + K_M^*(x_2, m, 1-\gamma, \epsilon_2, a)\right) \\
e^{-mc(\eta)(1-\gamma)G(ax_1)} - e^{-mc(\eta)(1-\gamma)G(ax_2)} ; 0\} \tag{41}
\]
and \(c(\eta) = (1+\eta)\ln(1+\eta) - \eta\) is considered in Proposition A1. For a one-sided bound we get the inequality
\[
\Pr((1-\eta)G(ax_1) \leq S_m) \\
\geq \max\{1 - K_M(x_1, m, \gamma, \epsilon_1, a) - K_M^*(x_1, m, 1-\gamma, \epsilon_1, a) - e^{-mc(\eta)(1-\gamma)G(ax_1)} ; 0\}. \tag{42}
\]

Remark 12. Noting Remark 9 we see that the expressions involving \(K^F_M, K^G_M\) can be substituted with
\[
2 \min\left\{ \sum_{t \in A(m,x)} e^{-kT(t)\epsilon\left(\frac{1}{T}\right)} ; e^{-\frac{2}{m}(ax)^2}\right\},
\]
where \(T\) is either \(F\) or \(G\).

Proof of Theorem 4. The proof is immediate but we note that the assumption \(\left[ \frac{m\alpha}{m}\right] < x_{\alpha,\gamma}\) and the choice of \(\epsilon_1, \epsilon_2\) are such that Theorem 2 applies to the right-hand side of (38), which yields an inequality identical to (35). The rest is as in the proof of Theorem 3.

3. Workflow
In this section we consider the case when the non-uniform distribution of \((X_i)_{m_0 < j \leq m}\) is Beta with parameters \(a, b\), that is \(G(\cdot, a, b)\) is the cdf of \(B(a, b)\), where \(a < 1\) and \(b > 1\) are taken so that condition (M) is satisfied. In this case Lemma 1 is valid and condition (M) holds since elementary calculations yield that \(g'_{a,b} < 0\) on \((0, a)\) and \(\lim_{x \to a} g_{a,b}(x) = \infty\), where \(g_{a,b}\) is the probability density function of \(B(a, b)\).

For a given number of hypotheses \(m\), level of significance \(\alpha\) and proportion of significant hypotheses \(\gamma\), we are interested in finding a theoretical bound for the probability that the power is between a number \(l\) and 1, where of course only \(l < G(ax_{\alpha,\gamma}, a, b)\) makes sense to be considered. We shall only provide a step by step description of the implementation and we shall refrain from developing a particular, explicit theoretical example for given values of \(a, b\) since the latter involves tedious, non-instructive and lengthy computations. We proceed with describing the workflow that has been implemented numerically in R:

1. we calculate \(x_{\alpha,\gamma}\) by solving numerically Equation (22) wherein \(G(\cdot) = G(\cdot, a, b)\);
2. using the value of \(x_{\alpha,\gamma}\), we calculate the power limit of \(S_m\), see (33), i.e., \(G(ax_{\alpha,\gamma}, a, b)\), according to (36);
3. for given \(l < G(ax_{\alpha,\gamma}, a, b)\), any given \(\eta\) and \(x_1\) is such that \(\left[ \frac{mx_1}{m}\right] < x_{\alpha,\gamma}\) we have an estimate based on (42), that is
\[
\Pr((1-\eta)G(ax_1) \leq S_m) \\
\geq \max\{1 - K_M(x_1, m, \gamma, \epsilon_1, a) - K_M^*(x_1, m, 1-\gamma, \epsilon_1, a) - e^{-mc(\eta)(1-\gamma)G(ax_1)} ; 0\};
\]
4. fixing \(\eta\) we attempt to solve (numerically) in \(x_1\) the equation
\[
l = (1-\eta)G(ax_1, a, b);
\]
5. if \( \left\lfloor \frac{m x_1}{m} \right\rfloor < x_{\alpha, \gamma} \) then (42) represents a valid theoretical lower bound, that is

\[
p(\eta) = \max \left\{ 1 - K_F^{M}(x_1, m, \gamma, \epsilon_1, \alpha) - K_G^{M}(x_1, m, 1 - \gamma, \epsilon_1, \alpha) - e^{-mc(\eta)(1-\gamma)G(\alpha x_1, \epsilon_1, \alpha)} \right\}
\]

with \( P(l \leq S_m^*) \geq p(\eta) \);

6. finally, we optimize \( p(\eta) \) in \( \eta \) and choose \( \eta^* \) so that the estimated probability bound is the largest, that is \( P(l \leq S_m^*) \geq \sup_{\eta \in (0,1)} p(\eta) = p(\eta^*) \).

