ELSEVIER

Journal of Statistical Planning and Inference 46 (1995) 105–121

journal of statistical planning and inference

Maximin efficiency-robust tests and some extensions

Sudip Bose[a,*], Eric Slud[b,1]

[a] Department of Statistics/C & IS, The George Washington University, 2201 G. Street, NW Washington, DC 20052, USA

[b] University of Maryland, MD, USA

Received 7 September 1992; revised 9 March 1994

Abstract

The Maximin Efficiency-Robust Test idea of Gastwirth (1966) was to maximize the minimum asymptotic power (for fixed size) versus special local families of alternatives over some specially chosen families of score statistics. This approach is reviewed from a general decision-theoretical perspective, including some Bayesian variants. For two-sample censored-data rank tests and stochastically ordered but not proportional-hazard alternatives, the MERT approach leads to customized weighted-logrank tests for which the weights depend on estimated random-censoring distributions. Examples include statistics which perform well against both Lehmann and logistic alternatives or against families of alternatives which include increasing, decreasing, and 'bathtub-shaped' hazards.

AMS Subject Classification: 62G20, 62E20

Key words: Asymptotic relative efficiency; Bayesian procedures; Best-precedence test; Censored data; Completely positive matrices; Contiguous alternatives; Decision theory; Efficacy; Robustness to priors; Statistical decision problem; Survival analysis; Weighted rank statistic

1. Introduction

Consider large-sample testing of a null hypothesis $\mathbf{H}_0$ against contiguous alternatives $\mathbf{H}_n(v)$ indexed by $v \in \mathbb{L}$, where $\mathbb{L}$ is a finite- or infinite-dimensional 'tangent space' to an underlying parameter space Θ and where $\mathbf{H}_n(0) = \mathbf{H}_0$ corresponds to a fixed null hypothesis parameter $\vartheta_0 \in \Theta$. In parametric, nonparametric, and semiparametric

* Corresponding author.

[1] Research partially supported by Office of Naval Research contracts N00014-89-J-1051 and N00014-92-C0019.

SSDI 0378-3758(94)E00091-3

settings (Ibragimov and Hasminski, 1981; Hajek and Sidak, 1967; Gill, 1980), general regularity conditions are known under which there is an essentially unique most-powerful score-type statistic $S_{n,v}$ for each such contiguous alternative. Moreover, if $w \neq v$, then the Pitman asymptotic relative efficiency (ARE) of $S_{n,v}$ versus $S_{n,w}$ under $\mathbf{H}_n(w)$ is a function $A^2(v,w)$ such that $A(v,v)=1$, $A(v,w)=A(w,v)$, and $A(cv,w)=A(v,w)$. Here Pitman ARE for testing $\mathbf{H}_n$ vs. $\mathbf{H}_0$ is defined as the squared ratio of efficacies $(e_n(v)/e_n(w))^2$, where $e_n(w) \equiv E(S_{n,w}\,|\,\mathbf{H}_n)/\mathrm{Var}^{1/2}(S_{n,w}\,|\,\mathbf{H}_0)$. Typically the parameter set is a linear space Θ, and the contiguous alternatives are of the form $\vartheta = \vartheta_0 + v/\sqrt{n}$ for fixed v.

We want to find a test statistic with good power, assuming that the probability law generating our data is one of the contiguous alternatives $\mathbf{H}_n(v)$ for v in a finite or infinite subset $\mathbb{T} \subset \mathbb{L}$. Asymptotically for large samples, the unsigned figure of merit $|A(v,w)|$, not depending on n or the significance level α, quantifies the performance of $S_{n,w}$ as two-sided test statistic under $\mathbf{H}_n(v)$. (See Section 2 for modifications in the setting of one-sided tests.) If $v \in \mathbb{T}$ is a state-of-Nature variable, and $w \in \mathbb{L}$ is the decision or action we will take, then Wald's (1950) formulation of statistical decision problems recasts the choice of a test procedure as a Game Against Nature with loss function $1 - A^2(v,w)$.

In purely nonparametric settings, where $S_{n,w}$ is the optimally weighted rank-statistic described by Hajek and Sidak (1967) and Gill (1980), and v indexes sufficiently smooth alternatives to the null hypothesis in the two-sample problem, $A(v,w)$ is the asymptotic correlation between $S_{n,v}$ and $S_{n,w}$ under $\mathbf{H}_n(v)$, and the asymptotic covariance between $S_{n,v}$ and $S_{n,w}$ is bilinear in v, w. More generally, under regularity conditions ensuring the applicability of the theory of 'influence functionals' (Serfling, 1980), $A(\cdot,\cdot)$ is a normalized bilinear form in the weight-functions v, w. In the nonparametric setting of Hajek and Sidak (1976), Gastwirth (1966) introduced the ideas in the foregoing paragraph and defined the Maximin Efficiency Robust Test (MERT), as the test procedure S_{n,w^*} for the value w^* such that

$$\inf_{v \in T} A^2(w^*, v) = \sup_{w \in L} \inf_{v \in T} A^2(w, v). \tag{1.1}$$

The maximin idea makes sense even if the 'action space' $\mathbb{L}$, i.e. the allowable parametric class of test procedures, were different from the whole class of procedures $\{S_{n,w} : w \in \mathbb{L}\}$ (cf. Example 3.5). For special problems and sets $\mathbb{T}$ of alternatives to a null hypothesis $\mathbf{H}_0$, the MERT procedure may also have maximin power with respect to a large family of nonlocal alternatives. (See Doksum, 1967, for an important example.)

Our objective in this paper is first to explain why the MERT idea is more general than Gastwirth's (1966, 1985) use of it, and then to describe customized weighted-logrank and other MERT-type tests for right-censored survival data. The paper is organized as follows. Section 2 summarizes the theoretical results of Gastwirth on MERT, as well as extensions and algorithms related to existence and uniqueness of solutions when there are only finitely many alternatives. Section 3 formulates MERT problems for randomly right-censored survival data, accommodating censoring distributions as nuisance parameters, and gives examples in nonparametric censored two-sample problems. Section 4 presents the Bayesian approach to MERT considered as a statistical decision problem. Section 5 contains discussion and conclusions.

2. MERT versus finitely many contiguous alternatives

First restrict attention, as Gastwirth (1966) did, to problems where $A(v,w)$ is a normalized quadratic form

$$A(v,w)=\frac{v'\Sigma w}{\sqrt{v'\Sigma v\, w'\Sigma w}}, \quad v,w\neq 0, \tag{2.1}$$

where $'$ denotes transpose or adjoint, v and w lie in a *Hilbert* space $\mathbb{L}$ with inner product $(\cdot,\cdot)_L$, Σ is a bounded symmetric nonnegative definite linear operator on $\mathbb{L}$, and $v'\Sigma w\equiv(v,\Sigma w)_L$. By convention, when $(v,\Sigma v)_L=0$, expression (2.1) is taken to be 0. We treat parametric MERT problems here, and nonparametric problems in Section 3.

