Journal of Statistical Planning and Inference 108 (2002) 255–262
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Mathematical aspects of estimating two treatment effects and a common variance in an assured allocation design

Bruce Levin[a,*], Herbert Robbins[b], Cun-Hui Zhang[b]

[a] *Department of Biostatistics, The Joseph L. Mailman School of Public Health, Columbia University, 600 West 168th Street, New York, NY 10032, USA*

[b] *Department of Statistics, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA*

Abstract

We consider a doubly semi-parametric model for normally distributed random variables which arises in experiments with an assured allocation design. In settling a curious question about estimation of the model's variance parameter, a certain inequality arises that involves the normal probability density function and its first two integrals. The inequality is of mathematical interest in its own right, and is given a rigorous proof.

MSC: 62C12; 62F10; 62G05; 62P10

Keywords: Assured allocation design; Controlled clinical trials; Empirical Bayes estimation; Inequalities; Mixture models; Normal distribution; Semiparametric models; Variance estimation

1. Introduction

The assured allocation design has been proposed for controlled clinical trials in situations where a randomized design is impossible to implement because patients are unwilling to enroll or where clinicians refuse to randomize patients out of ethical concerns. Finkelstein et al. (1996a, b) demonstrate the feasibility of this non-randomized design in such situations, and discuss some specific examples. The appropriate statistical analysis for an assured allocation design rests on the specification of a semi-parametric model whose parameters are estimated by general empirical Bayes methods, as

* Corresponding author.
E-mail address: bl6@columbia.edu (B. Levin).

PII: S0378-3758(02)00312-9

developed in earlier writings by Robbins and Zhang (1988, 1989, 1991) and Robbins (1993). The reader is referred to the above papers for fuller discussion of the issues involved in the design and analysis of an assured allocation trial. In this paper we consider a particular doubly semi-parametric model for normally distributed random variables. Instead of elaborating on practical details of application of the model in a controlled clinical trial setting, however, we focus here on two interesting mathematical questions that arise from the model.

Let X be a quantitative variable which we observe pre-treatment. For simplicity, we may think of X as a baseline value for a similar measurement, denoted by Y, to be observed post-treatment. (More generally, we may think of X as a pre-treatment quantitative risk assessment, and Y as a quantitative endpoint of interest associated with X.) In the assured allocation design, treatment is assigned on the basis of X: all those patients whose X lies at or below a certain pre-specified threshold are given treatment 1 (e.g., a standard treatment), while all those patients whose X lies above the threshold are assured treatment 2 (e.g., an experimental treatment).

The variables X and Y are related through an unobservable, patient-specific, random variable θ, such that given θ, X has a normal distribution with mean θ and variance σ^2. We make no parametric assumption concerning the distribution of θ in the population of patients; we denote the arbitrary and unknown cumulative distribution function of θ by G. Because the distribution of θ is entirely arbitrary, without loss of generality we may assume the allocation threshold to be pre-specified at 0 (possibly after subtraction of a constant from X). The treatments are assumed to affect each patient's θ additively, such that the distribution of Y is normally distributed with a shifted mean $\theta + c_1$ for patients given treatment 1 (those with $X \leqslant 0$), or $\theta + c_2$ for patients given treatment 2 (those with $X > 0$). We also explicitly assume that the treatment does not affect the variance of the response. In symbols, then, our model is given by the following three assumptions:

A1. $\theta \sim G$,
A2. $X|\theta \sim N(\theta, \sigma^2)$,
A3. $Y|\theta, X \sim N(\theta + c_\omega, \sigma^2)$,

where $\omega = \omega(X) = 1 + I[X > 0]$, so that index ω equals 1 or 2, depending on the sign of X.

The statistical problem is to estimate the treatment effects c_1 and c_2 and the variance σ^2.

