

RESEARCH ARTICLE

Open Access



A P -value model for theoretical power analysis and its applications in multiple testing procedures

Fengqing Zhang^{1*} and Jiangtao Gou²

Abstract

Background: Power analysis is a critical aspect of the design of experiments to detect an effect of a given size. When multiple hypotheses are tested simultaneously, multiplicity adjustments to p -values should be taken into account in power analysis. There are a limited number of studies on power analysis in multiple testing procedures. For some methods, the theoretical analysis is difficult and extensive numerical simulations are often needed, while other methods oversimplify the information under the alternative hypothesis. To this end, this paper aims to develop a new statistical model for power analysis in multiple testing procedures.

Methods: We propose a step-function-based p -value model under the alternative hypothesis, which is simple enough to perform power analysis without simulations, but not too simple to lose the information from the alternative hypothesis. The first step is to transform distributions of different test statistics (e.g., t , chi-square or F) to distributions of corresponding p -values. We then use a step function to approximate each of the p -value's distributions by matching the mean and variance. Lastly, the step-function-based p -value model can be used for theoretical power analysis.

Results: The proposed model is applied to problems in multiple testing procedures. We first show how the most powerful critical constants can be chosen using the step-function-based p -value model. Our model is then applied to the field of multiple testing procedures to explain the assumption of monotonicity of the critical constants. Lastly, we apply our model to a behavioral weight loss and maintenance study to select the optimal critical constants.

Conclusions: The proposed model is easy to implement and preserves the information from the alternative hypothesis.

Keywords: Critical constants, Multiple testing procedures, Power analysis, p -value

Background

Power analysis is a key technique in the experimental design to reveal an effect of a given size. Traditional power calculation usually assumes a single hypothesis test, but it is quite common for researchers to test several hypotheses simultaneously. Clinical trials often require two or more hypotheses to be tested, and studies which involve comparing treatments using multiple outcome measures happen frequently in medical research [8]. The development of high-throughput biology leads to a dramatic increase in the number of hypothesis tests in genomics

[20]. However, there are a limited number of studies on power analysis in multiple testing procedures. For scientific studies with multiple hypotheses, in order to correctly control the false positives, multiplicity adjustments to p -values should be taken into account in power analysis. A consequence of multiplicity adjustments is the loss of power [21] and the change in sample size requirements [16].

The p -value is a tail probability given the null hypothesis is true. Under the null hypothesis, the p -value is a uniformly distributed random variable between 0 and 1. If the null hypothesis is false, the p -value's distribution depends

*Correspondence: fz53@drexel.edu

¹Department of Psychology, Drexel University, 3201 Chestnut Street, 19104 Philadelphia, USA

Full list of author information is available at the end of the article

on the alternative hypothesis, which usually satisfies the inequality

$$\Pr(P \leq p) \geq p, \tag{1}$$

i.e., the random variable P is less than a standard uniform random variable in the stochastic order, where the random variable P is the probability of rejecting the null hypothesis when the alternative hypothesis is true.

In order to perform a p -value based power analysis, certain distribution models are needed to describe the behavior of p -values under the alternative hypothesis. In general there are two approaches. The p -value models based on the original test statistics [17] or based on copulas [27] are usually with complex expressions, and further calculations or evaluations require integrations. So the theoretical analysis is difficult and numerical simulations are often needed. The p -value models based on Dirac function [10, 24] are over-simplified, which limits their application areas.

In this paper, we propose the step-function-based p -value models under the alternative hypothesis, which are simple enough to perform theoretical power analysis, but not too simple to lose the information from the alternative hypothesis. Two applications in multiple testing procedures are shown and one application in weight-loss treatment is given.

Methods

A widely used p -value model assumes that the statistic under the null hypothesis follows $N(0, 1^2)$, and the statistics under the alternative hypothesis follows $N(\delta, 1^2)$. The p -values are calculated based on one-sided test [17].

The density function of the normal-distribution-based p -value model is

$$h(p) = \frac{\phi(\Phi^{-1}(p) + \delta)}{\phi(\Phi^{-1}(p))}, \tag{2}$$

with mean and variance

$$E[P] = \int_{-\infty}^{+\infty} \Phi(x)\phi(x + \delta) dx,$$

$$\text{var}[P] = \int_{-\infty}^{+\infty} [\Phi(x)]^2 \phi(x + \delta) dx - \left(\int_{-\infty}^{+\infty} \Phi(x)\phi(x + \delta) dx \right)^2.$$

Alternatively, we propose a step-function-based p -value model under the alternative hypothesis, which has a density function

$$h(p) = \begin{cases} f, & 0 \leq p \leq \frac{1-g}{f-g}, \\ g, & \frac{1-g}{f-g} < p \leq 1, \end{cases} \tag{3}$$

with mean and variance

$$E[P] = \frac{1-2g+fg}{2(f-g)}, \quad \text{var}[P] = \frac{1+4fg^2+4f^2g-3f^2g^2}{12(f-g)^2},$$

where $0 \leq g \leq 1 \leq f$. The parameter (f, g) indicates the deviation of the random variable P under the step-function-based p -value model from a standard uniform random variable.