4. Results

The following Tables 1–4 give empirical results for some values of \( m, a, b, \gamma \). We have included different cases for the real power limit—taking large, moderate and small values. In all cases the significance level is \( \alpha = 0.05 \). These tables also show the empirical probability based on numerical simulations with 1 million repetitions. Empirical probability of 1 means that no simulation repetition had power lower than \( l \). Theoretical probability of 0 or empty value of \( \eta^* \) indicate that some of the conditions of the method were not met.

Table 1. Simulation results for \( B(0.1, 100) \), number genes = 20,000, proportion significant = 0.1, real power limit = 0.940359.

| Lower Bound \( l \) | Empirical Probability | Theoretical Estimate \( p(\eta^*) \) | \( \eta^* \) |
|---------------------|-----------------------|-----------------------------|------------|
| 0.750               | 1.000                 | 1.000                       | 0.098      |
| 0.760               | 1.000                 | 1.000                       | 0.097      |
| 0.770               | 1.000                 | 1.000                       | 0.096      |
| 0.780               | 1.000                 | 1.000                       | 0.096      |
| 0.790               | 1.000                 | 1.000                       | 0.095      |
| 0.800               | 1.000                 | 1.000                       | 0.094      |
| 0.810               | 1.000                 | 1.000                       | 0.094      |
| 0.820               | 1.000                 | 1.000                       | 0.093      |
| 0.830               | 1.000                 | 1.000                       | 0.093      |
| 0.840               | 1.000                 | 0.999                       | 0.085      |
| 0.850               | 1.000                 | 0.996                       | 0.078      |
| 0.860               | 1.000                 | 0.985                       | 0.068      |
| 0.870               | 1.000                 | 0.952                       | 0.058      |
| 0.880               | 1.000                 | 0.873                       | 0.048      |
| 0.890               | 1.000                 | 0.714                       | 0.037      |
| 0.900               | 1.000                 | 0.476                       | 0.027      |
| 0.910               | 1.000                 | 0.204                       | 0.016      |
| 0.920               | 1.000                 | 0.018                       | 0.005      |
| 0.930               | 0.967                 | 0.000                       | 0.000      |

Larger sample size would result in a smaller parameter \( a \) of the Beta distribution, thus squeezing the distribution towards 0 and increasing the statistical power and its theoretical estimates, e.g., as in Tables 2 and 3.

We have carried out numerical simulations for many values of the parameters and based on the results it appears that the theoretical bound works well when \( l \) is not close to the real power limit \( lG(\alpha x_1, \gamma, a, b) \). Link to our R code is provided in the Supplementary Materials section.

Interestingly, when \( m \) increases and all other parameters are kept fixed, the expected power decreases for most values of \( m \). This is shown in the following Figure 1. Perhaps this is due to the fact that the set for which we take the maximum in (2) increases its size for most values of \( m \) (unless the ceiling function jumps).
We note that as expected when either $a$ is closer to 1 or/and $\gamma$ is smaller then the theoretical estimates are getting significantly worse. This is natural as the closer the alternative distribution gets to the uniform the harder is to distinguish that the respective random variable $X$ originates from it. The same is valid if the lower bound $l$ is very close to the true limit and it is harder to account for the possible error. All these points reflect natural restrictions when one deals with universal, theoretical bounds, which are seemingly conducted in an optimal way.

Table 2. Simulation results for $B(0.25, 2)$, number genes = 20,000, proportion significant = 0.1, real power limit = 0.233737.

| Lower Bound $l$ | Empirical Probability | Theoretical Estimate $p(\eta^*)$ | $\eta^*$ |
|-----------------|-----------------------|-----------------------------------|---------|
| 0.190           | 1.000                 | 0.779                             | 0.105   |
| 0.200           | 0.997                 | 0.434                             | 0.072   |
| 0.210           | 0.972                 | 0.000                             | 0.001   |
| 0.220           | 0.868                 | 0.000                             | 0.001   |
| 0.230           | 0.626                 | 0.000                             | 0.001   |

Table 3. Simulation results for $B(0.1, 2)$, number genes = 20,000, proportion significant = 0.1, real power limit = 0.620014.

