Optimal allocation to maximize power of two-sample tests for binary response

David Azriel[1], Micha Mandel[1] and Yosef Rinott[1,2]

[1] Department of statistics, The Hebrew University of Jerusalem, Israel

[2] Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel

October 10, 2011

Abstract

We study allocations that maximize the power of tests of equality of two treatments having binary outcomes. When a normal approximation applies, the asymptotic power is maximized by minimizing the variance, leading to Neyman allocation that assigns observations in proportion to the standard deviations. This allocation, which in general requires knowledge of the parameters of the problem, is recommended in a large body of literature. Under contiguous alternatives the normal approximation indeed applies, and in this case Neyman allocation reduces to a balanced design. However, when studying the power under a non-contiguous alternative, a large deviations approximation is needed, and Neyman allocation is no longer asymptotically optimal. In the latter case, the optimal allocation depends on the parameters, but turns out to be rather close to a balanced design. Thus, balanced design is a viable option for both contiguous and non-contiguous alternatives. Finite sample studies show that balanced design is indeed generally quite close to being optimal for power maximization. This is good news as implementation of balanced design does not require knowledge of the parameters.

Keywords: Adaptive design, Asymptotic power, Bahadur efficiency, Neyman allocation, Pitman efficiency.

1 Introduction

Let A and B be two treatments with unknown probabilities of success, $p_A, p_B \in (0,1)$. A trial is planned with $n_A > 0$ and $n_B > 0$ subjects assigned to treatment A and B, respectively, where $n_A + n_B = n$. For each subject, a binary response, success or failure, is observed. Let $\nu_n = n_A/n$ be the proportion of subjects assigned to treatment A. We sometimes refer to ν_n as the *allocation*. Our focus is on the first question appearing in Chapter 2 of Hu & Rosenberger (2006): "what allocation maximizes power?" We study in detail the Wald test of the hypothesis $p_A = p_B$ versus one or two-sided alternatives; other tests will be discussed briefly.

For $i = A, B$ let $Y_i(m) \sim Bin(m, p_i)$ be the number of successes if m patients are assigned to treatment i and $\hat{p}_i = \hat{p}_i(n_i) = Y_i(n_i)/n_i$ be the estimator of p_i. The Neyman allocation rule, $\nu = \{p_A(1-p_A)\}^{1/2}/[\{p_A(1-p_A)\}^{1/2} + \{p_B(1-p_B)\}^{1/2}]$, minimizes the variance of the estimator $\hat{p}_B - \hat{p}_A$, and is often used for power maximization; see, e.g., Brittain & Schlesselman (1982), Rosenberger et al. (2001), Hu & Rosenberger (2003, 2006), Tymofyeyev et al. (2007), Zhu & Hu (2010), Biswas et al. (2010) and Chambaz & van der Laan (2011).

Let $W = n^{1/2}(\hat{p}_B - \hat{p}_A)\big/V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)$ be the Wald statistic, where $V(p_A, p_B, \nu_n) = p_A(1-p_A)/\nu_n + p_B(1-p_B)/(1-\nu_n)$. The normal approximation argument for power calculation is:

$$\begin{aligned}
&\mathrm{P}_{p_A,p_B}(W > z_{1-\alpha}) = \\
&\mathrm{P}_{p_A,p_B}\left[\frac{n^{1/2}\{\hat{p}_B - \hat{p}_A - (p_B - p_A)\}}{V^{1/2}(p_A, p_B, \nu_n)} > \frac{z_{1-\alpha}V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n) - n^{1/2}(p_B - p_A)}{V^{1/2}(p_A, p_B, \nu_n)}\right] \\
&\approx 1 - \Phi\left\{\frac{z_{1-\alpha}V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n) - n^{1/2}(p_B - p_A)}{V^{1/2}(p_A, p_B, \nu_n)}\right\} \\
&\approx 1 - \Phi\left\{z_{1-\alpha} - \frac{n^{1/2}(p_B - p_A)}{V^{1/2}(p_A, p_B, \nu_n)}\right\}, \qquad (1)
\end{aligned}$$

where Φ is the standard normal distribution function, and $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. Expression (1) is maximized when $V(p_A, p_B, \nu_n)$ is minimized, leading to Neyman allocation. This approximation is valid under contiguity conditions such as $n^{1/2}(p_B - p_A)/V^{1/2}(p_A, p_B, \nu_n) = O(1)$, i.e., when $p_B - p_A \approx n^{-1/2}$. However, for fixed $p_B - p_A$, the term $n^{1/2}(p_B - p_A)/V^{1/2}(p_A, p_B, \nu_n)$ is of order $n^{1/2}$, and the expression $\Phi\left\{z_{1-\alpha} - n^{1/2}(p_B - p_A)/V^{1/2}(p_A, p_B, \nu_n)\right\}$ is of asymptotic order that is smaller than the precision of the normal approximation. In this situation, a large deviations approximation is needed for calculating the power and optimal allocation.

For asymptotic power comparisons and evaluation of the relative asymptotic efficiency of certain tests, two different criteria are often used, related to the notions of Pitman and Bahadur efficiency (van der Vaart, 1998, Chapter 14). In our context, the Pitman approach looks at sequences of alternative probabilities $p_B^k > p_A^k$ that tend to a common limit at a suitable rate. The Bahadur approach considers a fixed alternative p_A and p_B, and approximates the power using large deviations theory.

We show in the next sections that the asymptotically optimal allocation ν^* corresponding to Pitman approach is always 1/2, while Bahadur optimal allocation depends on p_A and p_B and can be calculated in a way described below. Interestingly, computation of the Bahadur criterion for different values of p_A and p_B reveals that the optimal allocation is often close to 1/2, that is, a balanced design. Finite sample calculations lead to similar conclusions, that a balanced design is generally adequate with performance comparable to Neyman and Bahadur allocations. Balanced designs appear in a large body of literature; see, for example, Kalish & Harrington (1988) and Begg & Kalish (1984), where several estimation criteria and designs are considered, and balanced design is recommended as being close to optimal under various criteria.

A balanced design has the advantage that it can be implemented without any knowledge of the parameters of the problem. On the other hand, Neyman and Bahadur allocations require knowledge of the parameters p_A and p_B, and this is one of the reasons for conducting adaptive designs, in which information on these parameters is collected sequentially. There do exist other important reasons for adaptive designs, not discussed in this paper, such as the quality of treatment during the experiment; see, e.g., Rosenberger et al. (2001).

2 The Pitman approach

Pitman relative efficiency provides an asymptotic comparison of two families of tests applied to contiguous alternatives, that is, any sequence of alternatives satisfying $p_A - p_B \to 0$. Here we use the same idea to compare different allocation fractions.

In order to be specific, we fix a sequence of statistical problems indexed by k, and test $H_0 : p_A = p_B$ against the sequence of alternatives $p_A^k = p$, $p_B^k = p + k^{-1/2}$, for some $0 < p < 1$. The general case, where $k^{-1/2}$ is replaced by any positive sequence converging to zero, requires trivial modifications. Given an allocation sequence $\{\nu_n\}_{n=1}^{\infty}$, let $n_k = n_k(p, \alpha, \beta, \{\nu_n\})$ be the minimal number of observations required for a one-sided Wald test at significance level α and power at least β, for $\beta > \alpha$, at the point p_A^k, p_B^k respectively, where the observations are allocated according to $\{\nu_n\}$. Set $n_k = \infty$ if no finite number of observations satisfies these requirements. The next theorem implies that balanced allocation is asymptotically optimal.

