HIGH DIMENSIONAL NUISANCE PARAMETERS: AN EXAMPLE FROM PARAMETRIC SURVIVAL ANALYSIS

By

H. S. Battey
Department of Mathematics, Imperial College London, SW7 2AZ, UK

and

D. R. Cox
Nuffield College, University of Oxford, OX1 1NF, UK

ABSTRACT

Parametric statistical problems involving both large amounts of data and models with many parameters raise issues that are explicitly or implicitly differential geometric. When the number of nuisance parameters is comparable to the sample size, alternative approaches to inference on interest parameters treat the nuisance parameters either as random variables or as arbitrary constants. The two approaches are compared in the context of parametric survival analysis, with emphasis on the effects of misspecification of the random effects distribution. Notably, we derive a detailed expression for the precision of the maximum likelihood estimator of an interest parameter when the assumed random effects model is erroneous, recovering simply derived results based on the Fisher information in the correctly specified situation but otherwise illustrating complex dependence on other aspects. Methods of assessing model adequacy are given. The results are both directly applicable and illustrate general principles of inference when there is a high-dimensional nuisance parameter. Open problems with an information geometrical bearing are outlined.

Some key words. Conditional likelihood; Exponential distribution; Marginal likelihood; Matched pairs; Model comparison; Poisson process; Random effects.

1 Introduction

Statistical analysis when the number of unknown parameters is comparable with the number of independent observations may demand modification of maximum-likelihood-based methods (Bartlett, 1937). There are comparable difficulties with Bayesian analyses based on high dimensional "flat" priors. For an extreme example from a different perspective, see Stein (1956).

Yates (1935, 1936) has discussed these issues in depth both for factorial experiments and also for variety trials in connection with balanced and partially balanced incomplete block designs. His development, powerful and almost explanation free, hinges, especially for incomplete block designs, on the geometry of least squares and the distinction between error-estimating and effect-estimating subspaces. Qualitatively similar forms of argument implicitly underlie the present paper.

Later discussion of these issues has mostly been either in general terms (Barndorff-Nielsen and Cox, 1994, chapter 2) or has approached them from a more decision-oriented perspective (e.g. Tibshirani, 1996). In the present paper we show the considerations involved in the context of parametric analysis of matched pair survival data. Matched pair designs leading to a large number of nuisance parameters have been considered in various contexts, in particular by Cox (1958), Anderson (1970), Lindsay (1980), Kumon and Amari (1984) and Kartsonaki and Cox (2016). A key aspect is the way the potentially large number of nuisance parameters are represented. One is by a probability distribution parametrically specified. The second is as a set of unknown constants and the third is as independent and identically distributed

random variables with totally unknown distribution. The consequences of the last two are essentially identical; note that the second would be converted into the third by reordering the data at random. By contrast, if appropriate the stronger assumptions involved in the parametric formulation lead to formally more precise conclusions. We illustrate the considerations involved with a theoretical and empirical analysis of the effect of misspecification. Assessment of model adequacy is also discussed. The results aim both to be directly applicable and to illustrate general principles.

2 Issues of formulation

Consider the comparison of two treatments in a matched pair design. For each of n pairs of individuals, one of the pair is a control and the other receives a treatment, leading to observations of survival times for the ith pair represented by random variables C_i, T_i. We study analyses based on underlying exponential distributions, that is that the observations are in effect the first point events in individual Poisson processes. Study of the systematic variation between treatment and control is in general complicated by variation between pairs.

There are a number of ways to represent this simple situation. We specify them in terms of the rate parameter of the underlying Poisson processes, that is the reciprocal of the exponential means. The two key components specify the relation between C_i and T_i and the form of the inter-pair variation.

For a given pair, the Poisson rate under the treatment may be a constant multiple of that under the control. Alternatively the two rates may have a constant difference. There are other possibilities such as that the two mean survival times differ by a constant. The first two representations at least have a clear underlying interpretation in terms of a potential generating process and we largely concentrate on those.

In the formulation in terms of ratios, the rate parameters of C_i and T_i are written γ_i/ψ and $\gamma_i\psi$, and in the additive formulation are written $\rho_i - \Delta$ and $\rho_i + \Delta$. Thus γ_i and ρ_i are responsible for the inter-pair variation whereas ψ and Δ are key parameters of interest for understanding the effect of the treatment. There is a clear constraint on the parameter space in the second model and the two representations are in a formal sense rather similar to logistic and additive models for binary data.

To represent in general terms arbitrary systematic variation between pairs of individuals we either treat γ_i or ρ_i as constants, unknown parameters specific to each pair, or as realizations of random variables. The conceptual differences are considerable although the numerical implications are often minor when the sample is large.

An approach sometimes used in observational studies for which there is no natural pairing involves matching individuals based on the combination of a large number of background variables into a one-dimensional propensity score (Rosenbaum and Rubin, 1983). If background variables are available and not too numerous we favour using them directly for detailed interpretation. By contrast, the present paper focusses on situations in which component variables are not separately observed.

3 Exponential matched pairs with proportional rates

3.1 Nuisance parameters as arbitrary constants

For the representation involving ratios of rates let $Z_i = T_i/C_i$, removing dependence on γ_i. The density at z is

$$\psi^2/(1+\psi^2 z)^2. \tag{1}$$

Standard maximum likelihood theory based on the marginal distribution of the Z_i applies. In particular, the maximum likelihood estimator of ψ based on (1) is consistent and asymptotically normally distributed with variance given by the inverse of the Fisher information.

The Fisher information per observation is

$$(2 + 2 - 8/3)\psi^{-2} = (4/3)\psi^{-2}. \tag{2}$$

By eliminating the nuisance parameters in this way by marginalization, some information on the interest parameter is in general lost, because $(Z_1, \ldots, Z_n) = S$, say, is not sufficient for ψ. Further discussion of these issues is given in section 7.2. A smaller variance is achievable at the expense of stronger modelling assumptions, as demonstrated in section 3.2.

3.2 Nuisance parameters as random variables

Instead of regarding the pair effects as constants we now suppose that they are random variables independently gamma distributed of shape parameter α and rate β. Then the joint density function of T_i and C_i at (t, c) is

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \gamma^{\alpha+1} \exp\{-\gamma(\psi t + c/\psi + \beta)\} d\gamma = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{\beta^\alpha}{(\psi t + c/\psi + \beta)^{\alpha+2}}. \tag{3}$$

The Fisher information matrix per observation can be shown (see Appendix A.2) to be block diagonal with the relevant entry for inference about ψ equal to

$$\frac{2(\alpha+2)}{(\alpha+3)\psi^2}. \tag{4}$$

The two limits of this as $\alpha \to \infty$ and $\alpha \to 0$ are $2\psi^{-2}$ and $(4/3)\psi^{-2}$, the latter being (2), the Fisher information per observation obtained by treating the nuisance parameters as arbitrary constants. The variance depends on the relative dispersion of the nuisance parameters through α.

See section 7.1 for a formulation in terms of unobserved covariates involving a log normal distribution over the γ_i.

Equation (4) shows that the gamma random effects formulation is more efficient than the one in which nuisance parameters are treated as arbitrary constants, provided that the random effects specification is reasonable. The modelling assumption is more severe, but the following analysis of the misspecified situation shows that, provided ψ is bounded away from zero, the corresponding maximum likelihood estimator $\hat{\psi}$ obtained by assuming the gamma random effects model of section 3.2, converges almost surely to ψ as $n \to \infty$. Thus $\hat{\psi}$ remains consistent in spite of an arbitrary degree of misspecification in the assumed random effects distribution.