2. Solution via the *u–v* method

Let u_1 and u_2 be absolutely continuous functions of bounded variation, where u_1 vanishes outside the interval $x \leqslant 0$ and u_2 vanishes outside the interval $x > 0$. An integration by parts produces the well-known result (Stein's lemma) for $j = 1, 2$

$$E[u_j(X)(X - \theta)|\theta] = \sigma^2 E[u_j'(X)|\theta],$$

so that after taking expectations with respect to G and rearranging terms, we have the fundamental general empirical Bayes identity for normal random variables

$$E[\theta u_j(X)] = E[Xu_j(X)] - \sigma^2 E[u_j'(X)].$$

Now from A3

$$E[u_j(X)(Y - X)|X, \theta] = u_j(X)(\theta + c_j - X)$$

and therefore, unconditionally,

$$E[u_j(X)(Y - X)] = c_j E[u_j(X)] - \sigma^2 E[u_j'(X)].$$

Solving the pair of equations for c_j for given σ^2, we find

$$c_j = \left\{ \frac{E[u_j(X)(Y-X)]}{E[u_j(X)]} \right\} + \left\{ \frac{E[u_j'(X)]}{E[u_j(X)]} \right\} \sigma^2 = a_j + b_j\sigma^2 \quad \text{(say)} \tag{1,2}$$

and

$$2\sigma^2 = E[(Y - X - c_\omega)^2]. \tag{3}$$

Eq. (3) holds because

$$\begin{aligned} &E[(Y - X - c_\omega)^2] \\ &\quad = E[E\{I[X \leqslant 0](Y - X - c_1)^2|X,\theta\}] + E[E\{I[X > 0](Y - X - c_2)^2|X,\theta\}] \\ &\quad = E[I[X \leqslant 0]\{\sigma^2 + (X-\theta)^2\} + I[X > 0]\{\sigma^2 + (X-\theta)^2\}] = 2\sigma^2. \end{aligned}$$

For convenience, let $I_1 = I_1(X) = I[X \leqslant 0]$ and $I_2 = I_2(X) = I[X > 0]$. We now specialize u_1 and u_2 to the functions $u_1(X) = XI_1 = XI[X < 0]$, $u_2(X) = XI_2 = XI[X > 0]$, in which case the coefficients a_j and b_j are given by

$$a_j = \frac{E[u_j(X)(Y-X)]}{E[u_j(X)]} = \frac{E[(Y-X)XI_j]}{E[XI_j]} \tag{4}$$

and

$$b_j = \frac{E[u_j'(X)]}{E[u_j(X)]} = \frac{E[I_j]}{E[XI_j]}. \tag{5}$$

Substituting (1) and (2) for c_j in (3) yields a quadratic equation for σ^2 of the form $Q(\sigma^2) = 0$, where

$$Q(t) = E[(Y - X - a_\omega - b_\omega t)^2] - 2t = A - 2Bt + Ct^2 \tag{6}$$

with

$$A = E[(Y - X - a_\omega)^2], \quad B = 1 + E[(Y - X - a_\omega)b_\omega] \quad \text{and} \quad C = E[b_\omega^2]. \tag{7}$$

In practice the quantities in Eqs. (4), (5) and (7) can be estimated with strong consistency by sample averages based on n pairs of observations (X_i, Y_i) for $i = 1, \ldots, n$, e.g.,

$$\hat{a}_j = \frac{\sum_i (Y_i - X_i)X_i I_j(X_i)}{\sum_i X_i I_j(X_i)}, \text{ etc.}$$

3. A curious question

The roots of the equation $Q(t)=0$ are $\{B \pm \sqrt{B^2-AC}\}/C$. The discriminant of $Q(t)$, B^2-AC, is always non-negative, because σ^2 is one of the roots. However, both roots of the quadratic equation are positive, and therefore feasible solutions for σ^2. Which root should one choose? Obviously, the answer is whichever root corresponds to σ^2 in $Q(t)=0$, but which root is that, the smaller or larger root? And which root should one choose when using the sample version, in which expectations are replaced by sample averages?

Consider the special case $\sigma^2=0$, in which case the variables X and θ are the same, and $Y-X=c_\omega$. Then from (4) we have $a_\omega=c_\omega$, so that from (7), $A=0$ and $B=1$. Thus, the two roots are 0 and $2/C>0$. Since C is a quantity that depends only on the distribution of θ, we deduce that the *smaller* root is the correct solution for σ^2 in this case.