Consider the statistical problem of testing the null hypothesis $\mathbf{H}_0$: $\vartheta=\vartheta_0$, where $\vartheta\in\Theta\subset R^k$ denotes a finite-dimensional unknown parameter in a large-sample problem in which data $\underline{X}\equiv\underline{X}_n$ have underlying density $f_n(x_n;\vartheta)$ with respect to a reference measure μ. Here Θ is an open neighborhood of ϑ_0, and the densities are assumed to be such that the alternatives $\mathbf{H}_n(w)$: $\vartheta=\vartheta_0+w/\sqrt{n}$ are 'regular' (Ibragimov and Hasminskii, 1981), and in particular are contiguous to $\mathbf{H}_0$. A standard result ('Noether's theorem' for regular parametric problems, Ibragimov and Hasminskii, 1981, pp. 113–120) says that the ARE under $\mathbf{H}_n(v)$ of the two-sided test procedure of size α based on the statistic $w'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0)$ is the square of

$$A(v,w)\equiv\lim_{n\to\infty}\frac{E(v'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0)\nabla'_\vartheta\log f_n(\underline{X}_n;\vartheta_0)\,w)}{\sqrt{E(v'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0))^2\,E(w'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0))^2}} \tag{2.2}$$

and the asymptotic power of a one-sided test of size α is

$$\lim_{n\to\infty}\{\Phi(-z_\alpha+A(v,w)\sqrt{E(v'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0))^2/n})\} \tag{2.3}$$

with expectations taken under the null hypothesis, i.e., with respect to the measure $f_n(\underline{x}_n;\vartheta_0)\,\mu(d\underline{x})$. The tangent space $\mathbb{L}$ can be any subspace of $\mathbb{R}^k$, with the vector w identified asymptotically with the $L^2(\mathbb{R}^k, f_n(\underline{x}_n;\vartheta_0)\,\mu(d\underline{x}))$ function $w'\nabla_\vartheta\log f_n(\underline{x};\vartheta_0)$. A natural inner product on $\mathbb{L}$ is given by

$$(v,w)\equiv\lim_{n\to\infty}\frac{1}{n}E(v'\nabla_\vartheta\log f_n(\underline{X}_n;\vartheta_0)\,\nabla'_\vartheta\log f_n(\underline{X}_n;\vartheta_0)\,w) \tag{2.4a}$$

with norm defined by $\|v\|^2 \equiv (v, v)$. Equivalently, the linear operator Σ for which $(v, w) \equiv (v, \Sigma w)_L$ is the 'Fisher information matrix'

$$\Sigma \equiv \lim_{n\to\infty} \frac{1}{n} E(\nabla_\vartheta \log f_n(\underset{\sim}{X}_n; \vartheta_0) \nabla'_\vartheta \log f_n(\underset{\sim}{X}_n; \vartheta_0)). \tag{2.4b}$$

Assume without loss of generality that Σ given by (2.4b) is nonsingular. (Otherwise replace $\mathbb{R}^k$ by $\Sigma(\mathbb{R}^k)$, and Σ by its restriction to $\Sigma(\mathbb{R}^k)$.) Problem (1.1) now becomes

$$\max_{w\in\mathbb{R}^k\setminus\{0\}} \min_{v\in\mathbb{T}\setminus\{0\}} \frac{|(v, w)|}{\|v\| \, \|w\|}, \tag{2.5a}$$

where $\mathbb{T} \subset \mathbb{L}$ is the set of alternatives against which we want adequate power. Without loss of generality, from now on take $\mathbb{T}$ to consist of vectors v such that $\|v\| = 1$.

For each fixed v in $\mathbb{T}$, the objective function in (2.5a) only increases if w is replaced by its orthogonal projection under $(\cdot,\cdot)$ onto the linear span of $\mathbb{T}$. Thus the maximization in (2.5a) may be restricted to span($\mathbb{T}$) even in general non- or semiparametric settings with infinite-dimensional $\mathbb{L}$. In the finite-dimensional case, if span($\mathbb{T}$) is strictly smaller than $\mathbb{L}$, denote by Π the Euclidean orthogonal projection of elements of $\mathbb{L}$ onto span($\mathbb{T}$), and let $\Sigma = B + C + C' + D$ denote the Euclidean 'block-decomposition' of the Fisher information matrix Σ, with $B = \Pi\Sigma\Pi$, $C = \Pi\Sigma(I - \Pi)$, $D = (I - \Pi)\Sigma(I - \Pi)$. Removing the w-component orthogonal under $(\cdot,\cdot)$ to all elements of $\mathbb{T}$ leaves $w^* = \Pi w - D^{-1}C'w$, and it is well known (Silvey, 1970, Chapter 7) that the score statistic S_{n,w^*} is asymptotically equivalent under $\mathbf{H}_0$: $\vartheta = \vartheta_0$ to the score statistic in direction Πw with $(I - \Pi)(\vartheta - \vartheta_0)$ replaced by the restricted maximum likelihood estimator of $\vartheta - \vartheta_0$ on the parameter-set $\{\vartheta\colon \Pi(\vartheta - \vartheta_0) = 0\}$.

The value of (2.5a) is nonincreasing as $\mathbb{T}$ gets larger. Although there are special examples of infinite sets $\mathbb{T}$ for which (2.5a) can be calculated explicitly (Gastwirth, 1985; Zucker and Lakatos, 1990; Zucker, 1992), the key to applying MERT tests is calculating the value and optimal w in (2.5a) when $\mathbb{T}$ is finite.

2.1. Problems equivalent to MERT

Assume that $\mathbb{T}$ has m elements, $m < \infty$. Then (1.1) and (2.5a) have several equivalent formulations, first described by Gastwirth (1966). Let $v_1, \ldots, v_m$ be unit vectors in $\mathbb{L} = \mathbb{R}^k$ equal to the elements of $\mathbb{T}$, let $\boldsymbol{C}$ denote the $m \times m$ correlation matrix with (i,j) entry equal to $(v_i, v_j) = v_i'\Sigma v_j$, and denote by $\boldsymbol{e}_j$ the vector in $\mathbb{R}^m$ with jth entry 1 and other entries 0. Take $w = \sum_{i=1}^m a_i v_i$, with $\boldsymbol{a}'\boldsymbol{C}\boldsymbol{a} = 1$. For all j, $|(v_j, w)| = |\boldsymbol{a}'\boldsymbol{C}\boldsymbol{e}_j|$, and (2.5a) is equivalent to either of the following:

$$\max_{\{a\in\mathbb{R}^m : a'Ca = 1\}} \min_{1\leqslant j\leqslant m} |\boldsymbol{a}'\boldsymbol{C}\boldsymbol{e}_j|, \tag{2.5b}$$

$$\max_{\{b = (b_1, \ldots, b_m) \in \text{range}(C) : |b_j| \geqslant 1,\ j = 1, \ldots, m\}} 1/\sqrt{\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}}. \tag{2.5c}$$

Here the inverse in (2.5c) is a generalized inverse. To see the equivalence of (2.5c) to (2.5b) when the latter is not 0, put $\boldsymbol{b} \equiv \boldsymbol{C}\boldsymbol{a}/\min_{1 \leqslant j \leqslant m} |(\boldsymbol{C}\boldsymbol{a})_j|$, so that $\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b} = (\min_{1 \leqslant j \leqslant m} |\boldsymbol{a}'\boldsymbol{C}\boldsymbol{e}_j|)^{-2}$. The optimal w in (2.5a) is $\sum_{i=1}^m a_i v_i$ for the optimal $\boldsymbol{a}$ in (2.5b), and is proportional to $\sum_{i=1}^m (\boldsymbol{C}^{-1}\boldsymbol{b})_i v_i$ for the optimal $\boldsymbol{b}$ in (2.5c).