Theorem 1. *Fix $\alpha < \beta$ and $0 < p < 1$. Let $\{\nu_n\}_{n=1}^{\infty}$ be any sequence of allocations and let $\{\tilde{\nu}_n\}_{n=1}^{\infty}$ be another sequence of allocations satisfying $\tilde{\nu}_n \to 1/2$ as $n \to \infty$. Then*

$$\liminf_{k\to\infty} \frac{n_k(p, \alpha, \beta, \{\nu_n\})}{n_k(p, \alpha, \beta, \{\tilde{\nu}_n\})} \geq 1.$$

The theorem follows readily from the following lemma.

Lemma 1. *I. If $\nu_n \to \nu$ as $n \to \infty$ for $0 < \nu < 1$ then*

$$\lim_{k\to\infty} \frac{n_k}{k} = (z_{1-\alpha} - z_{1-\beta})^2 \frac{p(1-p)}{\nu(1-\nu)}.$$

II. If $\nu_n \to 0$ or $\nu_n \to 1$ as $n \to \infty$ then $\lim_{k\to\infty} n_k/k = \infty$, and for any sequence of allocations $\{\nu_n\}$,

$$\liminf_{k\to\infty} \frac{n_k}{k} \geq (z_{1-\alpha} - z_{1-\beta})^2 \frac{p(1-p)}{1/4}.$$

Pitman optimality of balanced designs of Theorem 1 holds also for two-sided tests. In general, a homoscedastic model leads to a balanced allocation if a normal approximation is suitable. The binomial case is heteroscedastic since the variances of the estimators depend on p_A^k and p_B^k, but when these parameters converge to the same p, the limiting model is homoscedastic, and the limiting Neyman allocation is $1/2$ regardless of p. In heteroscedastic models concerning the Normal distribution or similar cases where the variance is not a function of the mean, Neyman allocation is Pitman optimal, and does not reduce to $1/2$.

It can be argued that rather than considering sequences of statistical problems as above, one should optimize for fixed p_A and p_B. The next section deals with this case.

3 The Bahadur approach

3.1 Wald test

In this section, large deviations theory is used to approximate the power of the Wald test for fixed p_A and p_B. This power increases exponentially to one with n at a rate that depends on the allocation fraction ν. The aim is to find the optimal limiting allocation fraction ν^* for which the rate is maximized. The results apply also to selection problems: if treatment B is better than A when $p_B > p_A$, and treatment B is selected if $\hat{p}_B > \hat{p}_A$, then the expression in (2) below with $K = 0$ represents the limit of the probability of incorrect selection.

We start with the following large deviations result. Recall that $\hat{p}_A$ and $\hat{p}_B$ depend on both n and an allocation ν_n.

Theorem 2. *Define*

$$H(t,\nu) = \nu \log(1 - p_A + p_A e^{t/\nu}) + (1-\nu)\log\{1 - p_B + p_B e^{-t/(1-\nu)}\} \text{ and } \mathrm{g}(\nu) = \inf_{\mathrm{t}>0} \mathrm{H}(\mathrm{t}, \nu).$$

I. If $p_B > p_A$ and $\nu_n \to \nu$ as $n \to \infty$, where $0 < \nu < 1$, then for any $K \geq 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \left[1 - \mathrm{P}\left\{\frac{n^{1/2}(\hat{p}_B - \hat{p}_A)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K\right\}\right] = g(\nu). \tag{2}$$

II. If $p_B \neq p_A$ and $\nu_n \to \nu$ as $n \to \infty$, where $0 < \nu < 1$, then for any $K > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \left[1 - \mathrm{P}\left\{\frac{n^{1/2}|\hat{p}_B - \hat{p}_A|}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K\right\}\right] = g(\nu). \tag{3}$$

III. If $\nu_n \to 0$ or 1 as $n \to \infty$ then (2) and (3) hold with $g(0) = g(1) = 0$.

Since $\hat{p}_B - \hat{p}_A$ is not an average of n independent identically distributed random variables, Theorem 2 does not follow directly from the Cramér–Chernoff theorem (van der Vaart, 1998, p. 205); proofs are given in the Appendix.

By Theorem 2, $\nu^* = \nu^*(p_A, p_B) = \arg\min_\nu g(\nu)$ is the asymptotically optimal allocation. It is easy to prove directly that g is strictly convex, and the minimum is attained uniquely. More generally, it is readily shown by differentiation that if $M(t) = E(e^{tX})$ is a moment generating function, then $\nu \log M(t/\nu)$ and therefore H are convex in ν.

Theorem 3 below shows that the finite-sample optimal allocations in the one and two-sample tests converge to ν^*, and therefore it is reasonable to use ν^* as an approximation to the finite-sample optimal design for sufficiently large samples.

For each fixed n, let $\nu_n^{*(1)} = \nu_n^{*(1)}(p_A, p_B, K)$ be the allocation that maximizes the power of the one-sided test for a total sample size of n subjects, i.e,

$$\nu_n^{*(1)} = \arg\max_{\nu_n \in \{1/n, \ldots, (n-1)/n\}} \mathrm{P}\left\{\frac{n^{1/2}(\hat{p}_B - \hat{p}_A)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K\right\};$$

similarly, $\nu_n^{*(2)} = \nu_n^{*(2)}(p_A, p_B, K)$ is the optimal allocation of the two-sided test.

Theorem 3. *I. If $p_B > p_A$ then for any $K \geq 0$, $\nu_n^{*(1)} \to \nu^*$ as $n \to \infty$.*

II. If $p_B \neq p_A$ then for any $K > 0$, $\nu_n^{*(2)} \to \nu^*$ as $n \to \infty$.

Remark 1. *Another formulation of these results, for the one-sided case, say, is the following: if $p_B > p_A$ then for any sequence ν_n and constant $K \geq 0$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log\left[1 - \mathrm{P}\left\{\frac{n^{1/2}(\hat{p}_B - \hat{p}_A)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K\right\}\right] \geq g(\nu^*),$$

with equality if and only if $\nu_n \to \nu^*$ as $n \to \infty$.

3.2 Other tests

The power approximations of Section 3.1 apply to tests based on statistics of the form $n^{1/2}(\hat{p}_B - \hat{p}_A)/V_n$, where V_n is a bounded random variable that converges almost surely to a positive constant. The chi-square test for equality of proportions and the score test are of this kind. Moreover, consider a test statistic of the form $n^{1/2}s(\hat{p}_B, \hat{p}_A)/V_s^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)$ that rejects $p_A = p_B$ for large values, where $s(\hat{p}_B, \hat{p}_A)$ is a statistic that measures the discrepancy between the proportions and V_s/n is an estimator of its variance under H_0. Practical examples to which Theorem 4 below applies arise when $s(x, y)$ has the form $f(x) - f(y)$ where $f : (0,1) \mapsto \mathbb{R}$ is an increasing function. The choices $f(x) = \log(x)$ and $f(x) = \mathrm{logit(x)}$ yield tests based on the relative risk and odds ratio, and $f(x) = \arcsin(x^{1/2})$ is sometimes used as a variance stabilizing transformation. Next we show that the results of Theorems 2 and 3 hold for such statistics under certain conditions, with the same $g(\nu)$ and ν^* as above.