It is not obvious how to extend this conclusion for $\sigma^2>0$ with a simple yet rigorous argument; however, the following theorem does confirm the identification of σ^2 with the smaller root generally. In what follows, in order to abstract from finite sample concerns, we continue the discussion in terms of mathematical expectations.

Theorem. *The quadratic equation $Q(t)=0$ has distinct positive roots, unless G is degenerate at $\theta=0$. The parameter σ^2 corresponds to the smaller of the two roots.*

We prove the theorem by establishing three lemmas. The first states that σ^2 is the smaller root if and only if an interesting inequality holds. It involves three functions of θ related to the normal distribution, for which we introduce the following notation: Let $\varphi(z)$ be the standard normal probability density function

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}$$

and let $\Phi(z)$ be the standard normal cumulative distribution function,

$$\Phi(z)=\int_{-\infty}^{z}\varphi(u)\,\mathrm{d}u.$$

Let $\Psi(z)$ be the integral of $\Phi(z)$

$$\Psi(z)=\int_{-\infty}^{z}\Phi(u)\,\mathrm{d}u=\varphi(z)+z\Phi(z).$$

Properties of $\Psi(z)$ include:

(i) $\Psi'(z)=\Phi(z)>0$ and $\Psi''(z)=\varphi(z)>0$;
(ii) $\lim_{z\to-\infty}\Psi(z)=0$;
(iii) $\lim_{z\to-\infty}\Psi(z)/z=0$ (Mills' ratio);
(iv) $\lim_{z\to+\infty}\Psi(z)/z=1$; and
(v) $\Psi(-z)=\Psi(z)-z$.

Thus $\Psi(z)$ is a positive, increasing, convex function with $\Psi(z)>z$ for all z.

Lemma 1. *A necessary and sufficient condition for* σ^2 *to be the smaller root of the equation* $Q(t)=0$ *is the inequality*

$$\frac{E[\Phi(\theta)]\cdot E[\theta\Phi(\theta)]}{E[\Psi(\theta)]}+\frac{E[\Phi(-\theta)]\cdot E[-\theta\Phi(-\theta)]}{E[\Psi(-\theta)]}\geqslant 0 \tag{9}$$

or equivalently

$$\frac{E[\varphi(\theta)]\cdot E[\Phi(\theta)]}{E[\Psi(\theta)]}+\frac{E[\varphi(-\theta)]\cdot E[\Phi(-\theta)]}{E[\Psi(-\theta)]}\leqslant 1. \tag{10}$$

Proof of Lemma 1. The smaller root of $Q(t)=0$ is σ^2 if and only if $Q'(\sigma^2)\leqslant 0$. Evaluating the derivative, $\frac{1}{2}Q'(t)=-E[b_\omega(Y-X-a_\omega-b_\omega t)]-1$, so the necessary and sufficient condition is

$$E[b_\omega(Y-X-a_\omega-b_\omega\sigma^2)]+1\geqslant 0.$$

But $E[Y|\theta,X]=\theta+a_\omega+b_\omega\sigma^2$, so the condition reduces to $E[b_\omega(\theta-X)]+1\geqslant 0$; and $E[b_\omega X]=1$, because $Xb_\omega=X(b_1I_1+b_2I_2)=u_1(X)b_1+u_2(X)b_2$, and from (5) this is

$$u_1(X)\,\frac{P[X\leqslant 0]}{E[u_1(X)]}+u_2(X)\,\frac{P[X>0]}{E[u_2(X)]}$$

with unit expectation. Thus the necessary and sufficient condition reduces simply to $E[\theta b_\omega]\geqslant 0$.