For comparison between the normal-distribution-based p -value model and the step-function-based p -value model, the parameters (f, g) of the step-function-based p -value model were chosen to match the means and the variances for the normal-distribution-based p -value model with parameter δ , as shown in Table 1. For the simplified step-function-based p -value model with parameter f , means are matched with the normal model.

In addition, a simplified step-function-based p -value model with a single parameter $f \in [1, +\infty)$ is achieved when assuming $g = 0$. The corresponding density function is

$$h_0(p) = \begin{cases} f, & 0 \leq p \leq \frac{1}{f}, \\ 0, & \frac{1}{f} < p \leq 1. \end{cases} \tag{4}$$

with mean and variance

$$E[P] = \frac{1}{2f}, \quad \text{var}[P] = \frac{1}{12f^2}.$$

For the simplified step-function-based p -value model, a larger parameter f corresponds to a larger effect size. As a

Table 1 Normal-distribution-based p -value model and step-function-based p -value model

δ	Normal		Step-function			Simplified step-function	
	Mean	SD	f	g	$(1-g)/(f-g)$	f	$1/f$
0.2	0.444	0.286	1.249	0.795	0.451	1.127	0.888
0.5	0.362	0.274	1.728	0.555	0.380	1.382	0.724
1	0.240	0.236	2.937	0.288	0.269	2.086	0.480
2	0.079	0.130	9.061	0.059	0.105	6.357	0.157
3	0.017	0.049	37.143	0.007	0.027	29.503	0.034

special case, when $f = 1$, the distribution is uniform $[0, 1]$. The probability density functions and the cumulative distribution functions of the normal-distribution-based and the step-function-based p -value models are compared in Fig. 1. The step-function-based p -value models serve an approximation to the p -value models based on the original test statistics.

Based on the univariate model, the corresponding multivariate p -value model has a density function

$$h(p_1, \dots, p_n) = \prod_{i=1}^n h_i(p_i), \tag{5}$$

where different $h_i(\cdot)$'s may have different parameters f_i 's and g_i 's. In the following sections, the simplified step-function-based p -value model is applied to two problems in multiple testing procedures.

Results

Application: optimal choices of the critical constants

Assume n hypotheses $\{H_i\}_{i=1}^n$ with p -values $\{p_i\}_{i=1}^n$. Sort p -values as $p_{(1)} \leq \dots \leq p_{(n)}$, and $H_{(1)}, \dots, H_{(n)}$ are the corresponding null hypotheses. Consider testing the global null hypothesis $\cap_{i=1}^n H_i$ under the control of type I error [15]

$$\Pr(\cup_{i=1}^n \{p_{(n-i+1)} \leq c_i \alpha\}) \leq \alpha.$$

The global test compares $p_{(n-i+1)}$ with its corresponding critical constant $c_i \alpha$ for every $i = 1, \dots, n$. If for some i 's, $p_{(n-i+1)} \leq c_i \alpha$, then the global null hypothesis is rejected. Simes [25] proposed a test with $c_i = (n - i + 1) / n$, and other choices of c_i 's were proposed by Rom [22], Cai and Sarkar [7], Gou and Tamhane [11].

Among different choices of critical constants c_i 's, people usually run simulations [19] or rely on numerical calculations [14] to make power comparison in order to choose suitable sets of critical constants. Besides the existing computationally intensive methods, the step-function-based p -value model is an alternative choice to theoretically calculate the powers and make comparisons between different multiple testing procedures.

In this section we first show how the most powerful critical constants can be chosen using our proposed method for a global test with two hypotheses, and then we can apply the successive recursion process to calculate the power for a global test with n hypotheses. For a global test with two single hypotheses, the control of type I error under independence requires

$$\Pr(\{p_{(2)} \leq c_1 \alpha\} \cup \{p_{(1)} \leq c_2 \alpha\}) = \alpha,$$

where $c_1 \geq c_2$. This equality is equivalent to

$$(1 - 2c_2) + c_1(2c_2 - c_1)\alpha = 0. \tag{6}$$

From (6) we get $c_2 = \frac{1 - c_1^2 \alpha}{2(1 - c_1 \alpha)}$, and the derivative $\frac{dc_2}{dc_1} = -\frac{(c_1 - c_2)\alpha}{1 - c_1 \alpha}$, which is less than zero. So c_2 decreases when c_1 increases.

