Journal of Education and Culture Studies
ISSN 2573-0401 (Print) ISSN 2573-041X (Online)
Vol. 10, No. 2, 2026
www.scholink.org/ojs/index.php/jecs

Original Paper

A Teaching Note on Simes' Test and the BH Procedure

Chikun Li[1*] & Ziyi Zhen[1]

[1] Institute of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, China

[*] Corresponding author: Chikun Li, E-mail: 120240009@aufe.edu.cn

Received: March 1, 2026 Accepted: April 7, 2026 Online Published: April 13, 2026

doi:10.22158/jecs.v10n2p12 URL: http://dx.doi.org/10.22158/jecs.v10n2p12

Abstract

In many traditional graduate courses in mathematical statistics in China, the main emphasis is still on point estimation, interval estimation and classical single-hypothesis testing, while topics such as global testing and multiple hypothesis testing are often treated very briefly or omitted altogether. At the same time, global and multiple testing have become standard tools in modern applications such as genomics, brain imaging and large-scale A/B testing, where hundreds or thousands of hypotheses are tested simultaneously. This teaching note is designed to help bridge that gap. We present a self-contained proof of Simes' test based on order statistics. We then explain how Simes' test is related to the Benjamini–Hochberg (BH) procedure for controlling the false discovery rate (FDR) in multiple hypothesis testing. R implementations and simulations that can be reproduced by students are provided. The aim is pedagogical rather than innovative: to make global testing and multiple testing accessible in a graduate course.

Keywords

Global testing, Multiple hypothesis testing, Teaching statistics

1. Introduction

Suppose $p_1, \dots, p_n$ are p-values corresponding to null hypotheses $H_{0,1}, \dots, H_{0,n}$. Formally, this means that when $H_{0,i}$ is true, p_i is super-uniform, so

$$\mathbb{P}_{H_{0,i}}(p_i \le s) \le s \quad \text{for each } s \in [0,1].$$

In the global testing problem, we want to test

$$H_0 : H_{0,1}, \dots, H_{0,n} \text{ are all true} \qquad \text{vs.} \qquad H_1 : \text{at least one is false.}$$

Given a desired level α, Bonferroni's global test rejects the global null H_0 if $\min_i p_i \le \alpha/n$, and controls the Type I error rate below α by the union bound. Intuitively, Bonferroni's test is often more powerful against "sparse" alternatives, under which only a small number of the individual null hypotheses are false. In contrast, Fisher's combination test rejects for large values of $-2\sum_i \log p_i$ and tends to be more powerful for "dense" alternatives, where many hypotheses exhibit weak but coordinated signals. Simes' test, introduced as a refinement of Bonferroni's test by Simes (1986), is a less conservative global-testing procedure. We order the p-values as $p_{(1)} \leq \cdots \leq p_{(n)}$. Simes' idea is to extend Bonferroni's rule by rejecting H_0 if

$$p_{(i)} \leq \frac{i}{n}\alpha \quad \text{for some } i \in \{1, \ldots, n\}.$$

Equivalently, one can define the Simes statistic

$$T_n = \min_{1 \leq i \leq n} \left(p_{(i)} \frac{n}{i} \right),$$

and reject when $T_n \leq \alpha$. The factor n/i plays the role of an adjustment that is n for the smallest p-value and decreases as i increases.

This small sharpening of Bonferroni has had an enduring impact. First, Simes' test plays a key role in the theoretical foundation of Hochberg's step-up procedure (Hochberg (1988)) for controlling the familywise error rate (FWER). It serves as the local test within a closed testing framework, from which the Hochberg procedure can be obtained in simplified form. Second, exactly the same principle of ordering the p-values underlies the Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg (1995)), the workhorse method for controlling the false discovery rate (FDR) in large-scale testing. Third, recent expository work, notably Wang (2022), shows that several classical results on FDR control (BH under independence, BH under PRDS, etc.) can all be proved by very short arguments once one is comfortable with ordered p-values and super-uniform statistics. This note is organized to make that path transparent for a graduate audience that may not have seen these topics in their standard curriculum.