| Lower Bound $l$ | Empirical Probability | Theoretical Estimate $p(\eta^*)$ | $\eta^*$ |
|-----------------|-----------------------|-----------------------------------|---------|
| 0.500           | 1.000                 | 1.000                             | 0.119   |
| 0.510           | 1.000                 | 1.000                             | 0.117   |
| 0.520           | 1.000                 | 1.000                             | 0.116   |
| 0.530           | 1.000                 | 1.000                             | 0.115   |
| 0.540           | 1.000                 | 0.999                             | 0.106   |
| 0.550           | 1.000                 | 0.996                             | 0.097   |
| 0.560           | 1.000                 | 0.979                             | 0.081   |
| 0.570           | 1.000                 | 0.926                             | 0.067   |
| 0.580           | 1.000                 | 0.786                             | 0.052   |
| 0.590           | 0.994                 | 0.521                             | 0.036   |
| 0.600           | 0.953                 | 0.189                             | 0.02    |
| 0.610           | 0.803                 | 0.000                             | 0.001   |
| 0.620           | 0.511                 | 0.000                             |         |

Table 4. Simulation results for $B(0.1, 10)$, number genes = 20,000, proportion significant = 0.1, real power limit = 0.755055.

| Lower Bound $l$ | Empirical Probability | Theoretical Estimate $p(\eta^*)$ | $\eta^*$ |
|-----------------|-----------------------|-----------------------------------|---------|
| 0.600           | 1.000                 | 1.000                             | 0.109   |
| 0.610           | 1.000                 | 1.000                             | 0.108   |
| 0.620           | 1.000                 | 1.000                             | 0.107   |
| 0.630           | 1.000                 | 1.000                             | 0.106   |
| 0.640           | 1.000                 | 1.000                             | 0.105   |
| 0.650           | 1.000                 | 1.000                             | 0.104   |
| 0.660           | 1.000                 | 0.999                             | 0.096   |
| 0.670           | 1.000                 | 0.995                             | 0.085   |
| 0.680           | 1.000                 | 0.980                             | 0.074   |
| 0.690           | 1.000                 | 0.934                             | 0.062   |
| 0.700           | 1.000                 | 0.821                             | 0.049   |
| 0.720           | 0.999                 | 0.608                             | 0.037   |
| 0.730           | 0.991                 | 0.309                             | 0.024   |
| 0.740           | 0.924                 | 0.048                             | 0.009   |
| 0.750           | 0.693                 | 0.000                             | 0.001   |
5. Discussion

5.1. Related Work

There are a number of results concerning the statistical power of the Benjamini–Hochberg procedure that relate to our result. Paper [15] discusses different properties of the procedure and shows that when increasing the number of tests \( m \), while keeping the underlying distributions and the proportion of true significant tests the same, the statistical power under some conditions converges to a limit that can be found as a solution of a particular equation, see (3.8) of Corollary 3.3 of that article. A step forward is made in [13] wherein it is shown that the proportions related to the power of the test obey even a law of the iterated logarithm. Other notable contributions amongst others include the papers [18] which deal with various properties of the Benjamini–Hochberg procedure, including a Donsker-type of convergence. They are discussed in more detail in Remark 5. We wish to highlight explicitly that to our knowledge there are no bounds, albeit conservative, for the probability to uncover a given proportion of the truly significant tests (that is \( S_n/n \)) with respect to the limit of \( S_n/n \), as \( n \) goes to infinity. We achieve this by evaluating the speed of convergence, whereas the results under more general assumptions, including under weak dependence between the tests, consider limit results. This does not allow for the derivation of any concrete intervals, depending solely on the set of parameters, which give a certain probability for the proportion of correctly identified significant tests.

While we argue below that for genetic studies the assumption of independence is not binding at all, we wish to mention some achievements under the weak dependence assumption since it is natural to pursue as a future development the extension of our results to this scenario. The limit for the power of the Benjamini–Hochberg procedure is proved in [15]. An interesting contribution is [12] which shows that for a large number of tests the FDR is controlled even when there is some degree of dependence between the test statistics. A conservative estimator for the proportion of false nulls among the hypotheses (the quantity \( 1 - \gamma \) in our work) is found (the proof there uses the Dvoretsky–Kiefer–Wolfowitz inequality [16], as we do for different purposes in our work). Then the estimator is used in
a plug-in procedure in order to boost the power of the procedure while keeping the FDR controlled. An estimator for $\gamma$ may precede the application of our procedure.

We are not interested in improving the procedure itself, but rather in estimating the power of the original procedure, that as we mentioned in the Introduction section, is the most widely used among the similar procedures.