Theorem 4. *Assume that $s(x,y)$ satisfies the following three conditions: (i) for all (x,y), $s(x,y)$ and $x-y$ have the same sign; (ii) for all (x,y) there exists $C_0>0$ such that*

$$C_0|x-y| \le |s(x,y)|; \tag{4}$$

(iii) for any $\varepsilon>0$ there exists $C_1(\varepsilon)>0$ such that for all $(x,y)\in G_\varepsilon$

$$|x-y| \ge C_1(\varepsilon)|s(x,y)|, \tag{5}$$

where $G_\varepsilon=\{(x,y):\min(x,y)>\varepsilon \text{ and } \max(\mathrm{x,y})<1-\varepsilon\}$.

Assume that for any $\varepsilon>0$ there exist $C_2(\varepsilon), C_3(\varepsilon)>0$, such that for all n and all $(x,y)\in D_\varepsilon$,

$$C_2(\varepsilon) \le {V_s}^{1/2}(x,y,\nu_n) \le C_3(\varepsilon), \tag{6}$$

where $D_\varepsilon=\{(x,y):\max(x,y)>\varepsilon \text{ and } \min(\mathrm{x,y})<1-\varepsilon\}$.

Let $\nu_n\to\nu$ as $n\to\infty$, where $0<\nu<1$ and $K>0$. Then for $p_B>p_A$,

$$\lim_{n\to\infty}\frac{1}{n}\log\left[1-\mathrm{P}\left\{\frac{n^{1/2}s(\hat p_B,\hat p_A)}{V_s^{1/2}(\hat p_A,\hat p_B,\nu_n)}>K\right\}\right]=g(\nu), \tag{7}$$

and for $p_B\neq p_A$,

$$\lim_{n\to\infty}\frac{1}{n}\log\left[1-\mathrm{P}\left\{\frac{n^{1/2}|s(\hat p_A,\hat p_B)|}{V_s^{1/2}(\hat p_A,\hat p_B,\nu_n)}>K\right\}\right]=g(\nu). \tag{8}$$

Conditions *(i)-(iii)* on $s(x,y)$ hold in the examples preceding Theorem 4. For the case where $s(x,y)=\log(x)-\log(y)$, Conditions (4) and (5) are satisfied with $C_0=1$ and $C_1(\varepsilon)=\varepsilon$. Condition (6) is more involved. It holds, for this case when $V_s(\hat p_A,\hat p_B,\nu_n)=(1-\hat p)/\{\nu_n(1-\nu_n)\hat p\}$, where $\hat p=\nu_n\hat p_A+(1-\nu_n)\hat p_B$ is the pooled estimator. However, Condition (6) can fail when using an unpooled-type estimator, and then $g(\nu)$ is not the correct rate for certain values of p_A,p_B,ν. The logit case can be verified in the same way; the arcsin case is easier since the variance estimator is constant.

4 Numerical illustration

Somewhat tedious calculations show that the optimal allocation is

$$\nu^*=\log\left\{\frac{p_B\log(\frac{p_B}{p_A})}{(1-p_B)\log(\frac{1-p_A}{1-p_B})}\right\}\Big/\log\left\{\frac{p_B(1-p_A)}{p_A(1-p_B)}\right\}. \tag{9}$$

Table 1: Comparison of Bahadur and Neyman allocations for a hypothesis testing problem.

p_A	p_B	Bahadur	Neyman
0·5	0·65	0·504	0·512
0·5	0·8	0·518	0·556
0·5	0·9	0·542	0·625
0·7	0·75	0·505	0·514
0·7	0·85	0·521	0·562
0·7	0·9	0·535	0·604
0·85	0·95	0·541	0·621

Table 1 compares the asymptotic Bahadur optimal allocation, ν^*, and Neyman allocation for several pairs (p_A, p_B). These and further systematic numerical calculations indicate that Bahadur allocation is closer to 1/2 than Neyman allocation and that it is quite close to 1/2 unless p_A or p_B are extreme, e.g., $p_A = 0{\cdot}85$, $p_B = 0{\cdot}95$.

Figure 1 compares the exact power under Bahadur, balanced, and Neyman allocations for the two-sided Wald test with critical value $K = 2$ corresponding to $\alpha \approx 0{\cdot}05$. For large power Bahadur allocation is better for almost all parameters, and balanced allocation is usually better than Neyman. For power $\approx 0{\cdot}75$ the picture is more ambiguous; it seems that in most cases Neyman allocation outperforms a balanced design but the differences are usually minor unless p_A is large. In the latter case, Bahadur allocation is comparable to Neyman allocation for moderate power and better for large power. Overall, the differences in power among the three allocations are relatively small. These findings justify the use of balanced allocation rather than more complex designs unless the probabilities are expected to be extreme, and then Bahadur allocation should be considered. Recall that implementation of Bahadur, as well as Neyman allocation, requires some prior knowledge of the parameters, or adaptive allocation.

5 A dose finding problem

Dose finding studies are conducted as part of phase I clinical trials in order to find the maximal tolerated dose among a finite, usually very small, number of potential doses. The maximal tolerated dose is defined as the dose with the probability of toxic reaction closest to a pre-specified probability p_0. Recently, we showed that under certain natural assumptions, in order to estimate the desired dose consistently, one can consider experiments that eventually concentrate on two doses (Azriel et al., 2011). Thus, asymptotically, the allocation problem in these studies reduces to the problem of finding which of two probabilities of toxic reaction $p_B > p_A$, corresponding to the doses $d_B > d_A$, is closer to p_0.

Let $\hat{p}_A$ and $\hat{p}_B$ denote the proportions of subjects that experienced toxic reactions in doses d_A and d_B based on a sample of size n and an allocation ν_n. For large n, $\hat{p}_B > \hat{p}_A$, and a natural estimator for the maximal tolerated dose is $\hat{D} = d_A$ if $(\hat{p}_A + \hat{p}_B)/2 > p_0$ and $\hat{D} = d_B$ otherwise. Similar to the problems discussed in previous sections, an optimal design is an allocation rule of $n\,\nu_n$ and $n(1-\nu_n)$ individuals to doses d_A and d_B, respectively, such that $\mathrm{P}(\hat{D} = d_A) = \mathrm{P}\{(\hat{p}_A + \hat{p}_B)/2 > p_0\}$ is maximized if d_A is indeed the maximal tolerated dose.

For the current problem, the Pitman approach is translated to a comparison of designs under sequences of parameters p_A^k, p_B^k and p_0^k such that $|(p_A^k + p_B^k)/2 - p_0^k| \to 0$ and $p_A^k \to p_A$, $p_B^k \to p_B$; for convenience we specify $|(p_A^k + p_B^k)/2 - p_0^k| = k^{-1/2}$. Let $0 < \nu < 1$ and let $n_k = n_k(p_A^k, p_B^k, p_0^k, \beta, \{\nu_n\})$ be the minimal number of observations required such that the probability of correct estimation of the maximal tolerated dose is larger than β for the given parameters when the allocation for dose d_A is $n\,\nu_n$. As in Lemma 1, it can be shown that if $\nu_n \to \nu$ as $n \to \infty$ then

$$\lim_{k\to\infty} \frac{n_k}{k} = \frac{z_\beta^2}{4} \left\{ \frac{p_A(1-p_A)}{\nu} + \frac{p_B(1-p_B)}{1-\nu} \right\}.$$

Thus, the asymptotically optimal design uses Neyman allocation, as it minimizes the limit of n_k/k. Unlike the previous problem, now p_A^k and p_B^k do not converge to the same value under the Pitman approach as defined here, and Neyman allocation does not reduce to a balanced design.