Let γ_i $(i = 1, \ldots, n)$ be independent random variables with an arbitrary density function $f(\gamma)$. The associated joint distribution of T_i and C_i satisfies (see Appendix A.3)

$$E\Big\{\frac{T_i}{(T_i\psi + C_i/\psi + \beta)^j}\Big\} = \frac{1}{\psi^2} E\Big\{\frac{C_i}{(T_i\psi + C_i/\psi + \beta)^j}\Big\} \quad (j = 1, 2, 3, \ldots). \tag{5}$$

In view of the expressions for the cross partial derivatives of the log likelihood function (equation (28) in Appendix A.2), equation (5) establishes orthogonality of ψ to α and β whatever the random effects distribution. The interpretation of the notional parameters α and β under model misspecification is discussed below. The orthogonality justifies consideration of the marginal maximum likelihood estimating equation for ψ, i.e.

$$0 = \frac{1}{n}\sum_{i=1}^n \ell_{i,\psi}(\hat{\psi}) = \frac{1}{n}\sum_{i=1}^n \frac{C_i}{\hat{\psi}^2(T_i\hat{\psi} + C_i/\hat{\psi} + \beta)} - \frac{1}{n}\sum_{i=1}^n \frac{T_i}{T_i\hat{\psi} + C_i/\hat{\psi} + \beta}. \tag{6}$$

For any $\kappa > 0$ bounded away from zero, consider

$$\frac{1}{n}\sum_{i=1}^n \frac{C_i}{\kappa^2(T_i\kappa + C_i/\kappa + \beta)} - \frac{1}{n}\sum_{i=1}^n \frac{T_i}{T_i\kappa + C_i/\kappa + \beta}. \tag{7}$$

Under the random effects formulation, the summands are independent and identically distributed and a law of large numbers implies convergence of the averages to their expectations. The limiting value of the maximum likelihood estimator, as $n \to \infty$, is the value of κ that equalizes the two expectations. Appendix A.4 shows that the expectations exist and the value of κ that equalizes them is ψ. Thus $\hat{\psi}$ is consistent despite the misspecification.

An analysis of efficiency is harder. Let g_{θ^*} denote the density function of the true joint distribution of (T_i, C_i), where $\theta^* = (\lambda, \psi)$ and λ could be a finite or infinite dimensional nuisance parameter, but the proportional rates model of section 2 is assumed so that ψ captures the treatment effect. This joint density is determined by the marginal density function of the random effects distribution $f(\gamma)$ as

$$g_{\theta^*}(t, c) = \int_0^\infty \gamma^2 f(\gamma) \exp\{-\gamma(t\psi + c/\psi)\} d\gamma. \tag{8}$$

Thus if f is not parameterized, λ is f itself. Let Θ denote the parameter space for the erroneous gamma random effects model and let $f_\theta(x, y)$ denote the misspecified joint density function of each (T_i, C_i) at (x, y), given by equation (3). Thus we may define $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\psi})$ by $\text{argmax}_{v \in \Theta} \sum_{i=1}^n \log f_v(T_i, C_i)$, which converges almost surely (Appendix A.1) to

$$\theta = (\theta_1, \theta_2, \theta_3) = \underset{v \in \Theta}{\text{argmin}} \int_0^\infty \int_0^\infty \log \frac{g_{\theta^*}(x, y)}{f_v(x, y)} g_{\theta^*}(x, y) dx dy, \tag{9}$$

where, from the previous derivations, $\theta_3 = \psi$, the true treatment effect. Thus $\alpha = \theta_1$ and $\beta = \theta_2$ are the values that minimize the Kullback-Leibler divergence between the assumed (erroneous) model and the true model.

By the orthogonality established in (5), a discussion of efficiency requires consideration of the likelihood derivatives only with respect to ψ. In particular, by the established consistency, a mean value expansion and standard arguments, it can be shown that the asymptotic distribution of $n^{1/2}(\hat{\psi} - \psi)$ is Gaussian of zero mean and variance $[E\{\ell_{i,\psi\psi}(\theta)\}]^{-2} E\{\ell_{i,\psi}^2(\theta)\}$, leading to a variance of $R/(R-Q)^2$, where R and Q depend in a rather complicated way on the density function $f(\gamma)$ of the true random effects distribution. Specifically

$$\begin{aligned} R &= \frac{1}{\psi^2}\Big\{\frac{1}{3} - \frac{2\beta}{3}E(\gamma_i) - \frac{\beta^2}{3}E(\gamma_i^2) \\ &- \beta^2 \int_0^\infty \gamma^2 f(\gamma) e^{\gamma\beta} \text{Ei}(-\gamma\beta) d\gamma - \frac{\beta^3}{3} \int_0^\infty \gamma^3 f(\gamma) e^{\gamma\beta} \text{Ei}(-\gamma\beta) d\gamma \Big\}, \\ Q &= \frac{1}{\psi^2}\Big\{1 + \beta^2 \int_0^\infty \gamma^2 f(\gamma) e^{\gamma\beta} \text{Ei}(-\gamma\beta) d\gamma - \beta E(\gamma_i)\Big\}. \end{aligned} \tag{10}$$

Here $\text{Ei}(x)$ is the exponential integral (Olver, 1974, equation 3.07) thus, in equation (10),

$$\text{Ei}(-\gamma\beta) = -\int_{\gamma\beta}^\infty z^{-1} e^{-z} dz, \quad \gamma\beta > 0,$$

and because the γ_i are treated as totally random, $E(\gamma_i^\kappa) = \int_0^\infty \gamma^\kappa f(\gamma) d\gamma$.

In a correctly specified situation, $\{E(\ell_{i,\psi\psi})\}^{-2} E(\ell_{i,\psi}^2)$ is the inverse Fisher information. When the random effects are gamma distributed of parameter α and rate β, as assumed, $Q = 2(\alpha+2)^{-1}\psi^{-2}$ and $R = 2(\alpha+3)^{-1}(\alpha+2)^{-1}\psi^{-2}$ so that $R/(R-Q)^2$ is $2^{-1}(\alpha+2)^{-1}(\alpha+3)\psi^2$, i.e. the reciprocal of equation (4). While formula (10) does not seem amenable to detailed interpretation under misspecification, it serves to illustrate complicated dependence on key aspects of the formulation.

Table 4 of section 6.2 shows that the loss of efficiency in the gamma model for random effects can be severe when the sample size is not large and when the random effects distribution is misspecified. Thus, while the random effects formulation is in principle always feasible for nuisance parameter problems, the adequacy of the choice of random effects distribution, often made on the basis of mathematical convenience, needs consideration. A discussion in the context of the present example is in section 5.