If the MERT problem were posed as maximizing the minimum power (2.3) of *one-sided* score-tests versus alternatives $\mathbf{H}_n(v_j)$, then the objective function in (2.5a) becomes $(v, w)/(\|v\| \|w\|)$, the absolute-value signs are removed in Eqs. (2.5b) and (2.5c), yielding in place of (2.5c)

$$\max_{\{\boldsymbol{b} = (b_1, \ldots, b_m) \in \mathrm{range}(\boldsymbol{C}) : |b_j| \geqslant 1,\ j = 1, \ldots, m\}} 1/\sqrt{\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}}. \tag{2.6}$$

This is equivalent to minimizing the convex functional $\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}$ over the convex cone $\{\boldsymbol{b} = (b_1, \ldots, b_m) \in \mathrm{range}(\boldsymbol{C}) : b_j \geqslant 1, j = 1, \ldots, m\}$. The solution exists, is unique, and can be found by quadratic programming techniques as mentioned in Gastwirth (1966) (cf. Section 2.2 below).

MERT formulations (2.5a) and (2.6) are different in general, as can be seen even in dimension $k = 2$ by choosing vectors $v_1 = (1, 0)'$, $v_2 = (-1, \delta)'$ for small $\delta > 0$. Then $(1, 0)'$ is an approximately optimal w for (2.5a), with objective function value close to 1, while an optimal w for (2.6) is close to $(0, 1)'$, with (2.6) close to 0.

Gastwirth's (1966) and later work solved MERT problem (2.6) or equivalent forms, having formulated (2.5a), with vectors v_i such that $(v_i, v_j) \geqslant 0$ for all i, j. In that setting, Gastwirth (1966) claimed but did not prove that the optimal solution w arising from (2.6) is also optimal for (2.5a). We also cannot prove this assertion in general, but justify it generally enough to cover all previous MERT applications.

A nonnegative-definite symmetric $m \times m$ matrix $\boldsymbol{C}$ such that there exists vectors $v_1, \ldots, v_m$ within the positive orthant of $\mathbb{R}^k$ (i.e., which all have nonnegative coordinates) and such that $C_{ij} = (v_i, v_j)$ for all i, j, is called *completely positive* (Hall, 1967). These matrices have been actively studied. For $m \leqslant 4$, a covariance matrix $\boldsymbol{C}$ is completely positive if and only if all its entries are nonnegative (Maxfield and Minc, 1962). However, Hall (1967, p. 267) gives the example

$$\begin{pmatrix} 1 & 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0.75 & 0 & 0.5 \\ 0 & 0.75 & 1 & 0.5 & 0 \\ 0.5 & 0 & 0.5 & 1 & 0 \\ 0.5 & 0.5 & 0 & 0 & 1 \end{pmatrix}$$

of a nonnegative correlation matrix which is *not* completely positive. Hall (1967, Section 16.2) shows that the completely positive matrices form a closed convex cone. Recently Grone et al. (1991) have proved that the pattern of zeroes of the $m \times m$ matrix $\boldsymbol{C}$ characterizes it as completely positive if and only if the graph of m nodes, with an edge between i, j whenever $C_{ij} > 0$, has no complete cycle of length greater than 4.

Lemma 2.1. *If the matrix* $C=((v_i,v_j))_{1\leqslant i,j\leqslant m}$ *is completely positive, then an optimal solution* $w=\sum_{i=1}^{m}(C^{-1}\boldsymbol{b})_i v_i$ *to* (2.6) *is optimal for* (2.5*a*).

Proof. If $(v_i,u_j)\geqslant 0$ for $i,j=1,\ldots,m$, where $\{u_j\}$ is an orthonormal basis for $\mathbb{L}$ with respect to $\|\cdot\|$, and if $w\in\mathbb{L}$ is any unit vector, then

$$|(v_j,w)|/(\|v_j\|\,\|w\|)\leqslant(v_j,w^*)/(\|v_j\|\,\|w^*\|),\quad j=1,\ldots,m,$$

where $w^*\equiv\sum_{i=1}^{m}|(w,u_i)|\,u_i$ is also a unit vector. Thus an optimizing w in (2.5a) must already lie in the positive orthant of $\mathbb{L}=\mathbb{R}^k$ and coincides, as asserted, with an optimizing w in (2.6). □

Corollary 2.2. (a) *If* $m\leqslant 4$ *and* $C_{ij}\geqslant 0$ *for all i,j, then* (2.5*a*) *and* (2.6) *are equivalent.*
(b) *Suppose that* $C=((v_i,v_j))_{1\leqslant i,j\leqslant m}=(\langle\rho_i,\rho_j\rangle)_{1\leqslant i,j\leqslant m}$, *where* ρ_j *are elements and* $\langle\cdot,\cdot\rangle$ *the inner product of a space* $L^2(\mathbb{R},\mu)$ *for a Borel probability measure* μ *on the real line, where the functions* ρ_j *all have* μ*– almost everywhere the same signs, i.e.,*

$$\mu(\{t:\ \rho_j(t)\geqslant 0 \text{ for all } j \text{ or } \rho_j(t)\leqslant 0 \text{ for all } j\})=1\,.$$

Then C *is completely positive, and* (2.5*a*) *and* (2.6) *are equivalent.*

Proof. Except for the assertion that C in (b) is completely positive, this follows immediately from Lemma 2.1. Under the hypotheses of (b) on the functions ρ_j, for arbitrarily small $\delta>0$, let $A_1,\ldots,A_k$ denote a finite measurable partition of $(\mathbb{R},\mathscr{B})$ such that the functions ρ_j have the same constant sign $\varepsilon_i\in\{-1,1\}$ throughout each partition atom A_i, and such that each ρ_j differs by at most $\delta\sqrt{\langle\rho_j,\rho_j\rangle}$ in L^2 norm from a $\sigma(A_1,\ldots,A_k)$ function r_j for which $\varepsilon_i r_j(\cdot)\geqslant 0$ on A_i. Then $C_{ij}=\langle\rho_i,\rho_j\rangle$ differs from $\langle r_i,r_j\rangle$ by at most $(2\delta+\delta^2)\sqrt{\langle\rho_j,\rho_j\rangle\langle\rho_i,\rho_i\rangle}$, and the $m\times m$ matrix with entries $\langle r_i,r_j\rangle$ by at most $(2\delta+\delta^2)\sqrt{\langle\rho_j,\rho_j\rangle\langle\rho_i,\rho_i\rangle}$, and the $m\times m$ matrix with entries $\langle r_i,r_j\rangle$ is obviously completely positive since all inner products between functions r_j and elements of the orthogonal basis $\{\varepsilon_i\chi_{A_i}\}$ of $L^2(\mathbb{R},\sigma(A_1,\ldots,A_k),\mu)$ are positive. Since the completely positive $m\times m$ matrices form a closed convex cone (Hall, 1967, p. 267), the arbitrary choice of δ implies that C is completely positive. □

Gastwirth's (1966, 1985) nonparametric MERT tests all arose from alternative-families satisfying the hypotheses of Corollary 2.2(b), and the same is true of the alternatives considered by Zucker and Lakatos (1990) and Zucker (1992).

2.2. Characterization of MERT procedures. Computational algorithms

We break (2.5c) into 2^m 'restricted-sign' subproblems, in each of which the signs of the components of $\boldsymbol{b}$ match those of a fixed element of $\{-1,1\}^m$. Comparing the 2^m

objective-function values at the solutions of the restricted-sign subproblems provides a solution $\boldsymbol{b}$ to (2.5c).