For the case of fixed p_A, p_B, and p_0, assume that p_B is nearer than p_A to p_0, and consider the problem of minimizing the probability of selecting d_A. The following theorem, analogous to Theorems 2 and 3, gives the asymptotic optimal allocation rule in the current setting.

Table 2: Comparison of Bahadur and Neyman allocations for a dose finding problem.

p_A	p_B	p_0	Bahadur	Neyman
0·10	0·30	0·28	0·420	0·396
0·10	0·40	0·26	0·384	0·380
0·10	0·40	0·30	0·400	0·380
0·10	0·40	0·35	0·417	0·380
0·20	0·35	0·30	0·460	0·456
0·20	0·40	0·33	0·455	0·449
0·22	0·33	0·30	0·471	0·468
0·25	0·35	0·33	0·479	0·476

Theorem 5. *If $\nu_n \to \nu$ as $n \to \infty$, and $0 < \nu < 1$, then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\{(\hat{p}_A + \hat{p}_B)/2 \geq p_0\} = \psi(\nu),$$

where $\psi(\nu) = \inf_t\{\nu \log(1 - p_A + p_A e^{t/\nu}) + (1-\nu)\log(1 - p_B + p_B e^{t/(1-\nu)}) - 2p_0 t\}$.

Moreover, let $\nu^* = \arg\min \psi(\nu)$, and let ν_n^* be the value of the allocation minimizing $\mathrm{P}\{(\hat{p}_A + \hat{p}_B)/2 \geq p_0\}$ for a given n. Then, $\nu_n^* \to \nu^*$ as $n \to \infty$.

We calculated ν^* for several values of p_A, p_B, and p_0. We found that ν^* is often close to Neyman allocation, see Table 2. Both Bahadur and Pitman criteria yield quite similar results in this problem. Allocating subjects according to either improves the probability of correct estimation compared to balanced allocation for very large samples, as the optimal allocations according to Bahadur or Pitman are far from 1/2. Calculations not presented here show that for practical sample sizes for this problem, all three methods differ negligibly.

6 A general response

In previous sections, we dealt with the very important, though specific, case of a binary response. In this section, we consider the more general case where the response of an individual treated in group $i = A, B$ follows a distribution F_i having moment generating function $M_i(t)$, and find the optimal allocation according to the Bahadur approach. Let $\bar{Y}_i(m)$ denote the average of m responses of subjects having treatment i. Treatment B is considered better if $\int xF_B(dx) > \int xF_A(dx)$, and assume that the treatment with the largest average response is declared better. The following theorem, which can be proved in a similar way as Theorems 2 and 3, provides the Bahadur optimal allocation rule for correct selection.

Theorem 6. *Assume $\int xF_B(dx) > \int xF_A(dx)$, and $\nu_n \to \nu$ as $n \to \infty$ for $0 < \nu < 1$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left[\bar{Y}_A(n\,\nu_n) \geq \bar{Y}_B\{n(1-\nu_n)\}\right] = h(\nu),$$

where

$$h(\nu) = \inf_{t>0} \big(\nu \log\{M_A(t/\nu)\} + (1-\nu)\log[M_B\{-t/(1-\nu)\}]\big). \tag{10}$$

Moreover, let ν_n^* be the value of the allocation minimizing $\mathrm{P}\left[\bar{Y}_A(n\,\nu_n) \geq \bar{Y}_B\{n(1-\nu_n)\}\right]$*, and* $\nu^* = \arg\min_\nu h(\nu)$*. Then $\nu_n^* \to \nu^*$ as $n \to \infty$.*

When the responses in the two treatments are normally distributed, the Bahadur allocation agrees with Neyman allocation. This can be easily verified by using the moment generating functions of Normal variables in (10). However, for other distributions, the allocations may differ considerably. Table 3 compares the Bahadur and Neyman allocations for different Poisson and Gamma distributions. As in the Binomial case, the Bahadur allocation is closer to 1/2 than to Neyman allocation. Further study is required to determine if the improvement over balanced allocation, in terms of power or probability of correct selection, is significant. Anyway, optimality of Neyman allocation for non-normal distributions should be questioned, and may hold only under restrictive conditions.

Acknowledgment

We thank Amir Dembo for helpful comments and in particular for deriving the exact expression for ν^* in (9). We also thank an Associate editor and two referees for very insightful comments.

Table 3: Comparison of Bahadur and Neyman allocations for maximizing the probability of correct selection for different distributions.

F_A	F_B	Bahadur	Neyman
Poisson(1)	Poisson(2)	0·471	0·414
Poisson(2)	Poisson(3)	0·483	0·449
Poisson(3)	Poisson(4)	0·488	0·464
Poisson(4)	Poisson(5)	0·491	0·472
Gamma(0·5,0·5)	Gamma(0·5,0·6)	0·515	0·590
Gamma(0·5,0·5)	Gamma(0·5,0·7)	0·528	0·662
Gamma(0·5,0·5)	Gamma(0·5,0·8)	0·539	0·719
Gamma(0·5,0·5)	Gamma(0·5,0·9)	0·549	0·764

Appendix

Proofs

Proof of Lemma 1 part *I*. The proof uses arguments as in Theorem 14.19 in van der Vaart (1998), which is stated in terms of relative efficiency rather than allocation.

First note that $\lim_{k\to\infty} n_k = \infty$; otherwise, there exists a bounded subsequence of n_k on which the power converges to a value $\leq \alpha$, since as $k \to \infty$ we have $p_A^k - p_B^k \to 0$. This contradicts the definition of n_k and the assumption that $\alpha < \beta$.

By the Berry–Esseen theorem

$$\frac{n_k^{1/2}(\hat{p}_A^k - p)}{\{\frac{p(1-p)}{\nu_{n_k}}\}^{1/2}} \to N(0,1) \ \text{ and } \ \frac{n_k^{1/2}\{\hat{p}_B^k - (p + k^{-1/2})\}}{\left[\frac{(p+k^{-1/2})\{1-(p+k^{-1/2})\}}{1-\nu_{n_k}}\right]^{1/2}} \to N(0,1)$$

in distribution as $k \to \infty$, since the third moment is bounded. Here $\hat{p}_A^k = \hat{p}_A(\nu_{n_k} n_k) = Y_A^k(\nu_{n_k} n_k)/(\nu_{n_k} n_k)$, where $Y_A^k(m) \sim Bin(m, p_A^k)$ is the sum of m independent binary responses with probability p_A^k, and $\hat{p}_B^k$ is defined analogously.