4 Exponential matched pairs with additive rates

When the nuisance parameters ρ_i of the additive treatment effects model (see section 2) are treated as arbitrary constants, the inference is based on conditioning on the sufficient statistic for the nuisance parameter in each pair (Kartsonaki and Cox, 2016). We extend their results slightly by giving explicit expressions for the conditional and unconditional variances of the estimator. The likelihood contribution from the ith pair is

$$(\rho_i^2 - \Delta^2)\exp\{-\rho_i(t_i + c_i)\}\exp\{-\Delta(t_i - c_i)\}. \tag{11}$$

Thus $T_i + C_i$ is sufficient for ρ_i and this leads to inference based on the difference $T_i - C_i$, or equivalently T_i given the pairwise totals $T_i + C_i = S_i$, say. The density function of S_i at s is

$$(\rho_i^2 - \Delta^2)\{e^{-(\rho_i - \Delta)s} - e^{-(\rho_i + \Delta)s}\}/(2\Delta). \tag{12}$$

Some algebra shows that the conditional density function of T_i at t given $S_i = s_i$ is, for $\Delta > 0$,

$$\frac{2\Delta e^{-2\Delta t}}{1 - e^{-2\Delta s_i}}. \tag{13}$$

Let $\hat{\Delta}$ denote the maximum likelihood estimator of Δ based on the conditional density function (13). The Fisher information of $\hat{\Delta}$, conditional on $S_i = s_i$ is

$$\frac{n}{\Delta^2} - 4\sum_{i=1}^{n}\frac{s_i^2 e^{-2\Delta s_i}}{(1 - e^{-2\Delta s_i})^2} = \frac{n}{\Delta^2} - \textstyle\sum_{i=1}^{n} s_i^2 \sinh^{-2}(\Delta s_i), \tag{14}$$

where $s\sinh^{-1}(\Delta s) < \Delta^{-1}$ for all $s > 0$ and $\lim_{s\to 0}\{s\sinh^{-1}(\Delta s)\} = \Delta^{-1}$ so that the conditional Fisher information is non-negative. For planning, the unconditional Fisher information is relevant. This is used for determining the sample size required to achieve a pre-specified conditional efficiency with high probability, and is obtained by replacing the ith summand by

$$\begin{aligned} & \frac{(\rho_i^2 - \Delta^2)}{2\Delta}\int_0^\infty \frac{s^2(e^{\Delta s} - e^{-\Delta s})\exp(-\rho_i s)}{\sinh^2(\Delta s)}ds \\ = \quad & \frac{(\rho_i^2 - \Delta^2)}{\Delta}\int_0^\infty \frac{s^2\exp(-\rho_i s)}{\sinh(\Delta s)}ds = \frac{(\rho_i^2 - \Delta^2)}{4\Delta^4}\int_0^\infty \frac{t^2 e^{-qt}}{1 - e^{-t}}dt, \end{aligned} \tag{15}$$

where $q = (\rho_i + \Delta)/(2\Delta)$ and in the last line we have changed variables to $t = 2\Delta s$. The integral and summation representations of Riemann's generalized zeta function are (Whittaker and Watson, 1927, p265–66)

$$\zeta(z, q) = \frac{1}{\Gamma(z)}\int_0^\infty \frac{t^{z-1}e^{-qt}}{1 - e^{-t}}dt = \sum_{m=0}^{\infty}\frac{1}{(q + m)^z},$$

and the unconditional Fisher information is, from (15),

$$\frac{n}{\Delta^2} - \frac{1}{2\Delta^4}\sum_{i=1}^{n}(\rho_i^2 - \Delta^2)\zeta\{3, (\rho_i + \Delta)/(2\Delta)\}. \tag{16}$$

Section 6.1 confirms the above calculations by simulation.

Among other possibilities, the pair effects might be assumed to have a gamma distribution of parameter α and rate β starting at Δ, leading to a joint density function of T_i and C_i at (t, c) given by

$$\frac{\alpha\beta^\alpha\exp(-2\Delta t)}{(t + c + \beta)^{\alpha+1}}\Big\{\frac{\alpha + 1}{t + c + \beta} + 2\Delta\Big\}. \tag{17}$$

Standard maximum likelihood theory applies when the random effects distribution is correctly specified. An analysis of misspecification of this model is complicated by the fact that the parameters α and β are not orthogonal to Δ under arbitrary misspecification. Thus a full theoretical analysis of the kind developed in section 3.2 will not be explored for the maximum likelihood estimator $\tilde{\Delta}$ based on equation (17). However Table 5 of section 6.2 provides numerical evidence that severe loss of efficiency can result, relative to the version that treats the nuisance parameters as arbitrary constants. Consistency of $\tilde{\Delta}$ is also suspect. A referee asked whether there is any mathematically convenient distribution for the nuisance parameters that results in orthogonality of the nuisance parameters to the interest parameter Δ in the additive rates model in spite of possible misspecification. In principle, if the true distribution of T_i and C_i is known and given in terms of parameters (Δ, α, β), say, a reparameterization to (Δ, λ, η), say, can always be found such that λ and η are orthogonal to Δ. This entails solving the pair of differential equations

$$\begin{aligned} i^*_{\alpha\alpha}\frac{\partial\alpha(\Delta,\lambda,\eta)}{\partial\Delta} + i^*_{\beta\alpha}\frac{\partial\beta(\Delta,\lambda,\eta)}{\partial\Delta} &= -i^*_{\Delta\alpha} \\ i^*_{\alpha\beta}\frac{\partial\alpha(\Delta,\lambda,\eta)}{\partial\Delta} + i^*_{\beta\beta}\frac{\partial\beta(\Delta,\lambda,\eta)}{\partial\Delta} &= -i^*_{\Delta\beta}, \end{aligned}$$

initially to determine the dependence of α and β on Δ and ultimately choosing λ and η as detailed by Cox and Reid (1987). However, in the above display $i^*_{\alpha\beta}$, $i^*_{\Delta\beta}$, etc. are the expectations of the second cross partial derivatives of the assumed loglikelihood function, taken with respect to the true model. These expressions differ depending on the form of misspecification. An extension of the ideas of Cox and Reid (1987) to accommodate arbitrary misspecification is an important question which demands further study, ideally in full generality.

5 Assessment of model adequacy

In the above two models, exact tests of model adequacy are available. Sufficiency represents a separation of the information in the data into that relevant for estimating the parameters of a given model and that relevant for assessing the adequacy of the model (Barndorff-Nielsen and Cox, 1994, p.29). Suppose that the proportional treatment effect model of section 3 holds. The likelihood contribution from the ith pair is

$$\gamma_i^2 \exp(-\gamma_i c_i/\psi)\exp(-\gamma_i \psi t_i). \tag{18}$$

From this, for any given ψ, $C_i/\psi + T_i\psi = S_i(\psi)$, say, is sufficient for γ_i and has density function

$$f_{S_i(\psi)}(s) = \gamma_i^2 s \exp(-\gamma_i s), \tag{19}$$

i.e., $S_i(\psi)$ is gamma distributed with shape parameter 2 and rate parameter γ_i.

The model and an arbitrarily specified parameter value $\psi = \psi_0$ are jointly compatible with the data if the realization of T_i, say, is not extreme relative to the conditional density function of T_i given $S_i(\psi) = s_i(\psi)$, assuming $\psi = \psi_0$. The conditional density of T_i at t_i, given $S_i(\psi) = s_i(\psi)$, is

$$\frac{\gamma_i^2 \exp\{-\gamma_i s_i(\psi)\}}{\gamma_i^2 s_i(\psi)\exp\{-\gamma_i s_i(\psi)\}} = \frac{1}{s_i(\psi)}, \tag{20}$$

showing that $T_i \mid \{S_i(\psi) = s_i(\psi)\}$ is uniformly distributed between 0 and $s_i(\psi)$.