2. Simes' Global Test

Simes' global test at level α is

$$\text{reject } H_0 \quad \Leftrightarrow \quad T_n \leq \alpha,$$

where T_n is the Simes statistic defined above. The next theorem (Simes (1986)) shows that, under H_0 and when the p_i's are independent and exactly $\mathrm{Unif}(0,1)$, T_n has the $\mathrm{Unif}(0,1)$ distribution. In particular, the test has exact size α and provides a less conservative global alternative to the classical Bonferroni procedure. Let $U_1, \ldots, U_n \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0,1)$ (i.e., they are independent and identically distributed $\mathrm{Unif}(0,1)$ random variables) and let $U_{(1)} \leq \cdots \leq U_{(n)}$ be their order statistics. Define

$$S_n := \min_{1 \leq i \leq n} \frac{n}{i} U_{(i)}.$$

Theorem 2.1 (Simes 1986). *For $d \in [0,1]$, $\mathbb{P}(S_n \leq d) = d$. Equivalently, $S_n \sim \mathrm{Unif}(0,1)$.*

The crucial observation is that

$$S_n > d \quad \Leftrightarrow \quad U_{(i)} > \frac{i}{n} d \text{ for all } i.$$

The right-hand side describes a polytope obtained by intersecting the standard ordered simplex with the half-spaces $U_{(i)} > (i/n)d,\ i = 1, \dots, n$. For the case $n = 3$, this construction is illustrated in Figure 1.

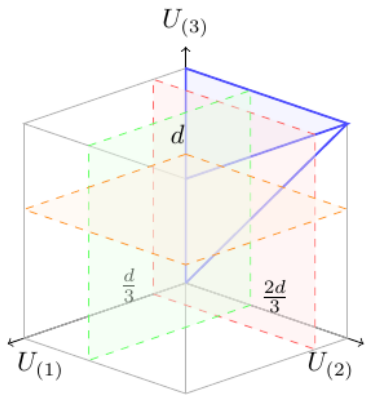

Figure 1. The ordered simplex $\{0 < U_{(1)} < U_{(2)} < U_{(3)} < 1\}$ (blue tetrahedron) inside the unit cube, together with the three truncating planes $U_{(1)} > \frac{d}{3}$, $U_{(2)} > \frac{2d}{3}$ and $U_{(3)} > d$.

In the following section, we compute the probability of this polytope by successive integration.

3. A Proof of Theorem 1 Based on Order Statistics

Proof. The joint density of $U_{(1)}, \dots, U_{(n)}$ is

$$f(x_1, \dots, x_n) = \begin{cases} n!, & 0 < x_1 < \cdots < x_n < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\mathbb{P}(S_n > d) = \mathbb{P}\left(U_{(1)} > \frac{d}{n},\ U_{(2)} > \frac{2d}{n},\ \dots,\ U_{(n)} > d\right).$$

$$\mathbb{P}(S_n > d) = n! \int_d^1 \int_{\frac{n-1}{n}d}^{x_n} \cdots \int_{\frac{2}{n}d}^{x_3} \int_{\frac{1}{n}d}^{x_2} d\,x_1\, dx_2 \cdots dx_{n-1}\, dx_n.$$

The inner $(n-1)$ integrals can be evaluated inductively; one obtains

$$\int_{\frac{n-1}{n}d}^{x_n} \cdots \int_{\frac{1}{n}d}^{x_2} n!\ dx_1 \cdots dx_{n-1} = n!\left(\frac{x_n^{n-1}}{(n-1)!} - \frac{d}{n}\frac{x_n^{n-2}}{(n-2)!}\right).$$