By using the simplified step-function-based p -value model, without loss of generality, we assume $1 \leq f_1 \leq f_2$, the probability of rejecting the global hypothesis is

Rejection probability

$$= \begin{cases} c_2(f_1 + f_2)\alpha + c_1(c_1 - 2c_2)f_1f_2\alpha^2, & \text{when } f_1 \leq f_2 \leq \frac{1}{c_1\alpha}, \\ (c_1f_1 + c_2f_2)\alpha - c_1c_2f_1f_2\alpha^2, & \text{when } f_1 \leq \frac{1}{c_1\alpha} \leq f_2 \leq \frac{1}{c_2\alpha}, \\ 1, & \text{when } f_1 \geq \frac{1}{c_1\alpha} \text{ or } f_2 \geq \frac{1}{c_2\alpha}. \end{cases}$$

When the global hypothesis contains two single hypotheses ($n = 2$), there are two configurations of

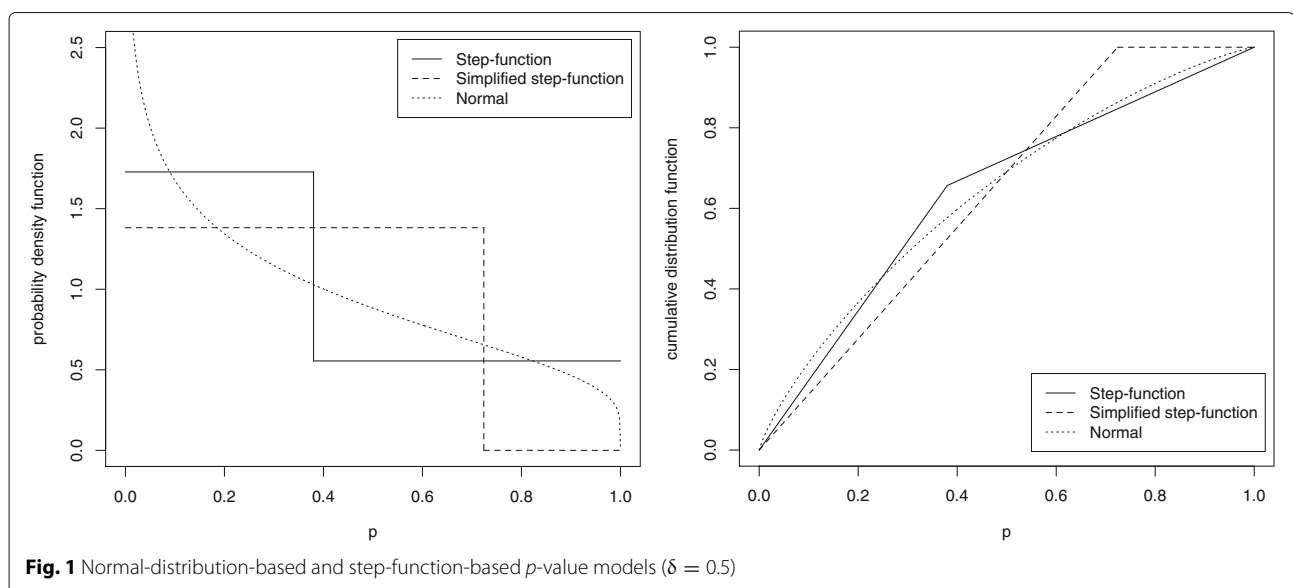


Fig. 1 Normal-distribution-based and step-function-based p -value models ($\delta = 0.5$)

true and non-true null hypotheses where the global null hypothesis is false: (1) one true null hypothesis ($n_0 = 1$) and one false null hypothesis ($m = 1$), and (2) two false null hypotheses ($m = 2$).

First, assume that one null hypothesis is true and the other is false, say, $f_1 = 1$ and $f_2 = f$, then the power is

$$\text{Power} = \begin{cases} c_2(1+f)\alpha + fc_1(c_1 - 2c_2)\alpha^2 & \text{when } f \leq \frac{1}{c_1\alpha}, \\ (c_1 + fc_2)\alpha - fc_1c_2\alpha^2, & \text{when } \frac{1}{c_1\alpha} \leq f \leq \frac{1}{c_2\alpha}, \\ 1, & \text{when } f \geq \frac{1}{c_2\alpha}. \end{cases}$$

When $f \leq \frac{1}{c_1\alpha}$, from (6) it follows that $c_1(c_1 - 2c_2)\alpha = 1 - 2c_2$, therefore

$$c_2(1+f)\alpha + fc_1(c_1 - 2c_2)\alpha^2 = (f - (f - 1)c_2)\alpha,$$

hence power increases when c_2 decreases (c_1 increases).