For any hypothesized value ψ_0 of ψ, compatibility of the proportional treatment effects model and ψ_0 with the data corresponds to compatibility of the realizations of $T_i/s_i(\psi_0) = U_i(\psi_0)$, say, with a uniform distribution on (0,1) for all $i = 1, \ldots, n$. This is a basis for checking consistency with the proportional model. More specifically, an α-level confidence set using Fisher's (1925, section 21.1) test is

$$\mathcal{C}(\alpha) \triangleq (\psi_0 \in \Psi : \min[F\{-2\textstyle\sum_{i=1}^n \log U_i(\psi_0)\}, 1 - F\{-2\sum_{i=1}^n \log U_i(\psi_0)\}] < \alpha), \tag{21}$$

where F is the distribution function of a χ^2 random variable with $2n$ degrees of freedom. If the confidence set is non-empty at a specified level, there are at least some values of ψ_0 for which the proportional treatment effects model is compatible with the data at this level.

For sufficiently large sample size, one might treat $\hat{\psi}$ as fixed and equal to ψ under the null hypothesis that the model is true. The adequacy of this assumption can then be assessed by checking the compatibility of the realizations of $T_i/s_i(\hat{\psi})$ for $i = 1, \ldots, n$ with a uniform distribution on $(0, 1)$.

The same ideas allow the adequacy of the a random effects model to be checked. In particular, for any given ψ, the collection of weighted sums $S_i(\psi)$ for $i = 1, \ldots, n$ is sufficient for the nuisance parameters α and β, as can be seen from equation (3). One could condition as above.

For sufficiently many pairs, however, a simpler option is available due to the small number of nuisance parameters in the random effects model. The distribution function at s of $S_i = T_i + C_i$ under the gamma random effects model is given by

$$1 - \frac{\beta^\alpha}{\psi^2 - 1} \left\{ \frac{\psi^2}{(\beta + s/\psi)^\alpha} - \frac{1}{(\beta + s\psi)^\alpha} \right\}. \tag{22}$$

Since the maximum likelihood estimators $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\psi}$ are consistent and completely specify the model, for sufficiently many individuals it may often be a reasonable approach to consider these as fixed and equal to the true values α, β and ψ under the null hypothesis that the gamma random effects model is correctly specified. Making this replacement in equation (22) and evaluating the distribution function at the points S_i for $i, \ldots, n$ leads to approximately standard uniformly distributed points under the null hypothesis, and Fisher's (1925, section 21.1) test is applicable.

Similar arguments apply to the additive effects model. Section 4 shows that $S_i = T_i + C_i$ is sufficient for the nuisance parameter ρ_i, so that the conditional density of T_i given $S_i = s_i$ is free of ρ_i and is given by equation (13). In section 4, this justified estimation of the treatment effect Δ by maximization of the conditional likelihood based on (13). To assess model adequacy it is necessary to condition on the jointly sufficient statistic for all unknown parameters. Thus, as in the proportional rates formulation, one must fix Δ at hypothesized values leading to a joint assessment of the adequacy of the additive effects model at an arbitrary but given value Δ_0 of the interest parameter. The model and a value Δ_0 are compatible with the data at a particular level if $T_1, \ldots, T_n$ are not extreme relative to what would be expected under their joint conditional density assuming $\Delta = \Delta_0$, i.e.,

$$\prod_{i=1}^{n} f_{T_i|S_i=s_i}(z_i; \Delta_0) = \prod_{i=1}^{n} \frac{2\Delta_0 e^{-2\Delta_0 z_i}}{1 - e^{-2\Delta_0 s_i}}.$$

As in the proportional rates model, For sufficiently large sample size, one might reasonably treat $\hat{\Delta}$ as fixed and equal to Δ under the null hypothesis that the additive rates model is true and proceed as above using $\hat{\Delta}$ in place of Δ_0 to assess model adequacy.

There are situations where exact tests of model adequacy based on these principles do not seem feasible. One example in the spirit of this work would be an exponential matched pair problem in which T_i and C_i have a stable difference in means. In section 7.2, we explain in more general terms how the structure of the inference problem dictates the appropriate strategy.

6 Empirical validation and numerical extensions

6.1 Fixed nuisance parameters

Throughout the following numerical work $\psi = \Delta = 2$. For several different values of n we generate $(\gamma_i)_{i=1}^n$ from a gamma distribution of shape $\alpha = 1$ and rate $\beta = 1$, and we define

$\rho_i = \Delta + \gamma_i$ so that $\rho_i - \Delta > 0$. The nuisance parameters $(\gamma_i)_{i=1}^n$ and $(\rho_i)_{i=1}^n$ are then fixed over Monte Carlo replications.

In each of $R = 1000$ Monte Carlo replications, $T_i^{(\mathrm{PR})}$ and $C_i^{(\mathrm{PR})}$ $(i = 1, \ldots, n)$ are generated independently from exponential distributions of rates $\gamma_i\psi$ and γ_i/ψ respectively, and $T_i^{(\mathrm{AR})}$ and $C_i^{(\mathrm{AR})}$ are generated from exponential distributions of rates $\rho_i + \Delta$ and $\rho_i - \Delta$. The parameter ψ in the proportional rates model is estimated by maximum likelihood based on the density function of $T_i^{(\mathrm{PR})}/C_i^{(\mathrm{PR})}$ of equation (1). Let $\hat{\psi}_n$ denote this estimator.

The sample variance of $\hat{\psi}_n$ over the 1000 Monte Carlo replications is reported in the second row of Table 1, with an estimate of its theoretical standard error in the third row. This is based on the χ^2 distribution with $R-1$ degrees of freedom of the sample variance. The theoretical variance of $\hat{\psi}_n$ is asymptotically (as $n \to \infty$) the inverse of the Fisher information. Its theoretical value obtained from equation (2) is reported below the row of standard errors. The values in the second and the fourth rows agree for large n.

We also report the results from fitting a gamma random effects model to $T_i^{(\mathrm{PR})}, C_i^{(\mathrm{PR})}$ for $i = 1, \ldots, n$. Let $\tilde{\psi}_n$ denote the corresponding maximum likelihood estimator of ψ. This model is misspecified but the efficiency of $\tilde{\psi}_n$ is high. However the model is not severely misspecified because the $(\gamma_i)_{i=1}^n$ are generated from a gamma distribution before being fixed across Monte Carlo replications. In section 6.2, we consider the effect of more severe misspecification of the random effects distribution.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated bias of $\hat{\psi}_n$	0.0217	0.0212	0.0078	0.0035	0.0039	0.0089	-0.0040
Simulated variance of $\hat{\psi}_n$	0.1513	0.0759	0.0486	0.0365	0.0258	0.0192	0.0144
Estimated standard error	0.0068	0.0034	0.0022	0.0016	0.0012	0.00087	0.00065
Inverse Fisher information for $\hat{\psi}_n$	0.1500	0.0750	0.0500	0.0375	0.0250	0.0187	0.0150
Simulated bias of $\tilde{\psi}_n$	0.0214	0.0164	0.0071	0.0038	0.0025	0.0103	-0.0036
Simulated variance of $\tilde{\psi}_n$	0.1384	0.0700	0.0435	0.0323	0.0229	0.0171	0.0125
Estimated standard error	0.0062	0.0031	0.0019	0.0014	0.0010	0.00077	0.00056

Table 1: the generating process is the proportional rates model with fixed $(\gamma_i)_{i=1}^n$. Simulated variance of the marginal maximum likelihood estimator and its estimated standard error, the associated inverse Fisher information and the simulated variance of the maximum likelihood estimator by erroneously assuming gamma random effects.