For $1 \le k \le n-1$, define

$$F_k(t) := \int_{\frac{k}{n}d}^{t} \int_{\frac{k-1}{n}d}^{x_k} \cdots \int_{\frac{1}{n}d}^{x_2} n!\ dx_1 \cdots dx_k.$$

We claim that, for all t,

$$F_k(t) = n!\left(\frac{t^k}{k!} - \frac{d}{n}\frac{t^{k-1}}{(k-1)!}\right).$$

Base case $k = 1$.

$$F_1(t) = \int_{\frac{1}{n}d}^{t} n\,!\;dx_1 = n!\left(t - \frac{d}{n}\right),$$

which agrees with the claimed formula for $F_k(t)$ when $k = 1$.

Induction step. Assume the claimed formula for $F_k(t)$ holds for some $k \in \{1, \dots, n-2\}$. Then

$$F_{k+1}(t) = \int_{\frac{k+1}{n}d}^{t} F_k\,(x_{k+1})\,dx_{k+1} = \int_{\frac{k+1}{n}d}^{t} n\,!\left(\frac{x_{k+1}^{k}}{k!} - \frac{d}{n}\,\frac{x_{k+1}^{k-1}}{(k-1)!}\right)dx_{k+1}.$$

Integrating term by term,

$$F_{k+1}(t) = n!\left[\frac{t^{k+1} - \left(\frac{k+1}{n}d\right)^{k+1}}{(k+1)!} - \frac{d}{n}\,\frac{t^{k} - \left(\frac{k+1}{n}d\right)^{k}}{k!}\right].$$

$$F_{k+1}(t) = n!\left[\frac{t^{k+1}}{(k+1)!} - \frac{d}{n}\,\frac{t^{k}}{k!} - C\right].$$

where

$$C := \frac{\left(\frac{k+1}{n}d\right)^{k+1}}{(k+1)!} - \frac{d}{n}\,\frac{\left(\frac{k+1}{n}d\right)^{k}}{k!} = 0.$$

So

$$F_{k+1}(t) = n!\left(\frac{t^{k+1}}{(k+1)!} - \frac{d}{n}\,\frac{t^{k}}{k!}\right),$$

which is exactly the same formula with k replaced by $k + 1$.

By induction, this formula holds for all $k = 1, \dots, n-1$. Taking $k = n-1$ and $t = x_n$ gives

$$\int_{\frac{n-1}{n}d}^{x_n} \cdots \int_{\frac{1}{n}d}^{x_2} n\,!\;dx_1 \cdots dx_{n-1} = F_{n-1}(x_n) = n!\left(\frac{x_n^{n-1}}{(n-1)!} - \frac{d}{n}\,\frac{x_n^{n-2}}{(n-2)!}\right),$$

as claimed.

Substituting and simplifying gives

$$\mathbb{P}(S_n > d) = \int_{d}^{1} [n x^{n-1} - d(n-1)x^{n-2}]\,dx = (1 - d^n) - d(1 - d^{n-1}) = 1 - d.$$

Hence $\mathbb{P}(S_n \le d) = d$, and the proof is completed. □

4. From Simes' Test to the Benjamini–Hochberg Procedure

Rather than only testing the single global null, in practice we often want to identify the false nulls among $H_{0,1}, \dots, H_{0,n}$, not just to know that "at least one" is false. If we reject every $H_{0,i}$ with $p_i \le \alpha/n$, then, under the assumption that the null p-values are super-uniform, the familywise error rate (FWER) is controlled at level at most α (Hommel, 1988). However, the price is high when n is large. Benjamini and Hochberg (1995) proposed to control instead the false discovery rate

$$\mathrm{FDR} = \mathbb{E}\left[\frac{V}{R \vee 1}\right],$$

where V is the number of false rejections and R is the total number of rejections. Their procedure looks very similar to Simes' test.