When $\frac{1}{c_1\alpha} \leq f \leq \frac{1}{c_2\alpha}$, from (6) we get $c_1c_2\alpha = \frac{1}{2}(2c_2 - 1 + c_1^2\alpha)$, consequently

$$\begin{aligned} (c_1 + fc_2)\alpha - fc_1c_2\alpha^2 &= \frac{\alpha}{2}(f + 2c_1 - f\alpha c_1^2) \\ &= \frac{1}{2f}(1 + f^2\alpha - (f\alpha c_1 - 1)^2). \end{aligned}$$

Note that $f\alpha c_1 \geq 1$, so power increases when c_1 decreases (c_2 increases).

We calculate the maximal power for different f 's. When we assume the alternative hypothesis has a small effect size, where $f \leq 1/\sqrt{\alpha}$, the maximal power is achieved when $c_2 = 0$ and $c_1 = 1/\sqrt{\alpha}$. When we assume the alternative hypothesis has a moderate effect size, where $1/\sqrt{\alpha} < f \leq (1 + \sqrt{1 - \alpha})/\alpha$, the maximal power is achieved when $c_1 = 1/(f\alpha)$ and $c_2 = (f^2\alpha - 1)/(2\alpha f(f - 1))$, so when f increases, we can follow the strategy to decrease c_1 (increase c_2) to achieve the maximal power. When we assume the alternative hypothesis has a large effect size, where $f > (1 + \sqrt{1 - \alpha})/\alpha$, the maximal power is achieved when we choose a test with $c_2 \geq 1/(f\alpha)$, so when f is large enough, different tests have similar power.

$$\text{max Power} = \begin{cases} f\alpha, & \text{when } f \leq \frac{1}{\sqrt{\alpha}}, \\ \frac{1+f^2\alpha}{2f}, & \text{when } \frac{1}{\sqrt{\alpha}} < f \leq \frac{1+\sqrt{1-\alpha}}{\alpha}, \\ 1, & \text{when } f > \frac{1+\sqrt{1-\alpha}}{\alpha}. \end{cases} \tag{7}$$

Second, assume that both null hypothesis are false with the same effect size, say, $f_1 = f_2 = f$, then the power is

$$\text{Power} = \begin{cases} 2fc_2\alpha + f^2c_1(c_1 - 2c_2)\alpha^2, & \text{when } f \leq \frac{1}{c_1\alpha}, \\ 1, & \text{when } f \geq \frac{1}{c_1\alpha}. \end{cases}$$

When $f \leq \frac{1}{c_1\alpha}$, from (6) we get $c_1(c_1 - 2c_2)\alpha = 1 - 2c_2$, then

$$2fc_2\alpha + f^2c_1(c_1 - 2c_2)\alpha^2 = f\alpha(f - 2(f - 1)c_2),$$

power increases when c_2 decreases (c_1 increases).

We calculate the maximal power for different f values. When we assume both hypotheses are false and with a small effect size, where $f \leq 1/\sqrt{\alpha}$, the maximal power is achieved when $c_2 = 0$ and $c_1 = 1/\sqrt{\alpha}$. When we assume both false hypotheses have a big effect size, where $f \geq 1/\sqrt{\alpha}$, the maximal power is achieved when the test satisfies $c_1 \geq 1/(f\alpha)$. By taking both $f \leq 1/\sqrt{\alpha}$ and $f \geq 1/\sqrt{\alpha}$ into account, it follows that $c_1 = 1/\sqrt{\alpha}$ and $c_2 = 0$ is the uniformly best choice when both null hypotheses are false.

$$\text{max Power} = \begin{cases} f^2\alpha, & \text{when } f \leq \frac{1}{\sqrt{\alpha}}, \\ 1, & \text{when } f > \frac{1}{\sqrt{\alpha}}. \end{cases} \tag{8}$$

In general, for a global test with n single hypotheses, where m of them are true significances (false null hypotheses), define the probabilities as

$$B_{n,m,i} = \begin{cases} \Pr_{n,m}(p_{(n)} \leq c_1\alpha), & i = 1, \\ \Pr_{n,m}(p_{(n)} > c_1\alpha, \dots, p_{(n-i+2)} > c_{i-1}\alpha, p_{(n-i+1)} \leq c_i\alpha), & i = 2, \dots, n, \\ \Pr_{n,m}(p_{(n)} > c_1\alpha, \dots, p_{(1)} > c_n\alpha), & i = n + 1, \end{cases}$$

where $\Pr_{n,m}$ indicates the probability for n hypotheses, where m is the number of the true significances, and $n_0 = n - m$ is the number of the true nulls.