The parameter Δ from the additive rates model is estimated using maximum likelihood based on the conditional density function of $T_i^{(\mathrm{AR})}$ given the realization of $T_i^{(\mathrm{AR})} + C_i^{(\mathrm{AR})}$. This is equation (13). Let $\hat{\Delta}_n$ denote this maximum likelihood estimator. The Monte Carlo variance of $\hat{\Delta}_n$ is reported in the second row of Table 2, with its estimated theoretical standard error in in the third row. The unconditional variance based on equation (16) is reported in the fourth row together with the Monte Carlo average of the conditional variances based on (14) in the fifth row. The two agree to a close approximation and they also agree with the Monte Carlo sample variances for sufficiently large n.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated bias of $\hat{\Delta}_n$	0.1154	0.0644	0.0569	0.0448	0.0287	0.0179	0.0142
Simulated variance of $\hat{\Delta}_n$	0.3895	0.1715	0.1144	0.0868	0.0554	0.0413	0.0307
Estimated standard error	0.0174	0.0077	0.0051	0.0039	0.0025	0.0018	0.0014
Inverse Fisher information	0.3040	0.1480	0.1033	0.0766	0.0523	0.0373	0.0308
Inverse conditional Fisher info	0.3072	0.1482	0.1034	0.0768	0.0523	0.0375	0.0308

Table 2: the generating process is the additive rates model with fixed $(\rho_i)_{i=1}^n$. Simulated variance of the conditional maximum likelihood estimator and its estimated standard error,

the inverse unconditional Fisher information based on equation (16), and the Monte Carlo average of the inverse conditional Fisher information based on equation (14).

6.2 Randomly generated nuisance parameters

The simulation studies are the same as in section 6.1 except that $(\gamma_i)_{i=1}^n$ and the $(\rho_i)_{i=1}^n$ are generated anew in each Monte Carlo replication. Thus the models in which these nuisance parameters are treated as arbitrary constants are misspecified. In particular, dependence between both versions of T_i and C_i is induced by the generating mechanism for γ_i and ρ_i.

Table 3 contains analogous information to the top three rows of Table 1 for the misspecified case. The theoretically true variances have not been calculated and so are not reported. However, the sample variances are very close to the theoretical asymptotic variances that would obtain if the nuisance parameters were arbitrary constants (cf fourth row of Table 1). We also report the Monte Carlo variance of $\tilde{\psi}_n$, now under a correctly specified model, and its theoretical asymptotic variance based on equation (4). Comparing the fifth and last rows of Table 3, these agree for sufficiently large n.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated bias of $\hat{\psi}_n$	0.0435	0.0204	0.0173	0.0131	0.0080	0.0095	0.0043
Simulated variance of $\hat{\psi}_n$	0.1563	0.0772	0.0544	0.0386	0.0271	0.0194	0.0154
Estimated standard error	0.0070	0.0035	0.0024	0.0017	0.0012	0.00087	0.00069
Simulated bias of $\tilde{\psi}_n$	0.0362	0.0198	0.0146	0.0106	0.0069	0.0047	0.0041
Simulated variance of $\tilde{\psi}_n$	0.1373	0.0709	0.0455	0.0339	0.0242	0.0173	0.0139
Estimated standard error	0.0061	0.0032	0.0020	0.0015	0.0011	0.00077	0.00062
Inverse Fisher info for $\tilde{\psi}_n$	0.1333	0.0667	0.0444	0.0333	0.0222	0.0167	0.0133

Table 3: the generating process is the proportional rates model with gamma distributed (α=1, β=1) random effects $(\gamma_i)_{i=1}^n$. Simulated variances of the two estimators of ψ and the asymptotic theoretical variance of $\tilde{\psi}_n$.

To assess the efficiency of $\tilde{\psi}_n$ under fairly extreme misspecification of the random effects distribution, we conduct the same experiment but with the $(\gamma_i)_{i=1}^n$ drawn from a log normal distribution with scale parameter $\tau = 10$. For comparison, the Monte Carlo variances of $\hat{\psi}_n$ are also reported in Table 4. The conclusion from this analysis is that while $\hat{\psi}_n$, justified under the assumption that the nuisance parameters are arbitrary constants, has a stable variance when the nuisance parameters are drawn from a rather extreme random effects distribution, the variance of $\tilde{\psi}_n$ is appreciably larger when the random effects distribution is misspecified in this way.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated bias of $\hat{\psi}_n$	0.0200	0.0099	0.0150	0.00026	-0.0036	0.0119	0.0010
Simulated variance of $\hat{\psi}_n$	0.1567	0.0729	0.0498	0.0375	0.0262	0.0185	0.0145
Estimated standard error	0.0070	0.0033	0.0022	0.0017	0.0012	0.00083	0.00065
Simulated bias of $\tilde{\psi}_n$	0.1029	0.0406	0.0551	0.0139	0.0298	0.0440	0.0065
Simulated variance of $\tilde{\psi}_n$	0.4367	0.1125	0.1389	0.0593	0.1645	0.1528	0.0208
Estimated standard error	0.0195	0.0050	0.0062	0.0027	0.0074	0.0068	0.00093

Table 4: the generating process is the proportional rates model with log normally distributed (τ=10) random effects $(\gamma_i)_{i=1}^n$. Simulated variances of the two estimators of ψ.

We now consider the effect of misspecification of the random effects distribution in the additive rates model by comparing the estimator $\hat{\Delta}$ of section 4 to the maximum likelihood

estimator $\tilde{\Delta}$ obtained by erroneously assuming that the joint density function of T_i and C_i is given by equation (17). Rather than being a gamma distribution starting at Δ, the true distribution of the ρ_i is a log normal distribution of scale parameter $\tau = 10$ starting at Δ. Although the theoretical variance of $\hat{\Delta}$ has not been calculated under the random effects formulation, the ones based on equations (16) and (14) are reported in the fourth and fifth rows of Table 5. As before, the estimated standard errors in the third and eighth rows are based on a χ^2 distribution with $R-1$ degrees of freedom for the sample variance, where R is the number of Monte Carlo replications.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated bias of $\hat{\Delta}_n$	0.2320	0.1335	0.0635	0.0491	0.0560	0.0347	0.0139
Simulated variance of $\hat{\Delta}_n$	0.6640	0.2681	0.1721	0.1198	0.0815	0.0539	0.0395
Estimated standard error	0.0297	0.0120	0.0077	0.0054	0.0036	0.0024	0.0018
Inverse Fisher information	0.3961	0.1899	0.1130	0.0880	0.0769	0.0511	0.0342
Inverse conditional Fisher info	0.4291	0.2043	0.1342	0.0998	0.0661	0.0495	0.0397
Simulated bias of $\tilde{\Delta}_n$	0.3540	0.1556	0.0578	0.0693	0.1454	0.1851	0.1614
Simulated variance of $\tilde{\Delta}_n$	1.1209	0.4147	0.2209	0.2465	0.2932	0.4872	0.4455
Estimated standard error	0.050	0.019	0.0099	0.011	0.013	0.022	0.020

Table 5: the generating process is the additive rates model with log normally distributed (τ=10) random effects $(\rho_i)_{i=1}^n$ shifted by Δ. Simulated variance of the conditional maximum likelihood estimator and its estimated standard error, the inverse unconditional Fisher information based on equation (16) of the paper, and the Monte Carlo average of the inverse conditional Fisher information based on equation (14). The seventh row presents the simulated variance of $\tilde{\Delta}$, the maximum likelihood estimator of Δ by erroneously assuming a gamma distribution starting at Δ for the random effects.