For fixed $\boldsymbol{\varepsilon}$ in $\{-1,1\}^m$, the restricted-sign subproblem is to find the unique local minimum for the positive definite strictly convex quadratic form $\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}$ on the closed convex set $\{\boldsymbol{b}\in\text{range}(\boldsymbol{C})\colon b_i\varepsilon_i \geqslant 1 \text{ for all } i\}$. Denote by E the diagonal matrix with the elements of $\boldsymbol{\varepsilon}$ along its principal diagonal. If $\boldsymbol{C}$ is singular of rank $r<m$, then define D to be a $(m-r)\times m$ matrix with rows orthogonal to each other and to the column space of $\boldsymbol{C}$. By Theorem 2 of Wolfe (1959), $\boldsymbol{b}\in\text{range}(\boldsymbol{C})$ minimizes $\boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}$ subject to $b_i\varepsilon_i\geqslant 1$ for all i, if and only if there exists an $(m-r)$-vector $\boldsymbol{u}$ such that for all i,

$$\varepsilon_i(\boldsymbol{C}^{-1}b+D'\boldsymbol{u})_i\geqslant 0, \quad \text{and} \quad (b_i\varepsilon_i-1)(\boldsymbol{C}^{-1}\boldsymbol{b}+D'u)_i=0. \tag{2.7}$$

Clearly these conditions are also necessary for the solution $\boldsymbol{b}$ to the overall problem where we do not fix the signs of the components b_i.

Wolfe (1959) describes a simplex-type procedure to solve the restricted-sign quadratic programming problem formulated above. Package software implementing such algorithms can be found in IMSL (subroutine QPROG) or SAS (subroutine LCP of PROC IML, for the problem presented in the form (2.7)). Another approach, which successively selects the canonical basis vectors which must be combined convexly to obtain the MERT, is described by Gastwirth (1985), p. 381) and works in many problems but not always. Alternatively, if the dimension m of the MERT problem is small, as in the cases where we advocate application of MERT tests, then (for C nonsingular) an easily-programmed brute-force search solves for vectors $\boldsymbol{b}$ satisfying (2.7) in restricted-sign subproblems. Computer code for this search, written in Turbo-Pascal and running in a few seconds on personal computers for dimensions up to 5 or 6, will be provided to interested readers.

3. MERT for censored survival data

Survival analysis (specifically the two-sample problem, with right-censored data) is rich in potential applications of MERT. The estimation of nuisance parameters related to group-0 hazard function and right-censoring distribution(s) in censored two-sample testing is no obstacle to the implementation of MERT.

Suppose that a random sample of right-censored survival data $(\min(X_i,C_i), I_{[X_i\leqslant C_i]}, Z_i\colon i=1,\ldots,n)$ is observed, where Z_i is a binary treatment-group indicator, and X_i and C_i are respectively failure- and censoring-time variables which are conditionally independent given Z_i. The group-0 failure intensity $\lambda(t)$, the probability $\gamma=P(Z_i=1)$, and the censoring survival functions $L_j(t)\equiv P(C_1\geqslant t\,|\,Z_1=j)$ for $j=0,1$, are nuisance parameters. The desired inference concerns the group-1 failure intensity $\lambda_1(t)$ or ratio $\lambda_1(t)/\lambda(t)$ specified by the family of hypotheses

$$\mathbf{H}_{n,U}\colon \lambda_1(t)=\lambda(t)\,\mathrm{e}^{U(t,S(t))/\sqrt{n}}$$

contiguous to the null hypothesis $\mathbf{H}_0$ $(=\mathbf{H}_{n,0})$: $\lambda_1(t)=\lambda(t)$. The function U is nonrandom and continuous in its second argument. Thus $\lambda_1(\cdot)$ depends continuously upon the nuisance-function λ, through $S(t)=\exp(-\int_0^t \lambda(u)\,\mathrm{d}u)$. If the data are summarized through the processes

$$N_k(t)=\sum_{i=1}^{n} I_{[X_i \leq \min(C_i,t)]} I_{[Z_i=k]}, \quad N(t)=N_1(t)+N_0(t),$$

$$Y_k(t)=\sum_{i=1}^{n} I_{[\min(C_i,X_i)\geq t]} I_{[Z_i=k]}, \quad Y(t)=Y_1(t)+Y_0(t),$$

then the asymptotically optimal test statistics S_n take the form

$$Z_W \equiv \int W(s,S_{KM}(s))\left(\mathrm{d}N_1(s)-\frac{Y_1(s)}{Y(s)}\,\mathrm{d}N(s)\right) \Bigg/ \sqrt{\int W^2(s,S_{KM}(s))\,\frac{Y_1(s)\,Y_0(s)}{Y(s)^2}\,\mathrm{d}N(s)},$$

where $S_{KM}(t)$ denotes the left-continuous Kaplan–Meier survival-curve estimator for $S(t)$, and $W(s,u)$ is a nonrandom weight-function which optimally takes the form $W \equiv U$.

Under $\mathbf{H}_n, U$, the ARE of the test Z_W versus the first-order optimal test Z_U, corresponding to the functional $A^2(W,U)$ of previous sections, is given (Gill, 1980) by the squared correlation

$$\rho^2(Z_W,Z_U)=\frac{(\int W(s)\,U(s)\,\psi(s)\,\mathrm{d}s)^2}{\int W^2(s)\,\psi(s)\,\mathrm{d}s \int U^2(s)\,\psi(s)\,\mathrm{d}s}.$$

where for brevity $U(t)$, $W(t)$ replace $U(t,S(t))$, $W(t,S(t))$, and

$$\psi(s)=\gamma(1-\gamma)\,L_1(s)\,L_0(s)\,S(s)\,\lambda(s)/\{\gamma\,L_1(s)+(1-\gamma)\,L_0(s)\}.$$

Let functions $U(\cdot)\equiv U(\cdot,S(\cdot))\in L^2([0,\infty),\psi(s)\,\mathrm{d}s)$ have unit norm: $\int U^2\psi = \|U\|^2 = 1$. As in Section 2, $\mathbb{T}$ is a finite set of alternatives indexed by U. The objective function depends upon the nuisance parameters γ, $\lambda(\cdot)$, $L_j(\cdot)$ only through ψ, becoming

$$A(W,U)=\int W(s)\,U(s)\,\psi(s)\,\mathrm{d}s \Big/ \sqrt{\int W^2(s)\,\psi(s)\,\mathrm{d}s}.$$

The effect of censoring upon $A(W,U)$ is simply to introduce the factor $L_1(s)\,L_0(s)/\{\gamma L_1(s)+(1-\gamma)\,L_0(s)\}$ into ψ.

The function ψ will usually be unknown and must be estimated. The empirical estimator for $\Psi(t)\equiv\int_0^t \psi(s)\,\mathrm{d}s$ is $\hat{\Psi}(t)\equiv n^{-1}\int_0^t Y_1 Y_0 Y^{-2}\,\mathrm{d}N$ and is consistent in the sense that, by contiguity to $\mathbf{H}_0$,

$$\sup\left\{\left|\int U(t)\,\mathrm{d}\hat{\Psi}(t)-\int U(t)\psi(t)\,\mathrm{d}t\right|:\ U\in\mathbb{T}\right\}\to 0$$

in probability under each alternative. Here we either rely on finiteness of $\mathbb{T}$ or, via empirical-process theory, allow $\mathbb{T}$ to be a family of bounded continuous functions or special infinite family such as a Vapnik–Cervonenkis class (Pollard, 1986). When, for all $U\in\mathbb{T}$, $\int U\,\mathrm{d}\hat{\Psi}$ based on large-sample data is substituted for $\int U\,\mathrm{d}\Psi$ in the entries $A(W,U)$ of the matrix $\boldsymbol{C}$, the MERT procedure solving (2.5a) is asymptotically the same, at least for finite $\mathbb{T}$, as if ψ had been known. Thus MERT ideas apply generally to two-sample survival data.