4.1 The Benjamini–Hochberg (BH) Procedure

Let $p_{(1)} \le \cdots \le p_{(n)}$ and fix a target FDR level $0 < \alpha < 1$. Define

$$k_{\max} = \max\{k: 1 \le k \le n,\ p_{(k)} \le (k/n)\alpha\}.$$

(with the convention $\max\emptyset = 0$). The BH procedure rejects $H_{0,(1)}, \dots, H_{0,(k_{\max})}$ and does not reject the rest.

Thus, BH and Simes use exactly the same sequence of thresholds $\frac{k}{n}\alpha$. The difference is that Simes stops as soon as one ordered p-value is below its line, while BH rejects all ordered p-values up to the largest one that is below its line. This lets us tell a natural teaching story: Simes certifies that "there is something to see"; BH then takes as many discoveries as possible while keeping a global risk small.

4.2 FDR Control under Independence

We now state the standard guarantee for the BH procedure and outline a short proof in the style of Wang (2022). For brevity and when there is no confusion, we will refer to the BH procedure simply as BH.

Theorem 4.1 (Benjamini and Hochberg (1995), independent case). *Suppose the p-values corresponding to true nulls are i.i.d.* $\mathrm{Unif}(0,1)$ *and are independent of the non-null p-values. Then BH procedure at level α satisfies* $\mathrm{FDR} = \frac{n_0}{n}\alpha$*, where n_0 is the number of true nulls.*

Proof sketch. Let $\mathcal{N}$ be the index set of true nulls. The key "leave-one-out" idea is to compare the BH procedure on the original p-values with a modified procedure in which one true-null p-value is set to zero. For each $k \in \mathcal{N}$, define R_k as the number of rejections we would get if we replaced p_k by 0 and left all other p-values unchanged. Then, as shown in Wang (2022),

$$\mathbb{E}\left[\frac{V}{R \vee 1}\right] = \sum_{k\in\mathcal{N}} \sum_{r=1}^{n} \frac{1}{r}\, \mathbb{P}\left(p_k \le \frac{r}{n}\alpha,\ R_k = r\right).$$

By independence, p_k is independent of R_k, and $\mathbb{P}(p_k \le (r/n)\alpha) = (r/n)\alpha$. Hence

$$\mathbb{E}\left[\frac{V}{R \vee 1}\right] = \sum_{k\in\mathcal{N}} \sum_{r=1}^{n} \frac{1}{r}\frac{r}{n}\alpha\, \mathbb{P}(R_k = r) = \frac{\alpha}{n}\sum_{k\in\mathcal{N}} 1 = \frac{n_0}{n}\alpha. \quad \square$$

Note that if $n_0 = n$, then $\mathrm{FDR} = \alpha$, and the event that BH makes at least one rejection coincides with the Simes rejection event. In this sense, Simes' test can be regarded as the global-null counterpart of BH.

5. Dependence, BY, and PRDS

Real p-values are rarely independent. Benjamini and Yekutieli (2001) showed that the BH procedure still controls the FDR under the positive regression dependence on a subset (PRDS) condition. For arbitrary dependence they proposed to scale the BH thresholds by the harmonic number $c_n = \sum_{i=1}^{n} 1/i$:

$$p_{(k)} \le \frac{k}{n}\frac{\alpha}{c_n} \quad \Rightarrow \quad \text{reject } H_{0,(1)}, \dots, H_{0,(k)},$$

and this BY procedure controls FDR at level α without any independence assumptions. Sarkar (1998) gave Simes-type inequalities for MTP_2 variables that also imply validity of Simes-type and BH-type rules beyond independence.

Wang (2022) gives particularly short proofs of these results and shows that the same "replacement" trick used above for BH can be adapted to PRDS, to e-values, and to BY. This is nice for teaching: students learn one pattern and see it reused in several theorems.