6.3 Assessment of model adequacy in the proportional rates model

To illustrate the ideas in section 5 we consider the data generating process corresponding to Table 1 with $\psi = 1$. This is the value of ψ that equalises the distributions of responses for treated individuals and controls. In each of 1000 Monte Carlo replications we calculate $T_i/s_i(\psi_0) = U_i(\psi_0)$ for all ψ_0 between zero and three in increments of 0.01 and for $i = 1, \ldots, n$ with the values of n reported in Table 6. We use the composite of these values to produce a confidence set for ψ as in equation (21). Table 6 reports the simulated coverage probabilities of the α-level confidence sets for $\alpha \in \{0.01, 0.05\}$. While the confidence sets need not be intervals in general, they turned out to be intervals in all our Monte Carlo replications, thus we report the mean lower and upper boundaries of these confidence intervals, averaged over Monte Carlo replications.

	sample size (n)						
	20	40	60	80	120	160	200
Simulated coverage probability ($\alpha = 0.01$)	0.996	0.993	0.990	0.994	0.989	0.991	0.992
Simulated mean lower boundary ($\alpha = 0.01$)	0.2766	0.3529	0.3804	0.4155	0.4570	0.4805	0.4993
Simulated mean upper boundary ($\alpha = 0.01$)	2.9861	2.8872	2.7308	2.5430	2.2913	2.1542	2.0615
Simulated coverage probability ($\alpha = 0.05$)	0.957	0.942	0.956	0.959	0.949	0.946	0.953
Simulated mean lower boundary ($\alpha = 0.05$)	0.3542	0.4075	0.4402	0.4685	0.5037	0.5151	0.5387
Simulated mean upper boundary ($\alpha = 0.05$)	2.8517	2.5349	2.3250	2.1951	2.0300	1.9514	1.8763

Table 6: the generating process is the proportional rates model with fixed $(\gamma_i)_{i=1}^n$ and $\psi = 1$. Monte Carlo coverage probabilities and mean upper and lower boundaries of α-level confidence intervals constructed according to equation (21).

The interpretation of the numbers in Table 6 is that the proportional rates model with fixed nuisance parameters is compatible with the data at level α for any value of ψ_0 taking values in $\mathcal{C}(\alpha)$ defined by equation (21).

7 Discussion and open problems

7.1 A synthesis with earlier literature

The choice of random effects distribution in section 3.2 was primarily one of mathematical convenience. It coincides with typical usage in applications and raises conceptual issues: (i) To what extent is the random effects formulation a plausible representation of the data generating mechanism? (ii) Are there statistical advantages of assuming a parametric random effects model even if the formulation is physically implausible? (iii) Are there statistical advantages of treating nuisance parameters as arbitrary constants when there is a probabilistic generating mechanism for them?

Our analysis has shown the need to be wary of assumptions made for mathematical convenience with no substantive basis. The following example shows how a different distribution for the random effects may be more plausible, leading to the situation considered in Table 4. The comparison to Table 1 shows that the approach in which nuisance parameters are treated as arbitrary constants is noticeably preferable to the approach in which the incorrect parametric random effects distribution is used.

Suppose, in the notation of section 3, that one models the nuisance parameters as $\gamma_i = \exp(x_i^T\theta)$, where the x_i are covariates that one could have, but did not, measure. If individuals are sampled completely at random from a larger population, it is not unreasonable to treat the covariates as realizations of random variables X_i, assumed to be i.i.d. copies of X, a p-dimensional normally distributed random vector of mean zero and covariance matrix $\Sigma = Q\Lambda Q^T$, where Q is a matrix whose columns are the unit-length eigenvectors of Σ. To derive the induced distribution over the γ_i, write $W \triangleq \theta^T X = \theta^T Q\Lambda^{1/2}V$, where V is a standard normally distributed random vector. We have $W = \|\theta^T Q\Lambda^{1/2}\|_2\|V\|_2 R$, where R is the cosine of the angle between V and $\Lambda^{1/2}Q^T\theta$, whose density function is given by (Fisher, 1915)

$$f_R(r) = \frac{\Gamma(p/2)}{\sqrt{\pi}\Gamma\{(p-1)/2\}}(1-r^2)^{(p-3)/2}, \quad -1 < r < 1,$$

and $\|V\|_2^2$ is a chi squared random variable with p degrees of freedom, so that $D \triangleq \|V\|_2$ has density function

$$f_D(\delta) = \frac{\delta^{p-1}\exp(-\delta^2/2)}{2^{(p/2)-1}\Gamma(p/2)}, \quad \delta \geq 0.$$

The characteristic function of W is

$$\phi_W(t) = E_R\{\phi_D(\|\theta^T Q\Lambda^{1/2}\|_2 tR)\} = E_D\{\phi_R(\|\theta^T Q\Lambda^{1/2}\|_2 tD)\},$$

where for any random variable Y, $\phi_Y(t) = E_Y(e^{itY})$. Let $s = \|\theta^T Q\Lambda^{1/2}\|_2 t$. Direct calculation gives

$$\begin{aligned}\phi_W(t) &= K^{-1}\int_0^\infty\int_{-1}^1 \exp\{-\delta^2/2 + is\delta r\}\delta^{p-1}(1-r^2)^{(p-3)/2}drd\delta \\ &\simeq K^{-1}\int_0^\infty \exp\{-\delta^2/2\}\delta^{p-1}\int_{-1}^1 \exp\{is\delta r - (p/2)r^2\}drd\delta, \quad p\to\infty\end{aligned}$$

where $K = \sqrt{\pi}2^{(p/2)-1}\Gamma\{(p-1)/2\}$. Since $\int_{-1}^1 \exp\{-(p/2)r\}\sin(s\delta r)dr = 0$,

$$\begin{aligned}&\int_{-1}^1 \exp\{is\delta r - (p/2)r^2\}dr = \int_{-1}^1 \exp\{-(p/2)r^2\}\cos(s\delta r)dr \\ = \;& \frac{\exp\{-(s\delta)^2/2p\}\sqrt{2\pi}}{p^{1/2}} - \left(\int_{-\infty}^{-1} + \int_1^\infty\right)\exp\{-(p/2)r^2\}\cos(s\delta r)dr,\end{aligned}$$

and the remainder terms are ignored for $p\to\infty$, leading to

$$\phi_W(t) \simeq \frac{\{1+(s^2/p)\}^{-p/2}\Gamma(p/2)\sqrt{2}}{\Gamma\{(p-1)/2\}p^{1/2}}, \quad p\to\infty$$

Using Stirling's formula in the form $\Gamma(k+a)/\Gamma(k) \simeq k^a$ for large k,

$$\phi_W(t) \simeq (1+s^2/p)^{-p/2} \simeq e^{-s^2/2} \quad (p \to \infty),$$

where $e^{-s^2/2} = e^{-(\|\Lambda^{1/2}Q^T\theta\|_2^2/2)t^2}$ is the characteristic function of a centred normal random variable with standard deviation $\tau \triangleq \|\Lambda^{1/2}Q^T\theta\|_2$. Under this generating mechanism for the covariates, γ_i are thus log-normally distributed, with density function

$$(\tau\gamma)^{-1}\phi(\log\gamma/\tau), \tag{23}$$

where $\phi(\cdot)$ is the standard normal density.