Now, if $\lim \nu_{n_k} = \nu$ as $k \to \infty$ we have

$$U_k = \frac{{n_k}^{1/2}(\hat{p}_B^k - \hat{p}_A^k) - (\frac{n_k}{k})^{1/2}}{\left\{\frac{p(1-p)}{\nu(1-\nu)}\right\}^{1/2}} \rightarrow N(0,1) \tag{11}$$

in distribution as $k \to \infty$. Since $\lim_{k\to\infty} n_k = \infty$, the critical value for the level α one-sided Wald test is $z_{1-\alpha} + o(1)$; then

$$\mathrm{P}_{p_A^k, p_B^k}\{W > z_{1-\alpha} + o(1)\} = \mathrm{P}\left\{\frac{n_k^{1/2}(\hat{p}_B^k - \hat{p}_A^k)}{V^{1/2}(\hat{p}_A^k, \hat{p}_B^k, \nu_{n_k})} > z_{1-\alpha} + o(1)\right\}$$

$$= \quad \mathrm{P}\left[U_k > \{z_{1-\alpha} + o(1)\}\left\{\frac{V(\hat{p}_A^k, \hat{p}_B^k, \nu_{n_k})}{\frac{p(1-p)}{\nu(1-\nu)}}\right\}^{1/2} - \left\{\frac{n_k/k}{\frac{p(1-p)}{\nu(1-\nu)}}\right\}^{1/2}\right].$$

Because,

$$\frac{V(\hat{p}_A^k, \hat{p}_B^k, \nu_{n_k})}{\frac{p(1-p)}{\nu(1-\nu)}} \rightarrow 1$$

in probability as $k \to \infty$, and since the limiting power is exactly β we have due to (11)

$$z_{1-\alpha} - \left\{\frac{\lim_{k\to\infty} n_k/k}{\frac{p(1-p)}{\nu(1-\nu)}}\right\}^{1/2} = z_{1-\beta};$$

hence part I holds.

Proof of part *II*. We only prove the case $\nu_n \to 0$, as $\nu_n \to 1$ is similar. If $n\nu_n$ is bounded, then the power converges to α and $n_k = \infty$ for large k.

Assume now that $n\,\nu_n \to \infty$ as $n \to \infty$; by the Berry–Esseen theorem and Slutsky's Lemma we have

$$(n_k\,\nu_{n_k})^{1/2}(\hat{p}_A^k - p) \rightarrow N\{0, p(1-p)\} \;\text{ and }\; (n_k\,\nu_{n_k})^{1/2}\{\hat{p}_B^k - (p + k^{-1/2})\} \rightarrow 0$$

in distribution as $k \to \infty$. This implies that

$$(n_k\,\nu_{n_k})^{1/2}(\hat{p}_B^k - \hat{p}_A^k) - \left(\frac{n_k\,\nu_{n_k}}{k}\right)^{1/2} \rightarrow N\{0, p(1-p)\}$$

in distribution as $k \to \infty$, and by arguments as in the first part we have

$$\lim_{k\to\infty} \frac{n_k\,\nu_{n_k}}{k} = (z_{1-\alpha} - z_{1-\beta})^2\, p(1-p).$$

Because $\nu_{n_k} \to 0$, $\lim_{k\to\infty} \frac{n_k}{k} = \infty$.

For the second part of *II* notice that there exists a subsequence $\{k'\}$ such that $\lim_{k'\to\infty} \nu_{n_{k'}} = \nu'$ for some ν' and

$$\liminf_{k\to\infty} \frac{n_k}{k} = \lim_{k'\to\infty} \frac{n_{k'}}{k'} = (z_{1-\alpha} - z_{1-\beta})^2 \frac{p(1-p)}{\nu'(1-\nu')},$$

where the second equality follows by part *I*. If $\nu'(1-\nu') = 0$ we interpret the limit as ∞; since $\nu'(1-\nu') \leq \frac{1}{4}$, the second part of *II* follows. $\square$

Proof of Theorem 2 parts I and II. The proof follows known large deviations ideas, but certain variations are needed for the present non-standard case. Notice that the probability in (2) is larger than the probability of and (3). Therefore, it is enough to show that for any $K \geq 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log \left[1 - \mathrm{P}\left\{ \frac{n^{1/2}(\hat{p}_B - \hat{p}_A)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K \right\} \right] \leq g(\nu), \tag{12}$$

and for any $K > 0$

$$\liminf_{n\to\infty} \frac{1}{n} \log \left[1 - \mathrm{P}\left\{ \frac{n^{1/2}|\hat{p}_B - \hat{p}_A|}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K \right\} \right] \geq g(\nu).$$

Instead of the latter inequality we prove a stronger result, namely

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left\{ 0 \leq \frac{n^{1/2}(\hat{p}_A - \hat{p}_B)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} \leq K' \right\} \geq g(\nu), \tag{13}$$

for all $K' > 0$, which is also used to show that $g(\nu)$ is a lower bound for the case of $K = 0$ in (2), when $K' = \infty$.

Starting with (12), since $V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)$ is bounded, for any $\varepsilon > 0$ and large enough n,

$$1 - \mathrm{P}\left\{ \frac{n^{1/2}(\hat{p}_B - \hat{p}_A)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} > K \right\} = \mathrm{P}\left\{ \hat{p}_A - \hat{p}_B \geq -n^{-1/2} K V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n) \right\} \leq \mathrm{P}\left(\hat{p}_A - \hat{p}_B \geq -\varepsilon \right).$$

Standard large deviations arguments for the upper bound (van der Vaart, 1998, p. 205) yield

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left\{ \hat{p}_A - \hat{p}_B \geq -n^{-1/2} K V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n) \right\} \leq g_\varepsilon(\nu),$$

where $g_\varepsilon(\nu) = \inf_{t>0}\{\varepsilon t + H(t, \nu)\}$. As this is true for any $\varepsilon > 0$, and by the continuity of $g_\varepsilon(\nu)$ in ε (12) is verified.

To prove (13), assume without loss of generality that $p_B > p_A$; define

$$T_n = \hat{p}_A(n\,\nu_n) - \hat{p}_B\{n(1-\nu_n)\} = \frac{Y_A(n\,\nu_n)}{n\,\nu_n} - \frac{Y_B\{n(1-\nu_n)\}}{n(1-\nu_n)}.$$

The cumulant generating function of T_n is

$$\log E(e^{tT_n}) = n\,\nu_n \log\{1-p_A+p_A e^{t/(n\,\nu_n)}\}+n(1-\nu_n)\log[1-p_B+p_B e^{-t/\{n(1-\nu_n)\}}] = nH(\frac{t}{n},\nu_n). \quad (14)$$

We have $\partial/\partial t H(0,\nu_n) = E(T_n) = p_A - p_B < 0$, by (14). Also, $H(0,\nu_n) = 0$ and $H(\cdot,\nu_n)$ is strictly convex being the log of a moment generating function, up to a constant. Since $\mathrm{P}(T_n > 0) > 0$ it follows that $H(t,\nu_n) \to \infty$ as $t \to \infty$ and therefore, $\arg\min_{t>0} H(t,\nu_n) = t_0^{(n)}$ is a unique interior point and $\partial H(t_0^{(n)},\nu_n)/\partial t = 0$. Let t_0 be the minimizer of $H(\cdot,\nu)$; we show that $t_0^{(n)} \to t_0$. If there is a subsequence $\{t_0^{(n_k)}\}$ that converges to $t_1 \leq \infty$ then $H\{t_0^{(n_k)},\nu_{n_k}\} \leq H(t_0,\nu_{n_k})$ as $t_0^{(n_k)}$ is the minimizer, and continuity of H implies $H(t_1,\nu) \leq H(t_0,\nu)$. Now $t_1 = t_0$ by uniqueness of the minimizer.