Under assumptions A1–A3, b_1 and b_2 can be evaluated explicitly in terms of expectations of functions of θ alone, as follows. Because the distribution of θ is arbitrary, without loss of generality we may assume that $\sigma^2=1$ (possibly after division of X by a constant), and we do so henceforth. Now $P[X>0]=EP[X>0|\theta]=E[\Phi(\theta)]$ and $P[X\leqslant 0]=E[\Phi(-\theta)]$. Thus

$$\begin{aligned}E[u_2(X)]&=E[XI_2(X)]=E\{E[(X-\theta)I_2(X)|\theta]+E[\theta I_2(X)|\theta]\}\\&=E\left[\int I_2(x)(x-\theta)\varphi(x-\theta)\,\mathrm{d}x+\theta\Phi(\theta)\right]=E[\varphi(\theta)+\theta\Phi(\theta)]\\&=E[\Psi(\theta)]\end{aligned}$$

and then

$$E[u_1(X)]=E[X]-E[u_2(X)]=E[\theta]-E[\Psi(\theta)]=-E[\Psi(-\theta)].$$

Therefore,

$$b_2=\frac{P[X>0]}{E[u_2(X)]}=\frac{E[\Phi(\theta)]}{E[\Psi(\theta)]}$$

and similarly,

$$b_1=\frac{P[X\leqslant 0]}{E[u_1(X)]}=-\frac{E[\Phi(-\theta)]}{E[\Psi(-\theta)]}.$$

Consequently,

$$E[\theta b_\omega|\theta] = E[b_1\theta I_1 + b_2\theta I_2|\theta] = b_1\theta\Phi(-\theta) + b_2\theta\Phi(\theta)$$

from which (9) follows by taking expectations with respect to θ. The equivalent form (10) follows from (9) by adding and subtracting $\varphi(\theta)$ from terms containing $\theta\Phi(\theta)$, using $\Psi(\theta) = \varphi(\theta) + \theta\Phi(\theta)$ to simplify, and rearranging terms. □

4. Proof of inequality (10)

Lemma 2. *For any distribution G satisfying A1*

$$E\left[\frac{\varphi(\theta)\cdot\Phi(\theta)}{\Psi(\theta)}\right] \geqslant \frac{E[\varphi(\theta)]\cdot E[\Phi(\theta)]}{E[\Psi(\theta)]}.$$

Lemma 3. *Inequalities* (9) *and* (10) *are true for all degenerate distributions*, *i.e.*, *for any* θ

$$\frac{\varphi(\theta)\cdot\Phi(\theta)}{\Psi(\theta)} + \frac{\varphi(-\theta)\cdot\Phi(-\theta)}{\Psi(-\theta)} \leqslant 1$$

with equality if and only if θ *is degenerate at* 0.

Lemmas 2 and 3 imply that the necessary and sufficient condition of Lemma 1 holds, and thus the Theorem, because

$$\frac{E[\varphi(\theta)]\cdot E[\Phi(\theta)]}{E[\Psi(\theta)]} + \frac{E[\varphi(-\theta)]\cdot E[\Phi(-\theta)]}{E[\Psi(-\theta)]}$$
$$\leqslant E\left[\frac{\varphi(\theta)\cdot\Phi(\theta)}{\Psi(\theta)} + \frac{\varphi(-\theta)\cdot\Phi(-\theta)}{\Psi(-\theta)}\right] \leqslant 1,$$

where the first inequality follows from an application of Lemma 2 to the distribution of θ and to that of $-\theta$, and where the second inequality follows from Lemma 3 with equality if and only if $\theta = 0$ almost everywhere. The rest of the paper is devoted to a proof of Lemmas 2 and 3.

Proof of Lemma 2. After a change of measure to

$$\mathrm{d}G^*(\theta) = \frac{\Phi(\theta)\,\mathrm{d}G(\theta)}{\int \Phi(\theta)\,\mathrm{d}G(\theta)},$$

we are to show that $E^*[\varphi/\Psi] \geqslant E^*[\varphi/\Phi] \div E^*[\Psi/\Phi]$, i.e., that $\mathrm{Cov}^*(\Psi/\Phi, \varphi/\Psi) \leqslant 0$.