While this formulation is to some extent physically justifiable, the integral (8) does not appear to have an analytic solution when $f(\gamma)$ is given by (23). This illustrates that random effects models are likely to be driven by mathematical convenience, highlighting the importance of studies of misspecification.

After completing this paper, we were made aware of a related contribution by Lindsay (1985). The work showed that straight maximum likelihood estimation (without preliminary manoeuvres based on the factorizability of the likelihood function) is consistent in a particular class of incidental parameter models. Specifically, those models for which there is a complete sufficient statistic $S_i(\psi)$ for the nuisance parameter λ_i, with ψ treated as fixed. This situation covers the exponential matched pairs problems with multiplicative treatment effect on the rates (section 3.1) and with additive treatment effect on the rates (section 4) but not the exponential matched pairs problem with additive treatment effect on the means. Despite consistency of the maximum likelihood estimator, the standard estimator of the variance of the maximum likelihood estimator is seriously distorted in these settings, the true variance typically being appreciably larger than that based on the supposed inverse Fisher information.

Lindsay considered estimation of the interest parameter by parametric random effects models and showed that the efficiency achievable by the resulting estimator is higher than straight maximum likelihood provided that a reasonable choice of parametric model for the random effects is used, even if this random effects distribution is misspecified. The appropriate conditions are essentially that the parameters of the parametric random effects distribution be orthogonal in the sense of Cox and Reid (1987). Parameter orthogonality arose in our derivations in sections 3.2 via equation (5) and its derivation in Appendix A.3. Lindsay (1985) does not discuss the potential for appreciable loss of efficiency over conditional or marginal likelihood, as opposed to full maximum likelihood, by erroneously assuming a parametric random effects model. This potential loss of efficiency is illustrated by our equation (10). The synthesis of Lindsay's analysis and ours is that, while a random effects formulation can lead to increased precision over straight maximum likelihood even when the random effects distribution is misspecified, provided that the parameters of the random effects distribution are orthogonal to the interest parameters, there is potential for appreciable loss of efficiency over marginal and conditional likelihood when the corresponding factorizations of the likelihood function are available.

7.2 Open problems

Issues connected with an appreciable number of nuisance parameters are likely to arise whenever a relatively complicated model is needed. In principle, analyses similar to those of sections 3 – 5 could be performed for other distributions. See Cox (1958) for a binary responses formulation that parallels the proportional rates model of section 3. Our existing work does not, however, generalize readily and the detailed calculation required for other distributional assumptions is likely to be considerable. Nevertheless, some general principles can be extracted from the previous discussion. Let ψ be an interest parameter and λ be a nuisance parameter. Either or both may be vectors. One starts from an arbitrary pair of observations (T, C), or more generally an arbitrary partition, and makes a bijective transformation $(T, C) \to (S, R)$ such that one of factorizations (i)–(v) holds, where:

(i) $f_{S,R}(s,r;\psi,\lambda) = f_{R|S}(r|s;\lambda)f_S(s;\psi)$, (iv) $f_{S,R}(s,r;\psi,\lambda) = f_{R|S}(r|s;\lambda,\psi)f_S(s;\psi)$,

(ii) $f_{S,R}(s,r;\psi,\lambda) = f_{R|S}(r|s;\psi)f_S(s;\lambda)$, (v) $f_{S,R}(s,r;\psi,\lambda) = f_{R|S}(r|s;\psi)f_S(s;\psi,\lambda)$.

(iii) $f_{S,R}(s,r;\psi,\lambda) = f_R(r;\lambda)f_S(s;\psi)$,

Factorization (i) requires marginalization with S sufficient for ψ, (ii) requires conditioning on S, which is now the sufficient statistic for λ. In (iii) the jointly sufficient statistic is two independent sufficient statistics so that conditioning reduces to marginalization. Marginalization is applicable in (iv), in which $R|S$ is sufficient for λ, and conditioning in (v), in which S is sufficient for λ, but information on ψ is lost in either case. The exponential proportional rates model and the exponential additive rates model are examples of factorizations (iv) and (ii) respectively.

Our suggestion of section 5 provides a unified approach to assessing the joint compatibility of a model and its parameter values with the data, and is justified in any situation for which one of factorizations (i)–(v) holds exactly. An important open question is the construction of appropriate factorizations, exact or approximate, in greater generality. We conclude by an outline of the considerations involved.

For an arbitrary pair (t,c) of jointly sufficient statistics, write the transformation equations as $s = s(t,c)$, and $r = r(t,c)$. The transformation is assumed to be bijective so that $t = t(s,r)$ and $c = c(s,r)$. For factorizations (i), (iii) or (iv) to be true, we require that $f_S(s;\psi,\lambda) = f_S(s;\psi)$, and similarly for (ii) and (v).

The general form of a solution to $f_S(s;\psi,\lambda) = f_S(s;\psi)$ is to express the unknown density of S in terms of the known joint density of T and C. For instance,

$$f_S(s;\psi,\lambda) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} \exp\{zs(t,c)\}T_\lambda(z)dz,$$

where τ is anywhere in the strip of convergence of the moment generating function of S and

$$T_\lambda(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-zs(x,y)\}f_{T,C}(x,y;\psi,\lambda)dxdy, \quad z \in \mathbb{C}.$$

The only contribution of λ comes from T_λ, so it is sufficient to choose the function $s(t,c)$ to make T_λ independent of λ, identically in z, ψ and λ. It would be sufficient that independence be achieved only at points z of singularity, but this is more difficult. There results the following integral equation, to be solved for $s(t,c)$, identically in z, ψ, and λ:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-zs(t,c)\}\left\{\frac{\partial}{\partial\lambda}f_{T,C}(t,c;\psi,\lambda)\right\}dtdc = 0. \tag{24}$$

In the exponential matched pair problem with proportional rates (section 3), equation (24) becomes

$$0 = \int_0^{\infty}\int_0^{\infty} \exp\{-zs(t,c)\}\left\{2\lambda - \lambda^2(\psi t + c/\psi)\right\}\exp(-\lambda\psi t)\exp(-\lambda c/\psi)dtdc. \tag{25}$$

While it is simple to show that $s(t,c) = t/c$ verifies equation (25), recovering the strategy of section 3.1, a general theory relies on a solution to the integral equation (24) when $s(t,c)$ is not known a priori.