Since

$$\begin{aligned} B_{n,m,n} &= \Pr_{n,m}(p_{(n)} > c_1\alpha, \dots, p_{(2)} > c_{n-1}\alpha, p_{(1)} \leq c_n\alpha) \\ &= (n - m)\Pr_{n-1,m}(p_{(n-1)} > c_1\alpha, \dots, p_{(1)} > c_{n-1}\alpha) \Pr_{1,0}(p \leq c_n\alpha) \\ &\quad + m\Pr_{n-1,m-1}(p_{(n-1)} > c_1\alpha, \dots, p_{(1)} > c_{n-1}\alpha) \Pr_{1,1}(p \leq c_n\alpha), \end{aligned}$$

we have the recurrence relation for $i = n$

$$B_{n,m,n} = (n - m)c_n\alpha B_{n-1,m,n} + m(fc_n\alpha \wedge 1)B_{n-1,m-1,n}. \tag{9}$$

Similarly, for general i , since

$$\begin{aligned} B_{n,m,i} &= \Pr_{n,m}(p_{(n)} > c_1\alpha, \dots, p_{(n-i+2)} > c_{i-1}\alpha, p_{(n-i+1)} \leq c_i\alpha) \\ &= \frac{n-m}{n-i+1}\Pr_{n-1,m}(p_{(n-1)} > c_1\alpha, \dots, p_{(n-i+1)} > c_{i-1}\alpha, p_{(n-i)} \leq c_i\alpha) \Pr_{1,0}(p \leq c_i\alpha) \\ &\quad + \frac{m}{n-i+1}\Pr_{n-1,m-1}(p_{(n-1)} > c_1\alpha, \dots, p_{(n-i+1)} > c_{i-1}\alpha, p_{(n-i)} \leq c_i\alpha) \Pr_{1,1}(p \leq c_i\alpha), \end{aligned}$$

we have the general recurrence relation for i

$$\begin{aligned} B_{n,m,i} &= \frac{n-m}{n-i+1}c_i\alpha B_{n-1,m,i} \\ &\quad + \frac{m}{n-i+1}(fc_i\alpha \wedge 1)B_{n-1,m-1,i}. \end{aligned} \tag{10}$$

Finner and Roters [9], Cai and Sarkar [7], and Gou and Tamhane [11] defined a special case of the probability $B_{n,m,i}$ for $m = 0$ to calculate the type I error under the

global null hypothesis. They proved a recurrence relationship among $B_{n,0,i}$'s. We generalize this result to $B_{n,m,i}$'s and have this recurrence relationship (10) under the simplified step-function-based p -value model for power analysis.

By starting from

$$\begin{aligned} B_{1,0,1} &= c_1\alpha, & B_{1,0,2} &= 1 - c_1\alpha, \\ B_{1,1,1} &= fc_1\alpha \wedge 1, & B_{1,0,2} &= 1 - fc_1\alpha \wedge 1, \end{aligned}$$

and using the recurrence relation (10), the power is calculated by

$$\text{Power}_{n,m} = \sum_{i=1}^n B_{n,m,i} \tag{11}$$

Note that

$$\sum_{i=1}^{n+1} B_{n,m,i} = 1,$$

and the control of type I error is satisfied if

$$\sum_{i=1}^n B_{n,0,i} \leq \alpha.$$

When f is specified and the set of critical constants $\{c_i\}_{i=1}^n$ is given, the exact power can be calculated by using (11). Since only arithmetic calculations are needed, the power can be computed very fast.

For theoretical analysis, we consider the situation where f is not too small.

$$\text{Power}_{n,m} = 1, \quad \text{when } f \geq \frac{1}{c_{n+1-m}\alpha}.$$

The largest possible c_{n+1-m} is achieved by using the set of critical constants which satisfies $c_1 = c_2 = \dots = c_{n+1-m} = c$, and $c_{n-m+2} = \dots = c_n = 0$. The control of type I error requires that

$$\sum_{i=0}^{n-m} \binom{n}{i} (c\alpha)^{n-i} (1 - c\alpha)^i \leq \alpha, \tag{12}$$

the largest possible c_{n+1-m} can be solved from (12). If we only take the leading term, we have an approximate solution

$$c_{n+1-m} \approx \frac{1}{\sqrt[m]{\binom{n}{m}\alpha^{m-1}}}$$

So when

$$f \gtrsim \sqrt[m]{\frac{\binom{n}{m}}{\alpha}}$$

the maximal power is achieved when the test satisfies $c_{n+1-m} \geq 1/(f\alpha)$.