As mentioned in [20], the results in [12] assume that $R_n/n$ is asymptotically larger than zero, where $R_n$ is the number of rejected tests out of all $n$ tests (the $R$ notation is consistent with ours through this work). A counterexample is given of a case in which $R_n/n$ converges asymptotically to zero with positive probability and in which the FDR is larger than $\alpha$. Such situation does not appear to be relevant to the RNA-seq data used for single-step differential expression methods that prompted our work. Moreover, paper [12] is concerned with controlling the FDR rather than estimates of the power, which is the scope of our work in which we are interested in the power of the procedure that the scientists predominantly use, regardless of whether the FDR is controlled.

5.2. Considerations about Our Work

Because this is the first result of this exact type that we are aware of, we start our considerations under independence assumption. We want to point out that, as mentioned above, there are more general results under convergence, but they do not include explicit bounds for given set of parameters. The independence assumption holds approximately in the case of RNA-seq expression level analysis, where there are groups of genes that are correlated within the group, but most pairs of genes are weakly or not correlated. Even in the case of weak dependence, having real-world RNA-seq expression data, one cannot reliably estimate or replicate the dependence unless the sample size is much larger than the one of the currently available datasets. In addition, more complicated assumptions for the structure of dependence would make the numerical procedure impossible to set up as there would be too many parameters in the input and these parameters cannot be easily estimated for small sample sizes that are typical for RNA-seq experiments.

Theoretical results under more complicated assumptions for the dependence could be a subject of a follow-up studies. Numerical simulations of RNA-seq data under different basic types of weak dependencies could also be considered in subsequent studies.

It is important to also emphasize that our considerations differ from the basic situations in which the statistical power is studied. In a case of a simple test the statistical power is typically considered as a function of the sample size. However, in a case of a complex tests such as the ones cited above, often it is not possible to estimate the distribution of the test statistic, even for a single test in the multiple comparison setup. In order to be able to model a general setup we model the test statistic under the alternative hypothesis using the Beta family of distributions. In order to use our current results, one would first have to approximately or exactly estimate the parameters of the Beta distribution as function of the sample size of the particular test. Because such estimates are test-specific and cannot be studied in general, this would be a subject of another article. One would also have to estimate the proportion of significant tests and, having achieved that, one would apply our results for that proportion and those parameters of the Beta distribution, taking into account the possible values of the number of tests. For example, in the differential expression analysis methods, mentioned above, the sample size corresponds to the number of individuals and the number of tests may correspond to the number of genes.

Supplementary Materials: The R script and instructions are available at the following page: http://www.math.bas.bg/~palejev/BH.

Funding: D.P. was supported by the European Union Horizon 2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under Grant Agreement No. 857155 and Operational Programme Science and Education for Smart Growth under Grant Agreement No. BG05M2OP001-1.003-0002-C01. M.S. was partially supported by the financial funds allocated to the Sofia University “St. Kliment Ohridski”, grant No. 80-10-87/2021. The computational resources and infrastructure were provided
Acknowledgments: The extensive numerical simulations were done on the Avitohol supercomputer that is described in [21].

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

Abbreviations
The following abbreviations are used in this manuscript:
FDR False Discovery Rate
PET Positron Emission Tomography
CT Computerized Tomography
fMRI Functional Magnetic Resonance Imaging

Appendix A. Petrov’s Inequality
In this part we state and outline the proof of a celebrated inequality which can be found in the monograph [22].

Proposition A1. Let $X$ is binomially distributed with parameters $n, p$. Then we have that for any $\epsilon > 0$
\[ P(X - np > \epsilon np) \leq e^{-n p c(\epsilon)}, \quad P(np - X > \epsilon np) \leq e^{-n p c(\epsilon)} \] (A1)
where $c(\epsilon) = (1 + \epsilon) \ln(1 + \epsilon) - \epsilon, \epsilon : (0, \infty) \rightarrow (0, \infty)$ is an increasing function and the limit
\[ \lim_{\epsilon \rightarrow 0} 2c(\epsilon)/\epsilon^2 = 1 \] holds. Finally, $c(\epsilon) \leq \epsilon^2$ and $c(\epsilon) \geq \min\left(1, \frac{\epsilon^2}{2}\right)$ for any $\epsilon > 0$.