Example 3.1 (*Logrank/Wilcoxon combination*). Take $m=k=2$, $\boldsymbol{C}$ the 2×2 correlation matrix with $C_{21}=\int S\,\mathrm{d}\Psi/\sqrt{\Psi(\infty)\int S^2\,\mathrm{d}\Psi}$, with $U_1(t)$, $U_2(t)$ respectively proportional to 1 and $S(t)/\sqrt{\int S^2\,\mathrm{d}\Psi}$. The statistics Z_{U_1} and Z_{U_2} are the well-known logrank and Peto-Prentice generalized Wilcoxon statistics (Gill, 1980). By symmetry, the optimal $\boldsymbol{b}$ in (2.5c) and $\boldsymbol{a}$ in (2.5b) are proportional to $\mathbf{1}=(1,1)'$. Therefore the MERT statistic has scores proportional to $1/\sqrt{\Psi(\infty)}+S_{KM}(t)/\sqrt{\int S_{KM}^2\,\mathrm{d}\Psi}$. In the uncensored case, ψ is proportional to λS and $C_{21}=\frac{1}{2}\sqrt{3}$, and the linear combination with weight 0.5176 on each of the normalized logrank and modified-Wilcoxon scores (respectively 1 and $\sqrt{3}S(t)$) is the MERT score. The maximim ARE is $\frac{1}{4}(2+\sqrt{3})=0.933$. Gastwirth (1985) discussed this example fully, in the uncensored case, and proposed that the same MERT statistic be used also when censoring is present. Gastwirth and Mahmoud (1986) urged that the MERT statistic weights be optimized for the right-censoring pattern actually observed, which accomplished in estimatng Ψ by $\hat{\Psi}$. If for instance there is 20% censoring with $L_i(t)=S^{1/4}(t)$ for $i=1,2$, then our MERT score is $1/\sqrt{\hat{\Psi}(\infty)}+S_{KM}(t)/\sqrt{\int S_{KM}^2\,\mathrm{d}\hat{\Psi}}$, yielding a test-statistic asymptotically equivalent to the ideal MERT with scores $0.5136(\sqrt{5}+S(t)\sqrt{13})$, and has maximin ARE $=0.948$. The presence of equal censoring in the two groups seems to increase the maximin ARE: when $L_i(t)=L(t)$ for $i=0,1$, so that $\Psi=\gamma(1-\gamma)SL$, the maximin ARE is at least 0.933 in all cases we have tried, with L an increasing function of $S(t)$. The ARE is probably lower than in the uncensored case for some unequal censoring patterns.

In Examples 3.2–3.4, our choices for U include cases which are increasing, decreasing, and 'bathtub-shaped' functions of S as well as log-hazard-ratios which are peaked or change sign, although one would not ordinarily insist on adequate power against all of these alternatives in any single application. The maximin ARE's decrease rather rapidly with the number m of alternatives in $\mathbb{T}$, reflecting our belief that MERT is useful in customizing weighted-logrank tests mainly when there are a few clear local alternatives to discriminate from the null hypothesis.

Example 3.2. Let $U_1 \equiv 1$, $U_2(t) = (-1/\sqrt{2})\ln^-(S(t)/\mathrm{d})$, $U_3(t) = (-1/\sqrt{2})\ln^-((1-S(t))/d)$, where $\ln^-(x) \equiv \min\{0, \ln(x)\}$, and where d is a constant parameter between 0.5 and 1. When $d > 0.5$, the optimal coefficient-vector a has $a_1 = 0$, $a_2 = a_3$, and the optimized ARE ranges from 0.589 when $d = 1$ down to 0.528 when $d = 0.8$ and 0.5 when $d = 0.5$. If there is no censoring, the MERT statistic has scores proportional to

$$-\{\ln^-(S^{KM}(t)/d) + \ln^-((1-S^{KM}(t))/d)\}$$

for each of these alternative-families indexed by $d \in (0.5, 1]$, and there is no constant component in the MERT scores.

Example 3.3 (*J, L, and bathtub-shaped hazards*). Consider U_1, U_2, U_3 as in the previous example with $d = 1$ and 'bathtub-shaped' score

$$U_4(t) = 0.6783 \max\left\{1, \frac{-1}{\sqrt{2}}\ln(S(t)), \frac{-1}{\sqrt{2}}\ln(1-S(t))\right\}. \tag{3.1}$$

In the uncensored case, the correlation matrix $\boldsymbol{C}$ for $m = 4$, is given by

$$\begin{pmatrix} 1 & 0.707 & 0.707 & 0.911 \\ 0.707 & 1 & 0.178 & 0.766 \\ 0.707 & 0.178 & 1 & 0.766 \\ 0.911 & 0.766 & 0.766 & 1 \end{pmatrix}.$$

The MERT statistic of Example 3.2 (for $d = 1$) is still MERT here, and the maximin ARE is again 0.589. Another choice for $U_4(t)$ (still with the same $U_i(t)$ for $i = 1, 2, 3$, and $m = 4$) is

$$U_5(t) = 2.4495\{1 - 5S(t)(1 - S(t))\}. \tag{3.2}$$

The alternative log-hazard-ratio (3.2) yields stochastically ordered survival distributions but is bathtub-shaped and dips below 0 when $|S(t) - 0.5| < 0.224$. In the uncensored case, the correlations between U_5 and U_i for $i = 1, \ldots, 4$ are 0.408, 0.529, 0.529, and 0.693. The MERT coefficient vector for $m = 4$, $\mathbb{T} = \{U_1, U_2, U_3, U_5\}$ is $(0, 0.582, 0.582, 0.147)$, with maximin ARE $= 0.582$. This MERT statistic is still MERT and has the same maximin ARE for $m = 5$, $\mathbb{T} = \{U_1, U_2, U_3, U_4, U_5\}$.

Example 3.4 (*Alternatives with peaked and crossing hazards*). We next conduct a MERT analysis for finite-dimensional alternative families with normalized log-hazard ratios $U_i(t)$ either of the form $U_{\beta\rho}(t) = C_{\beta\rho}(S(t) - \beta)(S(t))^\rho$, for specific choices of ρ in $(0, 1)$ and β in $(0, \rho/(1+\rho))$, or of the form $V_{\gamma\delta}(t) = c_{\gamma\delta}(1 - S(t))^\gamma (S(t))^\delta$ for positive γ and δ. The first form has crossing hazards with alternatives subject to stochastic ordering (i.e., $S_1(t) \geqslant S_0(t)$ for all t) and the second form allows peaked log-hazard ratios, as in Slud (1992).

One case, with $m=4$, based on the scores $U_{0,0}(t)$, $U_{2,2}(t)$, $U_{1,3}(t)$, and $V_{0.67,2}(t)$, leads to correlation matrix

$$\begin{pmatrix} 1 & 0.837 & 0.794 & 0.249 \\ 0.837 & 1 & 0.791 & -0.102 \\ 0.794 & 0.791 & 1 & -0.194 \\ 0.249 & -0.102 & -0.194 & 1 \end{pmatrix}.$$

with MERT ARE of 0.394 and vector $\boldsymbol{a}=(0, 0.247, 0.581, 0.765)$. The second and third of these alternatives have peaks in the region where the fourth log-hazard crosses from positive to negative, and the optimal coefficients $\boldsymbol{a}$ heavily weight such anomalous alternatives.