Let Z_n be the Cramér transform of T_n, that is, $\mathrm{P}(Z_n = z) = e^{-ng(\nu_n)}e^{zt_0^{(n)}n}\mathrm{P}(T_n = z)$. Now,

$$\begin{aligned}
& \mathrm{P}\left\{0 \leq \frac{n^{1/2}(\hat{p}_A - \hat{p}_B)}{V^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)} \leq K\right\} \\
= \; & \mathrm{P}\left\{0 \leq T_n \leq n^{-1/2}KV^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\right\} \\
= \; & E\left[I\{0 \leq Z_n \leq n^{-1/2}KV^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\}e^{-Z_n t_0^{(n)} n}\right] e^{ng(\nu_n)} \\
\geq \; & \mathrm{P}\left\{0 \leq Z_n \leq n^{-1/2}KV^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\right\} e^{-n^{1/2}K1/2\{1/\nu_n+1/(1-\nu_n)\}^{1/2}t_0^{(n)}} e^{ng(\nu_n)},
\end{aligned}$$

where the last inequality uses the upper bound on Z_n in the indicator function, and the upper bound on the variance obtained by replacing $\hat{p}_i(1-\hat{p}_i)$ by $1/4$ for $i = A, B$. It follows that

$$\begin{aligned}
& g(\nu_n) - \frac{1}{n}\log \mathrm{P}\left\{0 \leq \frac{n^{1/2}(\hat{p}_A - \hat{p}_B)}{V^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)} \leq K\right\} \\
\leq \; & \frac{-\log \mathrm{P}\left\{0 \leq Z_n \leq n^{-1/2}KV^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\right\}}{n} + \frac{K\frac{1}{2}(\frac{1}{\nu_n}+\frac{1}{1-\nu_n})^{1/2}t_0^{(n)}}{n^{1/2}}.
\end{aligned}$$

As $t_0^{(n)} \to t_0 < \infty$, and $\nu_n \to \nu \in (0,1)$, the second term on the right-hand side vanishes as $n \to \infty$; for the first, we claim that $n^{1/2}Z_n$ is asymptotically $N(0,\sigma^2)$, where $\sigma^2 = \partial^2 H(t_0,\nu)/\partial t^2$. Consequently $\mathrm{P}\left\{0 \leq Z_n \leq n^{-1/2}KV^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\right\} \to C$ for some constant $C > 0$, and (13) follows.

It remains to prove the asymptotic normality of $n^{1/2}Z_n$, which we do by considering its moment generating function. We have

$$\log E(e^{sn^{1/2}Z_n}) = -ng(\nu_n) + \log E\left\{e^{T_n(sn^{1/2}+t_0^{(n)}n)}\right\} = n\left\{-H(t_0^{(n)},\nu_n) + H(t_0^{(n)}+sn^{-1/2},\nu_n)\right\},$$

where the last equality follows from (14) and the identity $g(\nu_n) = H(t_0^{(n)}, \nu_n)$. By Taylor expansion of $H(\cdot, \nu_n)$ around $t_0^{(n)}$ we obtain

$$H(t_0^{(n)} + sn^{-1/2}, \nu_n) - H(t_0^{(n)}, \nu_n) = \frac{1}{2}\frac{s^2}{n}\frac{\partial^2}{\partial t^2}H(t_0^{(n)}, \nu_n) + O(n^{-3/2})$$

since the first derivative is 0, and therefore, as $n \to \infty$,

$$\log E\left(e^{sn^{1/2}Z_n}\right) \to \frac{s^2}{2}\frac{\partial^2}{\partial t^2}H(t_0, \nu) = \frac{s^2\sigma^2}{2};$$

thus $n^{1/2}Z_n$ is asymptotically $N(0, \sigma^2)$, and (13) is established.

Proof of part III. First note that (12) clearly holds with $g(\nu) = 0$ as $\log\{1 - \mathrm{P}(\cdot)\} \leq 0$, so it remains to prove (13) for $g(\nu) = 0$, that is, for any $K > 0$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left\{0 \leq \frac{n^{1/2}(\hat{p}_A - \hat{p}_B)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} \leq K\right\} \geq 0.$$

We only prove the case $\nu_n \to 0$, as $\nu_n \to 1$ is similar. If $n\,\nu_n \not\to \infty$ then $\hat{p}_A$ is inconsistent and the limit is easily seen to be zero. Assume now that $n\,\nu_n \to \infty$; since $V(\hat{p}_A, \hat{p}_B, \nu_n) \geq \hat{p}_A(1 - \hat{p}_A)/\nu_n$ we have

$$\mathrm{P}\left[0 \leq \frac{n^{1/2}(\hat{p}_A - \hat{p}_B)}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n)} \leq K\right] \geq \mathrm{P}\left[0 \leq \hat{p}_A - \hat{p}_B \leq \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2}}{(n\,\nu_n)^{1/2}}\right].$$

Now, for $\varepsilon = K\{p_A(1 - p_A)\}^{1/2}/2$,

$$\begin{aligned}
& \mathrm{P}\left[0 \leq \hat{p}_A - \hat{p}_B \leq \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2}}{(n\,\nu_n)^{1/2}}\right] \\
\geq\ & \mathrm{P}\left[\{0 \leq \hat{p}_A - \hat{p}_B \leq \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2}}{(n\,\nu_n)^{1/2}}\} \cap \{\hat{p}_B \in (p_B - \frac{\varepsilon}{(n\,\nu_n)^{1/2}}, p_B)\}\right] \\
\geq\ & \mathrm{P}\left[p_B \leq \hat{p}_A \leq p_B + \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2} - \varepsilon}{(n\,\nu_n)^{1/2}}\right] \mathrm{P}\left[\hat{p}_B \in (p_B - \frac{\varepsilon}{(n\,\nu_n)^{1/2}}, p_B)\right].
\end{aligned}$$

Taking logs and limits as $n \to \infty$ in the above product, we have to consider two parts. For the first, we have by Lemma 2 below

$$\lim_{n\to\infty} \frac{1}{n\,\nu_n} \log \mathrm{P}\left[p_B \leq \hat{p}_A \leq p_B + \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2} - \varepsilon}{(n\,\nu_n)^{1/2}}\right] = C,$$

for some constant C, and since $\nu_n \to 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left[p_B \leq \hat{p}_A \leq p_B + \frac{K\{\hat{p}_A(1 - \hat{p}_A)\}^{1/2} - \varepsilon}{(n\,\nu_n)^{1/2}}\right] = 0.$$

The limit of the log of the second part divided by n is 0, since by the Central Limit Theorem

$$\mathrm{P}\left\{\hat{p}_B \in (p_B - \frac{\varepsilon}{(n\,\nu_n)^{1/2}}, p_B)\right\} \geq \mathrm{P}\left[-\varepsilon \leq \{n(1-\nu_n)\}^{1/2}(\hat{p}_B - p_B) \leq 0\right] \to C' > 0.$$

□

Lemma 2. *Let $V_1, V_2, \ldots$ be independent identically distributed with $E(V_1) < 0$ and moment generation function $M(t)$, and let X_n be uniformly bounded random variables that satisfy $X_n \to K$ almost surely for a constant $K > 0$; then for $\bar{V}_n = \sum_{i=1}^n V_i/n$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\left(0 \leq \bar{V}_n \leq \frac{X_n}{n^{1/2}}\right) = \inf_{t>0} \{\log M(t)\}\,.$$

Proof of Lemma 2. The lemma follows by the same argument as in van der Vaart (1998), p. 206, replacing ε in that proof by $\tilde{K}n^{-1/2}$, where $\tilde{K}$ is the upper bound of X_n; see also the proof of parts I and II of Theorem 2, where a similar argument is used. □