Note that when m is relatively large (e.g., more than $n/2$), the bound $\sqrt[m]{\binom{n}{m}}/\alpha$ is small, and the global tests with large c_1, \dots, c_{n+1-m} and small c_{n+2-m}, \dots, c_n tend

to have large power. Similar observations were reported by Gou and Tamhane [11] based on simulations.

Note that in this application power is simply the probability of rejecting $H_0 = \cap_{i=1}^n H_i$ where at least one H_i is false. For testing multiple hypothesis, powers can be of different types: individual, average, disjunctive, and conjunctive, and the appropriate power concept is determined on a case-by-case basis [4]. These power definitions can also be used in the proposed method by using the step-function-based p -value model.

Power analysis can be complex when multiple hierarchical objectives are involved. Alosch and Huque [1] discussed the power for testing hierarchically ordered endpoints. The step-function-based p -value model can be applied to various clinical trials, e.g., group sequential designs [18], graphical procedures [5, 6].

Application: monotonicity of the critical constants

For multiple test procedures, critical constants are often required to satisfy [7, 12]

$$c_1 \geq c_2 \geq \dots \geq c_n \tag{13}$$

This requirement is called the monotonicity assumption of critical constants.

Suppose that we have a set of critical constants c_1^*, \dots, c_n^* , and $c_k^* < c_{k+1}^*$, so the monotonicity assumption is not satisfied. Note that

$$\begin{aligned} \Pr\left(\bigcup_{i=1}^n \{p_{(n+1-i)} \leq c_i^* \alpha\}\right) &= \Pr\left(\bigcup_{i=1, i \neq k}^n \{p_{(n+1-i)} \leq c_i^* \alpha\} \cup \{p_{(n+1-k)} \leq c_{k+1}^* \alpha\}\right) \end{aligned}$$

So if a test with critical constants $c_1^*, \dots, c_{k-1}^*, c_k^*, c_{k+1}^*, \dots, c_n^*$ controls type I error below α , then another test with critical constants $c_1^*, \dots, c_{k-1}^*, c_{k+1}^*, c_{k+1}^*, \dots, c_n^*$, which satisfies the monotonicity assumption, also controls type I error below α , and has the same power with the previous test which does not satisfy the monotonicity assumption. Hence, only the set of critical constants which satisfies the monotonicity assumption needs to be considered.

Many multiple tests have critical constants which satisfy a strict monotonicity assumption [11, 22, 25]

$$c_1 > c_2 > \dots > c_n. \tag{14}$$

Some multiple tests satisfy the monotonicity assumption (13), but do not satisfy the strict monotonicity assumption (14) [3, 26]. In general, these tests are not as powerful as the tests which satisfy the strict monotonicity assumption (14) [11]. Our step-function-based p -value models can explain that this assumption is necessary because the corresponding tests are generally more powerful than other tests which do not satisfy this assumption.

For multiple tests with two single hypotheses, by using the simplified step-function-based p -value model, we

have several observations: (1) when there is one true null hypothesis and one false null hypothesis, only if $f = (1 + \sqrt{1 - \alpha}) / \alpha$, the test which does not satisfy (14) can be more powerful than or as powerful as all the tests which satisfy (14), (2) when there are two false null hypotheses, the test does not satisfy (14) is less powerful than some of the tests which satisfy (14) for all f values. In general, on the parameter space of the effect size (under simplified step-function-based p -value model, the effect size is a function of parameter f), the tests which do not satisfy (14) have more power than all other tests which satisfy (14) only at a zero measure subspace of the effect size. This fact explains that usually people prefer multiple tests which satisfy the strict monotonicity property (14), because these tests are generally more powerful than the tests which do not satisfy (14).

A worked example

Annesi et al. [2] evaluated behavioral weight-loss treatments. They recruited 110 women whose BMI's are between 30 and 40 kg/m², and randomly assigned the participants to a comparison treatment with a print manual and telephone follow-ups, or an experimental treatment of the coach approach exercise-support protocol. The self-efficacy for controlled eating (SE-eating) is one of the psychological predictors of behavioral changes. Annesi et al. [2] reported that during the weight-loss phase (month 0-6), the SE-eating increases in the experimental group were significantly greater than the increases in the comparison group with $t = 2.88$, and there was no significant between-group difference during the weight-loss maintenance phase (month 6-24) with $t = -0.48$.