A better-behaved example is given, for $k=5$, by $U_{0,0}(t)$, $U_{3,1}(t)$, $U_{2,2}(t)$, $U_{1,3}(t)$, and $V_{0.33,1}(t)$, with correlation matrix

$$\begin{pmatrix} 1 & 0.794 & 0.837 & 0.794 & 0.628 \\ 0.794 & 1 & 0.791 & 0.4 & 0.760 \\ 0.837 & 0.791 & 1 & 0.791 & 0.376 \\ 0.794 & 0.4 & 0.791 & 1 & 0.095 \\ 0.628 & 0.760 & 0.376 & 0.095 & 1 \end{pmatrix}.$$

The MERT ARE is 0.548, with $\boldsymbol{a}$ vector$=(0,0,0,0.676,0.676)$. Again the optimal weight-vector heavily emphasizes the anomalous peaked and sign-switching alternatives, but the ARE is now higher.

Example 3.5. Tests which arise in other ways may have an interpretation as MERT procedures with restricted action space. Consider the case of contiguous Lehmann alternatives in the (right-censored) two-sample problem $\mathbf{H}_{n,U}$ with $U \equiv 1$. Against this single alternative, the MERT is the usual logrank test, i.e. the weighted-logrank test with constant score-function. Now restrict the allowable action space of weighted-logrank tests to those with weights proportional to $W(t) = Y(t)\,(Y_1(t)\,Y_0(t))^{-1}\,I_{[t \geqslant r]}$ for some $r>0$. The allowable test statistics are then proportional to

$$\int_0^r \frac{Y(s)}{Y_1(s)\,Y_0(s)}\left(\mathrm{d}N_1(s) - \frac{Y_1(s)}{Y(s)}\,\mathrm{d}N(s)\right) = \int_0^r \left\{\frac{\mathrm{d}N_1(s)}{Y_1(s)} - \frac{\mathrm{d}N_0(s)}{Y_0(s)}\right\}, \tag{3.3}$$

which are the differences at fixed times r between the Nelson-Aalen cumulative hazard estimators for groups 1 and 0 and yield tests asymptotically equivalent (by simple application of the 'delta method') to the test based upon the difference $S_1^{KM}(r) - S_0^{KM}(r)$ of groupwise Kaplan–Meier estimators at r. The restricted-action-space MERT versus $\mathbf{H}_{n,1}$ is obtained by minimizing the estimated asymptotic type-II error probability over r. This minimization was studied in Slud (1992) to obtain 'Best-Precedence Tests' for two-sample censored survival data. In the uncensored case, the optimal r turns out to be the 0.797 quantile of the empirical distribution function of the pooled two-sample data. In the presence of censoring, Slud (1992) provides an estimator for the optimal r, as well as a proof that the resulting test with estimated r is asymptotically equivalent to the test based on (3.3) with the ideal MERT constant r. In the uncensored case, this Best-Precedence Test has power only 0.65 against Lehmann alternatives compared with the logrank test, but in not-too-exotic alternatives with functions U which are very peaked or change sign (subject to the restriction to 'stochastically ordered' alternatives), the Best-Precedence Test behaves well.

4. Bayesian and pseudo-Bayesian approach to MERT

Suppose that one has a finite family of contiguous alternatives, indexed by $s \in \{1, 2, \ldots, m\}$, with difference from the null hypothesis measured by a common scaling factor $t > 0$. The alternative ('state of Nature') is given by $\vartheta_{s,n} = \vartheta_0 + tv_s/\sqrt{\ }$, *where* $v_s \in L$ is a normalized element of $\mathbb{T}$. The action-space consists of tests of size α based on score statistics $w' \nabla_\vartheta \log f_n(X; \vartheta_0)$, where $w = \sum_{i=1}^m a_i v_i$ for some $a \in \mathbb{R}^m$ such that $\boldsymbol{a}'C\boldsymbol{a} = 1$. The action-space of coefficient vectors $\boldsymbol{a}$ is identified with the surface of an ellipsoid in $\mathbb{R}^m$.

A natural decision-theoretic formulation would be to take as loss function the probability of type-II error. For one-sided tests,

$$L(s, \boldsymbol{a}) = \Phi\left(z_\alpha - t \frac{\boldsymbol{a}'Ce_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}} \right) \tag{4.1}$$

is a monotone decreasing function of $h(s, \boldsymbol{a}) \equiv \boldsymbol{a}'Ce_s/\sqrt{\boldsymbol{a}'C\boldsymbol{a}}$. Hence the MERT action, the vector $\boldsymbol{a}$ at which $\max_{\boldsymbol{a} \in \mathbb{R}^m} \min_{1 \leqslant s \leqslant m} h(s, \boldsymbol{a})$ is attained, is minimax.

For two-sided tests the loss function is

$$L(s, \boldsymbol{a}) = \Phi\left(z_{\alpha/2} - t \frac{\boldsymbol{a}'Ce_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}} \right) - \Phi\left(-z_{\alpha/2} - t \frac{\boldsymbol{a}'Ce_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}} \right),$$

which can be written as $g(h(s, \boldsymbol{a}))$, where h is as above and $g(x) \equiv \Phi(z_{\alpha/2} - tx) - \Phi(-z_{\alpha/2} - tx)$ is an even function of x, decreasing in $|x|$ and so possessing a unique maximum at $x = 0$. The two-sided MERT procedure which maximizes $\min_s |h(s, \boldsymbol{a})|$ over $\boldsymbol{a}$ is minimax.

The decision-theoretic formulation here differs from that of an early paper of Birnbaum and Laska (1967). Those authors derived Locally Most Powerful Rank

Tests (LMPRT) for an empirical-Bayes or mixture model based on local alternatives to a *composite* null hypothesis. In their nonparametric setting, the LMPRT's are closely related to the 'pseudo-Bayesian' procedures considered below.

4.1. Bayesian approach

In this context, a Bayesian approach to MERT uses randomized strategies in the Game Against Nature defined in the Introduction, with loss (4.1). A prior on the possible states of nature is a probability vector $\boldsymbol{p}=(p_1, \ldots, p_m)'$, i.e., a vector such that $\boldsymbol{p}\geqslant 0$ and $\boldsymbol{p}'\mathbf{1}=1$. For one-sided tests, the risk is

$$r(\boldsymbol{p},\boldsymbol{a})=E_p(L(s,\boldsymbol{a}))=\sum_{s=1}^{m} p_s \Phi\left(z_\alpha - t\frac{\boldsymbol{a}'C\boldsymbol{e}_s}{\sqrt{\boldsymbol{a}'C}}\right). \tag{4.2}$$

For two-sided tests

$$r(\boldsymbol{p},\boldsymbol{a})=\sum_{s=1}^{m} p_s\left\{\Phi\left(z_\alpha - t\frac{\boldsymbol{a}'C\boldsymbol{e}_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}}\right)-\Phi\left(-z_\alpha - t\frac{\boldsymbol{a}'C\boldsymbol{e}_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}}\right)\right\}.$$

The Bayes action $\boldsymbol{a}$ minimizes $r(\boldsymbol{p},\boldsymbol{a})$ over $\boldsymbol{a}$, and can be found with standard numerical minimization programs.

4.2. Pseudo-Bayesian approach

For one-sided tests, the type II error probability under the alternative $\theta_{s,n}$ is a monotone decreasing function of $h(s,\boldsymbol{a})\equiv\boldsymbol{a}'C\boldsymbol{e}_s/\sqrt{\boldsymbol{a}'C\boldsymbol{a}}$. For the decision problem with loss function $L_1(s,\boldsymbol{a})=1-h(s,\boldsymbol{a})$, the MERT procedure is still clearly minimax. If $\boldsymbol{p}$ is the vector of prior probabilities,

$$r_1(\boldsymbol{p},\boldsymbol{a})=E_p(L_1(s,\boldsymbol{a}))=\sum_{s=1}^{m} p_s\left(1-\frac{\boldsymbol{a}'C\boldsymbol{e}_s}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}}\right)=1-\frac{\boldsymbol{a}'C\boldsymbol{p}}{\sqrt{\boldsymbol{a}'C\boldsymbol{a}}}$$

The Cauchy–Schwarz inequality implies that $\min_a r_1(\boldsymbol{p},\boldsymbol{a})=r_1(\boldsymbol{p},\boldsymbol{p})$.