Proof of Theorem 3. We will prove part I; the proof of Part II is similar. Set $\nu_n' = \lfloor n\,\nu^* \rfloor/n$, where $\lfloor\cdot\rfloor$ is the floor function; we have $\nu_n' \to \nu^*$ as $n \to \infty$. Theorem 2 implies that

$$\lim_{n\to\infty} \frac{1}{n} \log\left\{1 - \mathrm{P}\left(\frac{n^{1/2}[\hat{p}_B\{n(1-\nu_n')\} - \hat{p}_A(n\,\nu_n')]}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_n')} > K\right)\right\} = g(\nu^*). \tag{15}$$

If there exists a subsequence $\{n_k\}$ such that $lim_{k\to\infty}\nu_{n_k}^{*(1)} = \tilde{\nu} \neq \nu^*$ then

$$\lim_{k\to\infty} \frac{1}{n_k} \log\left\{1 - \mathrm{P}\left(\frac{n_k^{1/2}[\hat{p}_B\{n_k(1-\nu_{n_k}^{*(1)})\} - \hat{p}_A(n_k\,\nu_{n_k}^{*(1)})]}{V^{1/2}(\hat{p}_A, \hat{p}_B, \nu_{n_k}^{*(1)})} > K\right)\right\} = g(\tilde{\nu}) > g(\nu^*). \tag{16}$$

It follows that for large enough n_k the probability appearing in (16) is smaller than that of (15), in contradiction to the definition of $\nu_{n_k}^{*(1)}$ as the maximizer of this probability for $n = n_k$. □

Proof of Theorem 4. Assume without loss of generality $p_B > p_A$. Using $\mathrm{P}(\hat{p}_A - \hat{p}_B \geq 0) \geq \mathrm{P}(\hat{p}_B = 0) = (1-p_B)^{(1-\nu_n)n}$ we obtain by (2) with $K = 0$, $g(\nu) = \lim_{n\to\infty} n^{-1} \log \mathrm{P}(\hat{p}_A - \hat{p}_B \geq 0) \geq (1-\nu)\log(1-p_B)$. Similarly $\mathrm{P}(\hat{p}_A - \hat{p}_B \geq 0) \geq \mathrm{P}(\hat{p}_A = 1)$ yields $g(\nu) \geq \nu \log(p_A)$. A standard calculation shows that as ε decreases to 0, the rates of the probabilities $\mathrm{P}(\hat{p}_B \leq \varepsilon)$ and $\mathrm{P}(\hat{p}_A \geq 1-\varepsilon)$ are $(1-\nu)\log(1-p_B)$ and $\nu\log(p_A)$, respectively, and as just noted both are bounded by $g(\nu)$. For $\varepsilon < 1/2$,

$$\begin{aligned}\mathrm{P}\{(\hat{p}_A, \hat{p}_B) \notin D_\varepsilon\} &= \mathrm{P}(\hat{p}_A \leq \varepsilon)\mathrm{P}(\hat{p}_B \leq \varepsilon) + \mathrm{P}(\hat{p}_A \geq 1-\varepsilon)\mathrm{P}(\hat{p}_B \geq 1-\varepsilon) \\ &\leq 2\max\{\mathrm{P}(\hat{p}_B \leq \varepsilon), \mathrm{P}(\hat{p}_A \geq 1-\varepsilon)\}\max\{\mathrm{P}(\hat{p}_A \leq \varepsilon), \mathrm{P}(\hat{p}_B \geq 1-\varepsilon)\}.\end{aligned}$$

Therefore, by the above discussion of the rates of expressions of this type,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log \mathrm{P}\{(\hat{p}_A,\hat{p}_B)\notin D_\varepsilon\} \le g(\nu)+\max\{\nu\log(1-p_A),(1-\nu)\log(p_B)\} < g(\nu).$$

Hence, there exists small enough ε such that for $D = D_\varepsilon$,

$$\lim_{n\to\infty}\frac{1}{n}\log \mathrm{P}\{(\hat{p}_A,\hat{p}_B)\notin D\} < g(\nu). \tag{17}$$

As in the proof of Theorem 2, it suffices to prove that $g(\nu)$ is an upper bound for the limit in (7) and a lower bound for the limit in (8). For the upper bound we have

$$\begin{aligned}
&1-\mathrm{P}\left\{\frac{n^{1/2}s(\hat{p}_B,\hat{p}_A)}{V_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)} > K\right\} = \mathrm{P}\left\{s(\hat{p}_B,\hat{p}_A)\le n^{-1/2}KV_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\right\}\\
&= \mathrm{P}\left[\{s(\hat{p}_B,\hat{p}_A)\le n^{-1/2}KV_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\}\cap\{(\hat{p}_A,\hat{p}_B)\in D\}\right]\\
&+ \mathrm{P}\left[\{s(\hat{p}_B,\hat{p}_A)\le n^{-1/2}KV_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\}\cap\{(\hat{p}_A,\hat{p}_B)\notin D\}\right]\\
&\le \mathrm{P}\{s(\hat{p}_B,\hat{p}_A)\le n^{-1/2}KC_3(\varepsilon)\} + \mathrm{P}\{(\hat{p}_A,\hat{p}_B)\notin D\}\\
&\le \mathrm{P}\left\{\hat{p}_B-\hat{p}_A\le n^{-1/2}KC_3(\varepsilon)/C_0\right\} + \mathrm{P}\{(\hat{p}_A,\hat{p}_B)\notin D\};
\end{aligned}$$

the first inequality follows from (6) and the second from (4). By (17), the second summand decreases exponentially faster than the first, which in turn has the rate $g(\nu)$ by (2), and the upper bound is established.

For the lower bound, we have for large enough n

$$\begin{aligned}
&1-\mathrm{P}\left\{\frac{n^{1/2}|s(\hat{p}_A,\hat{p}_B)|}{V_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)} > K\right\} \ge \mathrm{P}\left[\{|s(\hat{p}_A,\hat{p}_B)|\le n^{-1/2}KV_s^{1/2}(\hat{p}_A,\hat{p}_B,\nu_n)\}\cap\{(\hat{p}_A,\hat{p}_B)\in D\}\right]\\
&\ge \mathrm{P}\left[\{|s(\hat{p}_A,\hat{p}_B)|\le n^{-1/2}KC_2(\varepsilon)\}\cap\{(\hat{p}_A,\hat{p}_B)\in D\}\cap\{(\hat{p}_A,\hat{p}_B)\in G_{\varepsilon/2}\}\right]\\
&\ge \mathrm{P}\left[\{|\hat{p}_A-\hat{p}_B|\le n^{-1/2}KC_2(\varepsilon)C_1(\varepsilon/2)\}\cap\{(\hat{p}_A,\hat{p}_B)\in D\}\cap\{(\hat{p}_A,\hat{p}_B)\in G_{\varepsilon/2}\}\right]\\
&= \mathrm{P}\left[\{|\hat{p}_A-\hat{p}_B|\le n^{-1/2}KC_2(\varepsilon)C_1(\varepsilon/2)\}\cap\{(\hat{p}_A,\hat{p}_B)\in D\}\right]\\
&\ge \mathrm{P}\left\{|\hat{p}_A-\hat{p}_B|\le n^{-1/2}KC_2(\varepsilon)C_1(\varepsilon/2)\right\} - \mathrm{P}\left\{(\hat{p}_A,\hat{p}_B)\notin D\right\},
\end{aligned}$$

where the first inequality is trivial: we just intersected with an additional set. The second inequality follows from (6) which hold on D and another intersection. The third inequality is due to (5) on $G_{\varepsilon/2}$. Next, the equality is true since for large enough n, $\hat{p}_A$ and $\hat{p}_B$ are arbitrarily close and being in D implies that they are in $G_{\varepsilon/2}$. The last inequality is obvious. The rate of the first summand in the latter expression is $g(\nu)$ by (3) while the second summand decreases exponentially faster due to (17), and therefore can be ignored. $\square$

Being very similar to the proofs of Theorems 2 and 3, the proofs of Theorems 5 and 6 are omitted.