The increases of the self-efficacy for controlled eating were evaluated both during the weight-loss phase and during the weight-loss maintenance phase, so the multiplicity adjustment is advised to apply. To choose the optimal multiplicity correction based on the estimated effect size from the pilot study, we recommend our proposed step-function-based p -value model because it is easy to implement and preserves the information from the alternative hypothesis. If we take the weight-loss study by Annesi et al. [2] as a pilot study, we have the information that the standardized SE-eating increase during the weight-loss phase is normally distributed with mean $\delta = 2.88$ and variance 1, and the increase during the weight-loss maintenance phase is normally distributed with mean $\delta = 0$ and variance 1. To match the mean for the normal-distribution-based p -value model with parameter $\delta = 2.88$, the parameter f of the simplified step-function-based p -value model is 23.979. From (7) and by using the significance level $\alpha = 0.05$, the maximal power is $(1 + f^2\alpha)/(2f) = 62\%$, and the optimal choice of critical constants is $(c_1, c_2) = (1/(f\alpha), (f^2\alpha - 1)/(2\alpha f(f - 1))) = (0.8341, 0.5036)$.

So the larger p -value is compared with 0.8341α and the smaller p -value is compared with 0.5036α , and if any p -value is less than the corresponding critical value, the global null hypothesis will be rejected.

The step-function-based p -value models for power analysis simplify the theoretical analysis that is difficult in many situations. At the same time, information of loss remains at an acceptable level. Finner and Gontscharuk [10] and Sarkar et al. [24] used a tool called the Dirac-uniform configuration for power analysis, where all p -values under the false null hypotheses follow a Dirac distribution with point mass at 0. When the Dirac-uniform configuration is applied to Annesi et al.'s [2] study, the information of δ is lost, and any choice of positive critical constants (c_1, c_2) will result a claim of significance. Under the p -value model based on Dirac function, all choices of critical constants have the same power, and the optimal choice is unable to be located.

The Dirac-uniform configuration is too brief to include necessary information to choose the optimal critical constants. Hung et al. [17] discussed a p -value model based on normal distribution. By using Annesi et al.'s [2] research as a pilot study, one test statistic is $N(\delta, 1^2)$, and the other test statistic is $N(0, 1^2)$. The probability of rejecting the global null hypothesis is

$$\begin{aligned} \text{Power} &= \left[\int_{-\infty}^{\Phi^{-1}(c_1\alpha)} \int_{-\infty}^{\Phi^{-1}(c_1\alpha)} + \int_{-\infty}^{\Phi^{-1}(c_2\alpha)} \int_{\Phi^{-1}(c_1\alpha)}^{+\infty} \right. \\ &\quad \left. + \int_{\Phi^{-1}(c_1\alpha)}^{+\infty} \int_{-\infty}^{\Phi^{-1}(c_2\alpha)} \right] \\ &\quad \phi(x_1 + \delta)\phi(x_2) dx_1 dx_2 \\ &= c_1\alpha \cdot \Phi(\Phi^{-1}(c_1\alpha) + \delta) \\ &\quad + c_2\alpha (1 - \Phi(\Phi^{-1}(c_1\alpha) + \delta)) \\ &\quad + (1 - c_1\alpha) \Phi(\Phi^{-1}(c_2\alpha) + \delta), \end{aligned}$$

The optimal choice of critical constants is followed by solving

$$\begin{aligned} &\text{maximize}_{c_1, c_2} \text{Power}(c_1, c_2) \tag{15} \\ &\text{subject to } (1 - 2c_2) + c_1(2c_2 - c_1)\alpha = 0. \end{aligned}$$

This optimization problem has no explicit solution. When $\delta = 2.88$ and $\alpha = 0.05$, the optimal solution is $(c_1, c_2) = (1.0076, 0.4998)$. While the step-function based p -value model and the normal-distribution based p -value model produce similar choices of critical constants (c_1, c_2) , the model based on step function has an explicit solution of critical constants and requires little computational effort.

Discussion and conclusions

We have given a step-function-based p -value model and its simplified version. These p -value models are simple

and concise to perform theoretical power analysis. In addition, different test statistics, for example, t , chi-square or F , can be transformed to the p -value scale. These models can be applied to the field of multiple testing procedures to explain the assumption of monotonicity of the critical constants. We also use these p -value models to choose suitable sets of critical constants with more power. In this paper, we consider the independent p -values for the multivariate cases. Dependence structures can be brought into these p -value models, like Sarkar et al. [23] or Gou and Tamhane [13], and we will report the dependent multivariate p -value models in a separate paper. Finally, there are many applications of the step-function-based p -value models, and multiple testing procedure is an example in point.