6. R Implementations and Simulation Figures

6.1 Simes and BH in R

```
simes_test <- function(p, alpha = 0.05) {
    p <- sort(p)
    n <- length(p)
    simes_stat <- min((n/(1:n)) * p)
    list(statistic = simes_stat,
    reject = (simes_stat <= alpha))
}

bh_procedure <- function(p, alpha = 0.05) {
    o <- order(p)
    p_sorted <- p[o]
    n <- length(p)
    thresh <- (1:n) * alpha / n
    kstar <- max(which(p_sorted <= thresh), 0)
    reject <- rep(FALSE, n)
    if (kstar > 0) reject[o[1:kstar]] <- TRUE
    list(reject = reject, k = kstar,
    ordered_p = p_sorted, threshold = thresh)
}
```

6.2 Monte Carlo check of Simes

```
check_simes <- function(n = 5, B = 5000) {
    simes_vals <- numeric(B)
    for (b in 1:B) {
      u <- sort(runif(n))
      simes_vals[b] <- min((n/(1:n)) * u)
    }
    simes_vals
}
```

```
vals <- check_simes(n = 5, B = 5000)
pdf("simes_ecdf.pdf", width = 5, height = 4)
plot(ecdf(vals),
main = "Empirical CDF of Simes statistic",
xlab = "t", ylab = "F(t)")
abline(0, 1, lty = 2)
dev.off()
```

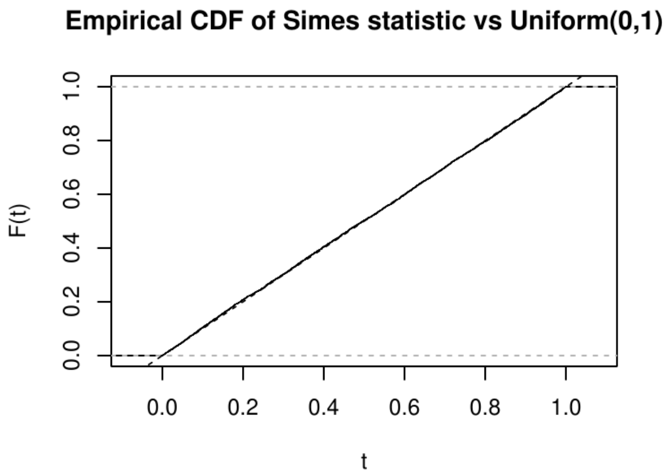

Figure 2. Empirical CDF of the Simes statistic from 5,000 simulations with $n = 5$, compared to the $\mathrm{Unif}(0, 1)$ CDF (dashed). The curves nearly coincide, confirming Theorem 2.1.

6.3 Monte Carlo Check of BH

For BH it is natural to check numerically that the empirical FDR is close to the theoretical value $(n_0/n)\alpha$ under independence. The following code fixes $n = 100$ hypotheses, with $n_0 = 80$ true nulls (uniform p-values) and $n_1 = 20$ non-null p-values drawn from a $\mathrm{Beta}(0.25,1)$ distribution (enriched near 0), and estimates the FDR by simulation.

```
bh_fdr_sim <- function(n = 100, n0 = 80, alpha = 0.05, B = 5000) {
    fdp <- numeric(B)
    for (b in 1:B) {
      # true null p-values: Unif(0, 1)
      p_null <- runif(n0)
      # non-null p-values: stochastically smaller than uniform
      p_alt   <- rbeta(n - n0, shape1 = 0.25, shape2 = 1)
      p <- c(p_null, p_alt)
      # indices of true nulls (first n0)
      null_idx <- 1:n0
      # apply BH
```

```
        res <- bh_procedure(p, alpha = alpha)
        R    <- sum(res$reject)
        V    <- sum(res$reject[null_idx])
        fdp[b] <- if (R == 0) 0 else V / R
    }
    list(fdp = fdp,
    mean_fdr = mean(fdp))
}

set.seed(123)
out <- bh_fdr_sim()
out$mean_fdr

pdf("bh_fdr.pdf", width = 5, height = 4)
hist(out$fdp,
breaks = 30,
main = "FDP under BH (n = 100, n0 = 80, B = 5000)",
xlab = "False discovery proportion")
abline(v = out$mean_fdr, lwd = 2, lty = 2)
dev.off()
```

Running this code prints the estimated mean FDR (which should be close to $(n_0/n)\alpha = 0.04$) and produces a histogram of the false discovery proportion (FDP) across simulations.