For this it suffices to show that the function $\Psi(\theta)/\Phi(\theta)$ is increasing in θ while $\varphi(\theta)/\Psi(\theta)$ is decreasing in θ. The justification of this assertion is as follows. If $g(u)$ is a non-increasing function with finite expectation with respect to a random variable U, then there exists a u^* such that $g(u) \geqslant E[g(U)]$ for $u < u^*$ and $g(u) \leqslant E[g(U)]$ for $u > u^*$. Then if $f(u)$ is non-decreasing, $\mathrm{Cov}(f(U), g(U)) = E[\{f(U) - f(u^*)\}\cdot\{g(U) - E[g(U)]\}]$, which is the expectation of the product of two expressions either zero or of opposite signs. An alternative justification, pointed out by a referee, follows from the relation

$$2\,\mathrm{Cov}(f(U), g(U)) = E[\{f(U) - f(U')\}\cdot\{g(U) - g(U')\}],$$

where U and U' are independent and identically distributed.

For the assertion that $\Psi(\theta)/\Phi(\theta)$ is increasing

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\Psi(\theta)}{\Phi(\theta)}\right) = \frac{\Phi(\theta)^2 - \Psi(\theta)\varphi(\theta)}{\Phi(\theta)^2},$$

so letting $f_1(\theta) = \Phi(\theta)^2 - \Psi(\theta)\varphi(\theta)$, we are to show $f_1(\theta) > 0$ for all θ. Each of the functions φ, Φ, and Ψ approaches 0 as $\theta \to -\infty$, so $\lim_{\theta\to-\infty} f_1(\theta) = 0$, and thus it suffices to show $f_1'(\theta) > 0$. Now $f_1'(\theta) = 2\varphi\Phi - \{\varphi\Phi - \theta\varphi\Psi\} = \varphi\{\Phi + \theta\Psi\}$. So letting $f_2(\theta) = \Phi(\theta) + \theta\Psi(\theta)$, it suffices to show $f_2(\theta) > 0$ for all θ. Again, $\lim_{\theta\to-\infty} f_2(\theta) = 0 + \lim_{\theta\to-\infty}\theta\{\varphi(\theta) + \theta\Phi(\theta)\} = 0$, because $\lim_{\theta\to-\infty}\theta^2\Phi(\theta) = \lim_{\theta\to-\infty}\theta\varphi(\theta) = 0$, and $f_2'(\theta) = \varphi(\theta) + \theta\Phi(\theta) + \Psi(\theta) = 2\Psi(\theta) > 0$.

Hence $f_2(\theta) > 0$ for all θ, hence $f_1'(\theta) > 0$ for all θ, and thus $f_1(\theta) > 0$ for all θ.

For the assertion that $\varphi(\theta)/\Psi(\theta)$ is decreasing, it suffices to show that $f_3(\theta) = \Psi(\theta)/\varphi(\theta)$ increases in θ. But, as just shown, $f_2(\theta) > 0$, so that $f_3'(\theta) = \{\varphi(\theta)\}^{-1}\{\Phi(\theta) + \theta\Psi(\theta)\} = \{\varphi(\theta)\}^{-1} f_2(\theta) > 0$. This concludes the proof of Lemma 2. □

Proof of Lemma 3. For degenerate G, inequality (9) takes the form $\theta\Phi(\theta)^2/\Psi(\theta) \geqslant \theta\Phi(-\theta)^2/\Psi(-\theta)$. Equality is obvious for $\theta = 0$, so we assume $\theta > 0$, and show that

$$\frac{\Phi(\theta)^2}{\Psi(\theta)} \geqslant \frac{\Phi(-\theta)^2}{\Psi(-\theta)}$$