Abbreviations

BMI: Body mass index; SE-eating: Self-efficacy for controlled eating

Acknowledgments

We thank the editor and referees for their comments which helped to improve the paper.

Funding

The researchers did not receive external sources of funding.

Availability of supporting data

Not applicable.

Authors' contributions

FZ and JG conceived and designed the study. FZ conducted the statistical analysis and drafted the manuscript. JG developed methods and revised the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Department of Psychology, Drexel University, 3201 Chestnut Street, 19104 Philadelphia, USA. ²Department of Mathematics and Statistics, Hunter College of CUNY, 695 Park Avenue, 10065 New York, USA.

Received: 24 March 2016 Accepted: 20 September 2016

Published online: 10 October 2016

References

- Alosh M, Huque MF. A consistency-adjusted alpha-adaptive strategy for sequential testing. *Stat Med*. 2010;29:1559–71.
- Annesi JJ, Johnson PH, Tennant GA, Porter KJ, McEwen KL. Weight loss and the prevention of weight regain: evaluation of a treatment model of exercise self-regulation generalizing to controlled eating. *Permanente J*. 2016;20:1–15.
- Bauer P. Sequential tests of hypotheses in consecutive trials. *Biom J*. 1989;31:663–76.
- Bretz F, Hothorn T, Westfall P. Multiple Comparisons Using R; 2010.
- Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28:586–604.
- Burman CF, Sonesson C, Guilbaud O. A recycling approach for the construction of Bonferroni-based multiple tests. *Stat Med*. 2009;28:739–61.
- Cai G, Sarkar SK. Modified Simes' critical values under independence. *Stat Probab Lett*. 2008;78:1362–78.
- Feise RJ. Do multiple outcome measures require p -value adjustment. *BMC Med Res Methodol*. 2002;2:1–4.
- Finner H, Roters M. On the limit behavior of the joint distribution of order statistics. *Ann Inst Stat Math*. 1994;46:343–9.
- Finner H, Gontscharuk V. Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *J R Stat Soc Ser B*. 2009;71:1031–48.
- Gou J, Tamhane AC. On generalized Simes critical constants. *Biom J*. 2014;56:1035–54.
- Gou J, Tamhane AC, Xi D, Rom D. A class of improved hybrid Hochberg-Hommel type step-up multiple test procedures. *Biometrika*. 2014;101:899–911.
- Gou J, Tamhane AC. Hochberg procedure under negative dependence. 2015. Technical report, Department of Statistics, Northwestern University, Evanston, Illinois.
- Hayter AJ, Tamhane AC. Sample size determination for step-down multiple test procedures: orthogonal contrasts and comparisons with a control. *J Stat Plan Infer*. 1991;27:271–90.
- Hochberg Y, Tamhane AC. *Multiple Comparison Procedures*. New York: John Wiley; 1987.
- Hsu JC. Sample size computation for designing multiple comparison experiments. *J Comput Stat Data Anal*. 1988;7:79–91.
- Hung HM, O'Neill RT, Bauer P, Köhne K. The behavior of the p -Value when the alternative hypothesis is true. *Biometrics*. 1997;53:11–22.
- Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. New York: Chapman and Hall/CRC; 2000.
- Jung S, Bang H, Young S. Sample size calculation for multiple testing in microarray data analysis. *Biostatistics*. 2005;6:157–69.
- Lazzeroni LC, Ray A. The cost of large numbers of hypothesis tests on power, effect size and sample size. *Mol Psychiatry*. 2012;17:108–14.
- Maxwell SE, Kelley K, Rausch RJ. Sample size planning for statistical power and accuracy in parameter estimation. *Annu Rev Psychol*. 2008;59:537–63.
- Rom DM. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*. 1990;77:663–5.
- Sarkar SK, Fu Y, Guo W. Improving Holm's procedure using pairwise dependencies. *Biometrika*. 2016;103:237–43.
- Sarkar SK, Guo W, Finner H. On adaptive procedures controlling the familywise error rate. *J Stat Plan Infer*. 2012;142:65–78.
- Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 1986;73:751–4.
- Šidák Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc*. 1967;62:626–33.
- Stange J, Bodnar T, Dickhaus T. Uncertainty quantification for the family-wise error rate in multivariate copula models. *ASTA Adv Stat Anal*. 2015;99:281–310.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