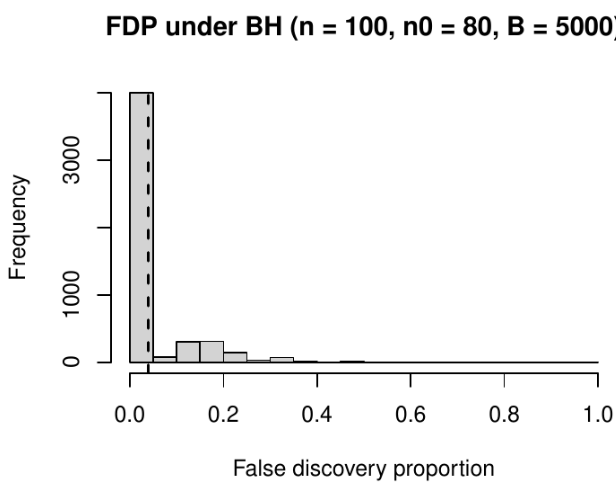

Figure 3. Empirical distribution of the false discovery proportion (FDP) for the BH procedure with $\mathbf{n = 100}$, $\mathbf{n_0 = 80}$, $\boldsymbol{\alpha = 0.05}$ and $\mathbf{B = 5000}$ replications. The dashed line marks the empirical mean FDR, which is close to $\mathbf{(n_0/n)\alpha}$.

7. Recent Literature and Applications

Because multiple testing shows up in so many modern applications, it is useful to point students to a few directions beyond the classical BH/BY setup:

• **High-dimensional biology and neuroimaging.** The original BH paper (Benjamini and Hochberg (1995)) was quickly adopted in microarray analysis; today fMRI studies and proteomics routinely report FDR-controlled lists of discoveries.

• **E-values and e-BH.** Vovk and Wang (2021) and Wang and Ramdas (2022) showed that one can replace p-values by e-values and still run a BH-like step-up rule with FDR control, even under arbitrary dependence.

• **Structured and online FDR.** Ramdas et al. (2019) gave the p-filter to enforce FDR across groups, layers, or time. This is important in large platforms that run many A/B tests and must allocate error across business lines.

These works all reuse the same two ideas emphasized in this note: order the evidence (by p-value or e-value), and control a simple super-uniform (or supermartingale) quantity.

8. Conclusion

We started from a very concrete order-statistic computation developed earlier in this note and showed how it leads naturally to Simes' global test. From there it is only a small step to the Benjamini–Hochberg procedure, whose proof under independence can be done with essentially the same level of probability theory. Once BH is in place, modern extensions (PRDS, BY, e-BH, structured and online FDR) can be understood as variations on the same theme. Because most of the traditional graduate courses in China still spend little time on global testing and multiple testing, we hope this note can be used directly as a teaching handout.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289-300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165-1188.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800-802.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, *75*(2), 383-386.

Ramdas, A., Barber, R. F., Wainwright, M. J., & Jordan, M. I. (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *Annals of Statistics*, *47*(5), 2790-2821.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: A proof of the Simes conjecture. *Annals of Statistics*, *26*(2), 494-504.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751-754.

Vovk, V., & Wang, R. (2021). E-values: Calibration, combination and applications. *Annals of Statistics*, *49*(3), 1736-1754.

Wang, R. (2022). Elementary proofs of several results on false discovery rate. arXiv:2201.09350.

Wang, R., & Ramdas, A. (2022). False discovery rate control with e-values. *Journal of the Royal Statistical Society: Series B*, *84*(3), 822-852.