Proof. We consider
\[ P(X > (1 + \epsilon)np) \leq e^{-\lambda np(1+\epsilon)} \mathbb{E}\left[e^{\lambda X}\right] = e^{n\ln(1-p+pe^\lambda)-\lambda np(1+\epsilon)}, \]
where the latter is simply the Markov’s inequality with some $\lambda > 0$. Then using $\ln(1 + \epsilon) \leq \epsilon, \epsilon > 0$, we deduce that
\[ P(X > (1 + \epsilon)np) \leq e^{np(\epsilon^3 - \lambda(1+\epsilon))}. \]

The minimum of $f(\lambda) = e^\lambda - 1 - \lambda(1 + \epsilon)$ is attained at $\lambda = \ln(1 + \epsilon)$ and thus we conclude that
\[ P(X > (1 + \epsilon)np) \leq e^{-np(1+\epsilon)\ln(1+\epsilon) - \epsilon}. \]

Similar computation works for $P(X > (1 - \epsilon)np)$. Clearly, setting the function
\[ c(\epsilon) = (1 - (1 + \epsilon) \times \ln(1 + \epsilon)) \]
we arrive at the limit
\[ \lim_{\epsilon \rightarrow 0} 2\frac{(\epsilon - (1 + \epsilon) \ln(1 + \epsilon))}{\epsilon^2} = 1 \]
and $(1 + \epsilon) \ln(1 + \epsilon) - \epsilon$ is increasing in $\epsilon$. Using again $\ln(1 + \epsilon) \leq \epsilon, \epsilon > 0$, we also conclude that $c(\epsilon) \leq \epsilon^2$ for any $\epsilon > 0$. Considering the properties of the function $f(\epsilon) = c(\epsilon) - \frac{\epsilon^2}{2\epsilon}$ for $\epsilon \in (0, e - 1)$ and noting that $c(\epsilon) \geq 1$ for $\epsilon > e - 1$ we deduce that $c(\epsilon) \geq \min\left(1, \frac{\epsilon^2}{2}\right)$. □
References

1. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, 57, 289–300. [CrossRef]

2. Love, M.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, 15, 550. [CrossRef]

3. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, 26, 139–140. [CrossRef] [PubMed]

4. Trapnell, C.; Hendrickson, D.G.; Sauvageau, M.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2012**, 31, 46–53. [CrossRef]

5. Chalkidou, A.; O’Doherty, M.J.; Marsden, P.K. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS ONE* **2015**, 10, e0124165. [CrossRef]

6. Bennett, C.M.; Wolford, G.L.; Miller, M.B. The principled control of false positives in neuroimaging. *Soc. Cogn. Affect. Neurosci.* **2009**, 4, 417–422. [CrossRef]

7. Miller, C.J.; Genovese, C.; Nichol, R.C.; Wasserman, L.; Connolly, A.; Reichart, D.; Hopkins, A.; Schneider, J.; Moore, A. Controlling the False-Discovery Rate in Astrophysical Data Analysis. *Astron. J.* **2001**, 122, 3492. [CrossRef]

8. Ferreira, J.A.; Zwinderman, A.H. Approximate Power and Sample Size Calculations with the Benjamini-Hochberg Method. *Int. J. Biostat* **2006**, 2, 8. [CrossRef]

9. Ferreira, J.A.; Zwinderman, A.H. On the Benjamini—Hochberg method. *Ann. Statist.* **2006**, 34, 1827–1849. [CrossRef]

10. Neuvial, P. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2008**, 2, 1065–1110. [CrossRef]

11. Neuvial, P. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2008**, 2, 1065–1110. [CrossRef]

12. Petrov, V.V. *Sums of Independent Random Variables*; Brown, A.A., Translator; Ergebnisse der Mathematik und ihrer Grenzgebiete; Springer: Berlin/Heidelberg, Germany, 1975; Volume 82.

13. Chi, Z. On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **2007**, 35, 1409–1431. [CrossRef]

14. Ferreira, J.A.; Zwinderman, A.H. Approximate Power and Sample Size Calculations with the Benjamini-Hochberg Method. *Int. J. Biostat* **2006**, 2, 8. [CrossRef]

15. Dvoretzky, A.; Kiefer, J.; Wolfowitz, J. Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. *Ann. Math. Stat.* **1956**, 27, 642–669. [CrossRef]

16. Pounds, S.; Morris, S.W. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics* **2003**, 19, 1236–1242. [CrossRef] [PubMed]

17. Gontscharuk, V.; Finner, H. Asymptotic FDR control under weak dependence: A counterexample. *Stat. Probab. Lett.* **2013**, 83, 1888–1893. [CrossRef]

18. Atanassov, E.; Gurov, T.; Karaivanova, A.; Ivanovska, S.; Drouchova, M.; Dimitrov, D. On the parallelization approaches for Intel MIC architecture. *AIP Conf. Proc.* **2016**, 1773, 070001. [CrossRef]