References

Azriel, D., Mandel, M. & Rinott, Y. (2011). The treatment versus experimentation dilemma in dose-finding studies. *J. Statist. Plan. Infer.* **141**, 2759–68.

Begg, C. B. & Kalish, L. A. (1984). Treatment allocation for nonlinear models in clinical trials: the logistic model. *Biometrics* **40**, 409–20.

Biswas, A., Mandal, S. & Bhattacharya, R. (2010). Multi-treatment optimal response-adaptive designs for phase iii clinical trials. *J. Korean Statist. Soc.* **40**, 33–44.

Brittain, E. & Schlesselman, J. J. (1982). Optimal allocation for the comparison of proportions. *Biometrics* **38**, 1003–9.

Chambaz, A. & van der Laan, M. J. (2011). Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate. *Int. J. Biostatist.* **4**, Article 10.

Hu, F. & Rosenberger, W. F. (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J. Am. Statist. Assoc.* **98**, 671–8.

Hu, F. & Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: Wiley.

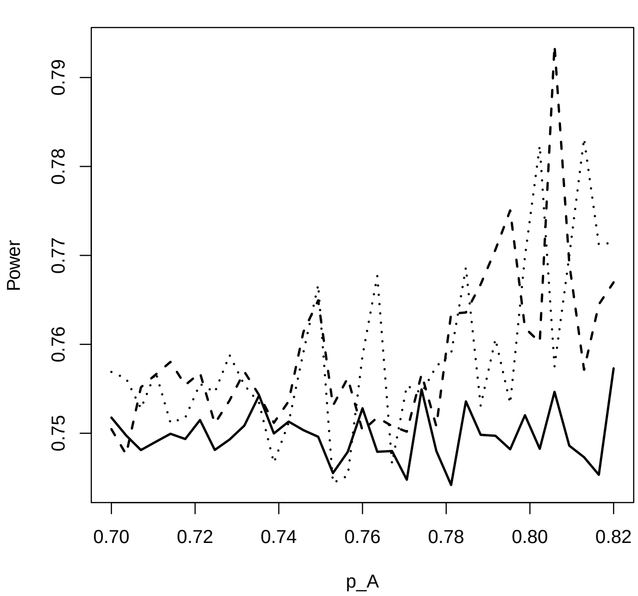
Kalish, L. A. & Harrington, D. P. (1988). Efficiency of balanced treatment allocation for survival analysis. *Biometrics* **44**, 815–21.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N. & Ricks, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57**, 909–13.

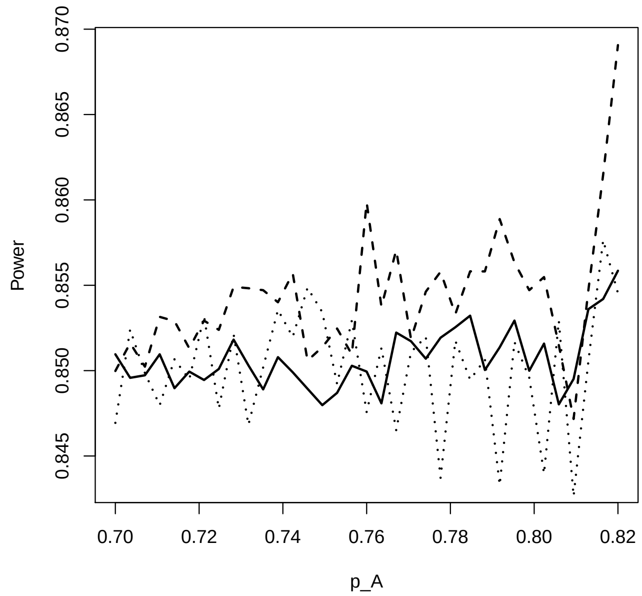
Tymofyeyev, Y., Rosenberger, W. F. & Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *J. Am. Statist. Assoc.* **102**, 224–34.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

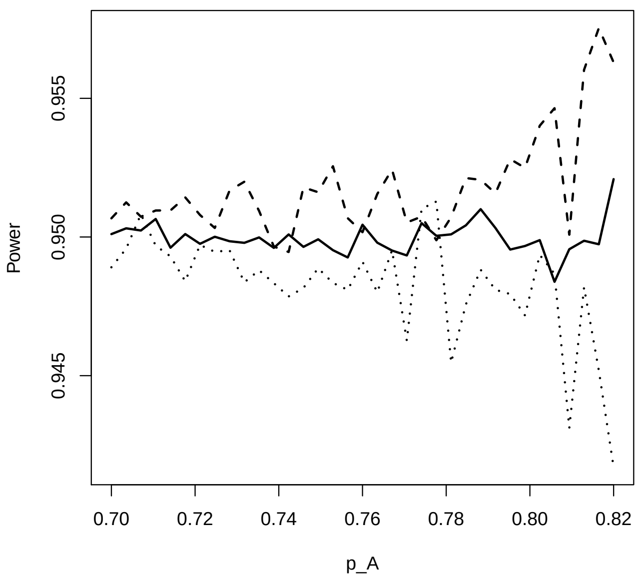
Zhu, H. & Hu, F. (2010). Sequential monitoring of response-adaptive randomized clinical trials. *Ann. Statist.* **38**, 2218–41.



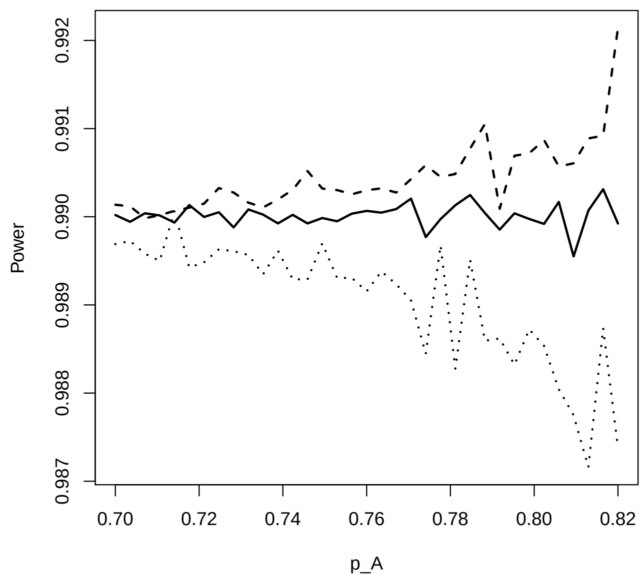
(a) Power ≈ 0·75



(b) Power ≈ 0·85



(c) Power ≈ 0·95



(d) Power ≈ 0·99

Figure 1: Comparison of the power of the two-sided Wald test with critical value $K = 2$ for Neyman (dotted), balanced (solid) and Bahadur allocation (dashed), where $p_B = p_A + 0{\cdot}15$. For each pair (p_A, p_B) the sample size is determined so that the power under balanced design is closest to the target power. The sample sizes range from 103 (power=0·75, p_A =0·82) to 563 (power=0·99, p_A =0·7).