By Taylor expansion of $r(\boldsymbol{p},\boldsymbol{a})$ with respect to t, as t approaches 0 the solution $\boldsymbol{a}$ becomes proportional to $\boldsymbol{p}$. Thus $\boldsymbol{p}$ is a reasonable starting point in the numerical minimization over $\boldsymbol{a}$ of (4.2).

For two-sided tests, with loss-function $L_2(s,\boldsymbol{a})=1-|h(s,\boldsymbol{a})|$, the MERT procedure is still minimax. The pseudo-Bayesian solution $\boldsymbol{a}$ based on this loss is not simple. However, a good starting point for numerical search is still $\boldsymbol{p}$, since it minimizes the first-order Taylor expansion in t of the risk.

4.3. Examples of Bayes solutions

We examine the results of Bayesian analyses of Examples 3.2 and 3.4. Calculations are carried out for one-sided tests using (4.2) with $\alpha=0.05$. In each example a finite set of alternatives and corresponding correlation matrix $\boldsymbol{C}$ is given. As in Section 4.1, for any given probability vector $\boldsymbol{p}$ of prior weights on the alternatives, one can find the corresponding Bayes action $\boldsymbol{B}$. We adopt the following approach to studying robustness with respect to misspecification of $\boldsymbol{p}$. Suppose that $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$ are two priors, with respective Bayes actions $\boldsymbol{B}^{(1)}$ and $\boldsymbol{B}^{(2)}$. A small value for the risk $r(\boldsymbol{p}^{(1)}, \boldsymbol{B}^{(2)})$ of using action $\boldsymbol{B}^{(2)}$ when the true prior is $\boldsymbol{p}^{(1)}$, as compared with the Bayes risk $r(\boldsymbol{p}^{(1)}, \boldsymbol{B}^{(1)}) \equiv \min_{\boldsymbol{a}} r(\boldsymbol{p}^{(1)}, \boldsymbol{a})$, is an indicator of robustness. Tables of risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ are provided for several prior vectors $\boldsymbol{p}^{(i)}$, and values of the risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)})$ achieved by using the pseudo-Bayes action $\boldsymbol{B}=\boldsymbol{p}^{(j)}$ under prior $\boldsymbol{p}^{(i)}$ are displayed, for comparison with the risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{a}^M)$ of using the MERT actions $\boldsymbol{a}^M$ under priors $\boldsymbol{p}^{(i)}$.

In Section 4.1, the scaling-factor t measures distance from the alternative hypotheses to the null. In the following examples, $t=3.0$ was chosen to yield $\min_{\boldsymbol{a}} r(\boldsymbol{p}, \boldsymbol{a})$ of roughly 0.2, or power of 0.8, for each prior vector $\boldsymbol{p}$. Table 1 uses the setting of Example 3.2, indicating both a high degree of robustness over misspecified prior weights and a good degree of approximation of the risks using pseudo-Bayes actions to the risks using Bayes solutions. In Table 1, $\boldsymbol{p}^{(1)}=(0.2, 0.3, 0.5)$ $\boldsymbol{p}^{(2)}=(0.3, 0.2, 0.5)$, and $\boldsymbol{p}^{(3)}=(0.5, 0.3, 0.2)$.

Table 2 used the second correlation-matrix $\boldsymbol{C}$ of Example 3.4 with $m=5$, and prior weight-vectors $\boldsymbol{p}^{(1)}=(0.3, 0.3, 0.1, 0.2, 0.1)$, $\boldsymbol{p}^{(2)}=(0.3, 0.1, 0.1, 0.3, 0.2)$, $\boldsymbol{p}^{(3)}=(0.1, 0.3, 0.1, 0.2, 0.3)$, $\boldsymbol{p}^{(4)}=(0.1, 0.3, 0.2, 0.3, 0.1)$, $\boldsymbol{p}^{(5)}=(0.1, 0.3, 0.2, 0.1, 0.3)$, $\boldsymbol{p}^{(6)}=(0.1, 0.2, 0.3, 0.3, 0.1)$, $\boldsymbol{p}^{(7)}=(0.3, 0.1, 0.3, 0.1, 0.2)$, $\boldsymbol{p}^{(8)}=(0.2, 0.1, 0.3, 0.1, 0.3)$. The results in this table can be compared with the risks under the uniform prior $\boldsymbol{p}^{(9)}$: in this case the risk $r(\boldsymbol{p}^{(9)}, \boldsymbol{B}^{(9)})=0.204$, $r(\boldsymbol{p}^{(9)}, \boldsymbol{a}^M)=0.230$, $r(\boldsymbol{p}^{(9)}, \boldsymbol{p}^{(9)})=0.205$, and the risks

Table 1
Risks (4.2) under correlations of Example 3.2

$\boldsymbol{a}$	$\boldsymbol{B}^{(1)}$ $(\boldsymbol{p}^{(1)})$	$\boldsymbol{B}^{(2)}$ $(\boldsymbol{p}^{(2)})$	$\boldsymbol{B}^{(3)}$ $(\boldsymbol{p}^{(3)})$	$\boldsymbol{a}^M$
$\boldsymbol{p}^{(1)}$	0.2233 (0.2259)	0.2263 (0.2355)	0.2434 (0.2518)	0.2307
$\boldsymbol{p}^{(2)}$	0.2028 (0.2008)	0.2000 (0.2027)	0.2308 (0.2383)	0.2183
$\boldsymbol{p}^{(3)}$	0.1976 (0.2079)	0.2080 (0.2244)	0.1816 (0.1828)	0.1935

Each $(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ cell contains the risk of using the Bayes action $\boldsymbol{B}^{(j)}$, and in parentheses the risk $r(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)})$ of using the pseudo-Bayes action $\boldsymbol{p}^{(j)}$. The last column gives risks for the MERT action $\boldsymbol{a}^M=(0, 0.652, 0.652)$

Table 2
Risks (4.2) under correlations of Example 3.4

a	1	2	3	4	5	6	7	8	a^M
$p^{(1)}$	0.181 (0.182)	0.187	0.193	0.184	0.203	0.189	0.185	0.192	0.213
$p^{(2)}$	0.214	0.208 (0.209)	0.227	0.217	0.245	0.219	0.217	0.227	0.221
$p^{(3)}$	0.234	0.240	0.220 (0.221)	0.249	0.223	0.260	0.225	0.221	0.248
$p^{(4)}$	0.202	0.208	0.226	0.199 (0.200)	0.239	0.200	0.210	0.223	0.243
$p^{(5)}$	0.221	0.232	0.201	0.238	0.198 (0.199)	0.251	0.209	0.201	0.243
$p^{(6)}$	0.201	0.205	0.230	0.194	0.246	0.193 (0.194)	0.210	0.226	0.243
$p^{(7)}$	0.188	0.193	0.190	0.195	0.195	0.201	0.185 (0.186)	0.187	0.212
$p^{(8)}$	0.214	0.219	0.203	0.228	0.205	0.237	0.205	0.202 (0.202)	0.230

Each $(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ cell contains the risk of using the Bayes action $\boldsymbol{B}^{(j)}$, and below in parentheses when $j=i$ the risk $r(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)})$ of using the pseudo-Bayes action $\boldsymbol{p}^{(j)}$. The final column gives risks for the MERT action $\boldsymbol{a}^M = (0, 0, 0, 0.676, 0.676)$

$r(\boldsymbol{p}^{(9)}, \boldsymbol{B}^{(j)})$ for $j=1, \ldots, 8$ range from 0.205 for $j=7$ to 0.219 for $j=5$. The differences among $r(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ as j varies are now not as small as in Table 1, ranging up to 0.068. As one might expect with as many as five disparate alternatives, the risk does depend noticeably on which is the correct vector of prior weights. Moreover, the MERT action $\boldsymbol{a}^M$ is now sometimes (for example, when the correct prior is $\boldsymbol{p}^{(4)}$, $\boldsymbol{p}^{(5)}$, $\boldsymbol{p}^{(6)}$, or $\boldsymbol{p}^{(7)}$) much worse than the Bayes action, indeed as bad as the worst Bayes actions for misspecified priors among $\boldsymbol{p}^{(i)}$.