(the opposite inequality holds for $\theta < 0$). This is equivalent to showing the positivity of the function

$$\begin{aligned} f_4(\theta) &= \Phi(\theta)^2\Psi(-\theta) - \Phi(-\theta)^2\Psi(\theta) \\ &= \Phi(\theta)^2\{\varphi(\theta) - \theta\Phi(-\theta)\} - \Phi(-\theta)^2\{\varphi(\theta) + \theta\Phi(\theta)\} \\ &= \varphi(\theta)\{\Phi(\theta)^2 - \Phi(-\theta)^2\} - \theta\Phi(-\theta)\Phi(\theta)\{\Phi(\theta) + \Phi(-\theta)\} \\ &= \varphi(\theta)\{2\Phi(\theta) - 1\} - \theta\Phi(\theta)\{1 - \Phi(\theta)\} \\ &= \varphi(\theta)\Phi(\theta) - \Psi(\theta)\{1 - \Phi(\theta)\}. \end{aligned}$$

Differentiating the final expression, we have

$$\begin{aligned} f_4'(\theta) &= \varphi(\theta)^2 - \theta\varphi(\theta)\Phi(\theta) + \Psi(\theta)\varphi(\theta) - \Phi(\theta)\{1 - \Phi(\theta)\} \\ &= 2\varphi(\theta)^2 - \Phi(\theta)\{1 - \Phi(\theta)\}. \end{aligned}$$

Note that $f_4'(0) = (1/\pi) - 1/4 > 0$. Since $f_4 \to 0$ as $\theta \to \infty$, it suffices to show that $f_4'(\theta)$ has at most one zero for $\theta > 0$. In fact, we prove that f_4'' has at most one positive zero, which implies that f_4' does too. Now

$$f_4''(\theta) = -4\theta\varphi(\theta)^2 - \varphi(\theta) + 2\Phi(\theta)\varphi(\theta) = \varphi(\theta)\{2\Phi(\theta) - 1 - 4\theta\varphi(\theta)\},$$

so it suffices to show that

$$f_5(\theta) = 2\Phi(\theta) - 1 - 4\theta\varphi(\theta)$$

has at most one positive root. But

$$f_5'(\theta) = 2\varphi(\theta) - 4\varphi(\theta) + 4\theta^2\varphi(\theta) = 2\varphi(\theta)\{2\theta^2 - 1\}$$

has exactly one zero for $\theta > 0$. Therefore, $f_5(\theta)$ can have at most one positive root (it has exactly one, since $f_5'(0) < 0$, and $\lim_{\theta\to\infty} f_5(\theta) = 1$), and this concludes the proof of Lemma 3. □

5. Remark

It might appear that an argument from continuity would allow generalization from the special case $\sigma^2 = 0$ considered above to show that σ^2 is always the smaller root of $Q(t) = 0$. It is unclear how to make this argument rigorous, however, because after rescaling, the case of any non-zero σ^2 is essentially the same as the case $\sigma^2 = 1$, whereas the case $\sigma^2 = 0$ cannot be rescaled, a rather discontinuous state of affairs. In addition, one would have to show that the two roots of the quadratic remained separated and did not "trade places" as σ^2 increased from zero to an arbitrary positive value. The analytic approach used in this paper does provide a rigorous proof, and reveals an inequality of some independent aesthetic appeal, although we would be delighted to learn if a simple yet rigorous continuity argument exists.

References

Finkelstein, M.O., Levin, B., Robbins, H., 1996a. Clinical and prophylactic trials with assured new treatment for those at greater risk. Part I—Introduction. Amer. J. Public Health 86, 691–695.

Finkelstein, M.O., Levin, B., Robbins, H., 1996b. Clinical and prophylactic trials with assured new treatment for those at greater risk. Part II—Examples. Amer. J. Public Health 86, 696–705.

Robbins, H., 1993. Comparing two treatments under biased allocation. La Gaz. Sci. Math. Que. 15, 35–41.

Robbins, H., Zhang, C.-H., 1988. Estimating a treatment effect under biased sampling. Proc. Nat. Acad. Sci. USA 85, 3670–3672.

Robbins, H., Zhang, C.-H., 1989. Estimating the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug. Proc. Nat. Acad. Sci. USA 86, 3003–3005.

Robbins, H., Zhang, C.-H., 1991. Estimating a multiplicative treatment effect under biased allocation. Biometrika 78, 349–354.