Another observation from Tables 1 and 2, and many other similar calculations not presented here, is that for risks (= type II error-probabilities) in the range 0.15–0.25, there is very small difference, almost always no more than 0.005, between risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ when Bayes actions are used and the corresponding risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)})$ for pseudo-Bayes actions, whether the prior is correctly specified ($i=j$) or misspecified ($i \neq j$). With only a very few exceptions, the pseudo-Bayes actions lead to slightly higher risks, as might be expected because Bayes actions have a tendency to shrink away from extremes of the simplex of probability vectors.

We conclude our Bayesian discussion of examples by considering somewhat large t (or power closer to 1) in Example 3.4. In Table 2, we saw for that example that risks for the pseudo-Bayes actions gave excellent approximations to the risks for the corresponding Bayes actions. When t is taken somewhat larger than the value of 3.0 used in that table, the risks $r(\boldsymbol{p}^{(i)}, \boldsymbol{p}^{(j)})$ can diverge somewhat from $r(\boldsymbol{p}^{(i)}, \boldsymbol{B}^{(j)})$ in relative terms, although not very much in absolute terms. For example, the pair of values $(r(\boldsymbol{p}^{(1)}, \boldsymbol{B}^{(1)}), r(\boldsymbol{p}^{(1)}, \boldsymbol{p}^{(1)}))$ which we found to agree up to three decimal places in Table 2 when $t=3.0$, has the values (0.0172, 0.0188) for $t=4.6$, (0.0099, 0.0114) for $t=4.9$, and (0.0010, 0.0015) for $t=6.0$, and other Bayes vs. pseudo-Bayes comparisons show much the same pattern with increasing t. In the last three cases, the risks $r(\boldsymbol{p}^{(1)}, \boldsymbol{a}^M)$ are respectively 0.0225, 0.0130, and 0.0012. Only for t so large that power is of order 0.999, do pseudo-Bayes actions approximate Bayes sufficiently badly to perform worse than MERT.

5. Summary and discussion

Maximin Efficiency Robust Tests are properly understood decision-theoretically as the minimax hypothesis tests for a specified asymptotic significance level within an action space of score-test statistics with loss function defined by the large-sample asymptotic power, where the state-of-Nature parameter is the parameter value according to which the underlying data is distributed. This formulation is applicable, as in Sections 2, and 3, in the presence of high-dimensional nuisance parameters. Bayesian approaches to the same decision problem were described in Section 4.

MERT procedures are appealingly customized to have good power in specific problems against the small set of contiguous alternatives deemed most important for particular applications. From this point of view, they deserve to become a standard element of the applied statistician's repertoire. See Burnett et al. (1989) for discussion of MERT, with dose replacing time in the dose–response experiments, as the independent variable for optimal score tests.

In the examples of Section 3 involving two-sample right-censored survival data, the maximin ARE can decrease rather rapidly with the number of distinct contiguous alternatives against which good power is required. A tentative conclusion is that most useful applications of MERT will involve no more than five contiguous alternatives, although the MERT statistic will have good behavior against other alternatives of interest which lie in the positive cone generated by these alternatives. However, as Gastwirth (1985), Zucker and Lakatos (1990), and Zucker (1992) have pointed out, there are special infinite-dimensional families of (contiguous) alternatives with respect to which the minimum ARE of a MERT test will still be reasonably good (say, 60–70%). Although Zucker and Lakatos (1990) and Zucker (1992) succeeded in finding the exact MERT tests for special problems, the techniques of those papers do not work generally. A very good approximation to the MERT solution will usually be given painlessly by specifying a reasonable prior distribution on the collection of alternatives and, in the spirit of our 'pseudo-Bayesian' approach in Section 4, choosing as test-statistic the prior-weighted average of normalized score statistic optimal for the alternatives.

With the examples of Section 3 in hand, we tentatively recommend the following as the proper role of MERT in survival analysis applications. The investigator should define a small family of alternatives against which adequate power is desired and find the MERT procedure (or Bayes or pseudo-Bayes procedure, if prior weights are available). If serious differences between conventional (logrank or modified-Wilcoxon) and MERT test p-values arise in analyzing particular datasets, perhaps both analyses need to be published!

Acknowledgment

We are grateful to Mike Boyle for very useful discussions involving the connection between MERT and completely positive matrices; to Tom Louis for the Burnett, Krewski, and Bleuer (1989) reference; and to the referees for constructive comments.

References

Birnbaum, A. and E. Laska (1967). Efficiency robust two-sample rank tests. *J. Amer. Statist. Assoc.* **62**, 1241–1251.

Burnett, R., D. Krewski and S. Bleuer (1989). Efficiency robust score tests for rodent tumorigenicity experiments. *Biometrika* **76**, 317–324.

Gastwirth, J. (1966). On robust procedures. *J. Amer. Statist. Assoc.* **61**, 929–948.

Gastwirth, J. (1985). The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J. Amer. Statist. Assoc.* **80**, 380–384.

Gill, R. (1980). *Censoring and Stochastic Integrals*. Math. Centrum, Amsterdam.

Grone, R., D. Hershkowitz, N. Kogan and T. Ando (1991). *Completely Positive Graphs*. Institute for Mathematics and Applications Conference on Linear Algebra, November 1991, Minneapolis.

Hajek, J. and Z. Sidak (1967). *Theory of Rank Tests*. Academic Press, New York.

Hall, M. (1967). *Combinatorial Theory*. Ginn, Blaisdell, Waltham.

Ibragimov, I. and R. Hasminskii (1981). *Statistical Estimation*. Springer, New York.

Maxfield, J. and H. Minc (1962/63). On the matrix equation $X'X = A$. *Proc. Edinburgh Math. Soc* **13**, 125–129.

Pollard, D. (1986). *Convergence of Stochastic Processes*. Springer, New York.

Rao, C.R. (1973). *Linear Statistical Inference and Applications*, 2nd ed. Wiley, New York.

Rosen, J. (1960). The gradient projection method for nonlinear programming, Part I: Linear constraints. *J. Soc. Ind. Appl. Math.* **8**.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Silvey, S.D. (1970). *Statistical Inference*. Penguin, Baltimore, MD.

Slud, E. (1992). Best precedence tests for right-censored data. *J. Statist. Plann. Inference* **31**, 283–293.

Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.

Wolfe, P. (1959). The simplex method for quadratic programming. *Econometrica* **27**, 382–398.

Zucker, D. and E. Lakatos (1990). Weighted linear rank statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* **77**, 853–862.

Zucker, D. (1992). The efficiency of a weighted log rank test under a percent error misspecification model for the log hazard ratio. *Biometrics* **48**, 893–900.