An alternative general formulation to that based on Laplace transforms uses the joint density function of S and R. Specifically, for factorization (i), (iii) or (iv) consider

$$\begin{aligned} f_S(s;\psi,\lambda) &= \int_{\mathcal{R}} f_{S,R}(s,r;\psi,\lambda)dr \\ &= \int_{\mathcal{R}} f_{T,C}\Big\{t(s,r),c(s,r);\psi,\lambda\Big\}|\det\{J_{(T,C)\to(S,R)}\}|dr, \end{aligned}$$

where $J_{(T,C)\to(S,R)}$ is the Jacobian of the transformation $(T,C) \to (S,R)$. Thus, for the marginal density to be independent of λ, we require the solution in $t(s,r)$ and $c(s,r)$ of the set of partial integro-differential equations:

$$\begin{aligned}\int_{\mathcal{R}} \left(\frac{\partial t(s,r)}{\partial s}\frac{\partial c(s,r)}{\partial r} - \frac{\partial t(s,r)}{\partial r}\frac{\partial c(s,r)}{\partial s} \right) \frac{\partial}{\partial \lambda} f_{T,C}\Big\{ t(s,r), c(s,r); \psi, \lambda \Big\} dr &= 0, \\ \int_{\mathcal{R}} \left(\frac{\partial t(s,r)}{\partial r}\frac{\partial c(s,r)}{\partial s} - \frac{\partial t(s,r)}{\partial s}\frac{\partial c(s,r)}{\partial r} \right) \frac{\partial}{\partial \lambda} f_{T,C}\Big\{ t(s,r), c(s,r); \psi, \lambda \Big\} dr &= 0, \quad (26)\end{aligned}$$

identically in λ and ψ.

In connection with these ideas there are a number of open problems with a differential geometrical bearing:

1. When there are nuisance parameters two approaches are to transform the data and marginalize or condition based on factorizations (i)–(v) above, or to find an interest-respecting orthogonal transformation as in Cox and Reid (1987). It is natural to expect there to be a connection between the two, and for this to be characterizable geometrically.

2. Is there a geometric representation of conditioning to evade nuisance parameters, and if so, how is this different geometrically to conditioning to ensure relevance (Amari, 1982)?

3. Differential geometric treatments of asymptotic inference (e.g. Amari, 1982; 1983; Amari and Kumon, 1983; Kumon and Amari, 1983; Barndorff-Nielsen *et al.*, 1986) hinge on looking locally in the parameter space of fixed number of dimensions as the amount of information becomes so large that interest is focused on a small region. As such it does not seem directly applicable when the dimension of the parameter space is itself very large which is the situation considered in the present paper. Is there an extension of these ideas suitable for the incidental parameter problems of the present paper?

The analysis of section 3.2 also hints at a more general analysis of model misspecification. There are important open questions. For instance: when is inference on an interest parameter relatively unaffected by misspecification of the nuisance part of the model? What type of misspecification is the inference robust to and how does this depend on the structure of the model and the loss function used for estimation? In what sense is the inference robust? For instance consistency may be achievable but efficiency lost.

Acknowledgements: the work was supported by a UK Engineering and Physical Sciences Research Council Fellowship to HSB.

Conflict of interest statement: On behalf of all authors, the corresponding author states that there is no conflict of interest.

APPENDIX

A Derivations of key results

A.1 Derivation of equation (9)

The argmax is unchanged by rescaling and subtraction of constants. Dividing by n and subtracting $n^{-1}\sum_{i=1}^{n} \log g_{\theta^*}(T_i, C_i)$ shows that

$$\hat{\theta} = \underset{v\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^{n} \log \frac{f_v(T_i, C_i)}{g_{\theta^*}(T_i, C_i)}.$$

The summands are identically distributed and of finite expectations, therefore $\hat{\theta}$ converges almost surely to

$$\theta = (\theta_1, \theta_2, \theta_3) = \underset{v \in \Theta}{\operatorname{argmin}} \int_0^\infty \int_0^\infty \log \frac{g_{\theta^*}(x,y)}{f_v(x,y)} g_{\theta^*}(x,y) dx dy.$$

A.2 Derivation of equation (4)

The second derivative of the log likelihood for the ith pair with respect to ψ is

$$\begin{aligned} \ell_{i,\psi\psi} &= -(\alpha+2)\Big\{ \underbrace{\frac{2C_iT_i}{\psi^2(\psi T_i + C_i/\psi + \beta)^2}}_{\mathrm{I}_1} - \underbrace{\frac{C_i^2}{\psi^4(\psi T_i + C_i/\psi + \beta)^2}}_{\mathrm{I}_2} \\ &\quad + \underbrace{\frac{2C_i}{\psi^3(\psi T_i + C_i/\psi + \beta)}}_{\mathrm{I}_3} - \underbrace{\frac{T_i^2}{(\psi T_i + C_i/\psi + \beta)^2}}_{\mathrm{I}_4} \Big\}, \end{aligned} \tag{27}$$

and the two cross-partial derivatives with respect to ψ are

$$\ell_{i,\psi\alpha} = \frac{T_i - C_i/\psi^2}{(T_i\psi + C_i/\psi + \beta)}, \quad \ell_{i,\psi\beta} = -\frac{(\alpha+2)(T_i - C_i/\psi^2)}{(T_i\psi + C_i/\psi + \beta)^2}. \tag{28}$$

The expectations of both terms in (28) are zero because, for any κ,

$$\int_0^\infty t \left\{ \int_0^\infty (t\psi + c/\psi + \beta)^{-\kappa} dc \right\} dt = \frac{\psi}{(\kappa-1)} \int_0^\infty t(t\psi+\beta)^{-(\kappa-1)} dt, \tag{29}$$

$$\psi^{-2} \int_0^\infty c \left\{ \int_0^\infty (t\psi + c/\psi + \beta)^{-\kappa} dt \right\} dc = \frac{1}{\psi^3(\kappa-1)} \int_0^\infty c(c/\psi+\beta)^{-(\kappa-1)} dc. \tag{30}$$

Changing variables to $z = t\psi$ and $z = c/\psi$ in (29) and (30) shows that both integrals are equal to

$$\frac{1}{\psi(\kappa-1)} \int_0^\infty z(z+\beta)^{-(\kappa-1)} dz,$$

so that terms cancel when taking expectations in (28). It follows that the Fisher information matrix per observation is block diagonal with the relevant block equal to the negative expectation of (27), specifically

$$(\alpha+2)\Big\{ \underbrace{\frac{2}{(\alpha+2)(\alpha+3)\psi^2}}_{E(\mathrm{I}_1)} - \underbrace{\frac{2}{(\alpha+2)(\alpha+3)\psi^2}}_{E(\mathrm{I}_2)} + \underbrace{\frac{2}{(\alpha+2)\psi^2}}_{E(\mathrm{I}_3)} - \underbrace{\frac{2}{(\alpha+2)(\alpha+3)\psi^2}}_{E(\mathrm{I}_4)} \Big\} = \frac{2(\alpha+2)}{(\alpha+3)\psi^2}.$$

This is (4).

A.3 Derivation of equation (5)

Consider $j = 1$ and let K be the normalizing constant for the joint density of T_i and C_i. Then

$$\mathrm{E}_1 \triangleq E\Big(\frac{T_i}{T_i\psi + C_i/\psi + \beta}\Big) = \frac{1}{K} \int_0^\infty \gamma^2 f(\gamma) \Big\{ \int_0^\infty e^{-\gamma c/\psi} \Big(\int_0^\infty \frac{t e^{-\gamma t\psi} dt}{t\psi + c/\psi + \beta} \Big) dc \Big\} d\gamma.$$

Direct calculation shows that the inner integral is

$$\psi^{-2}[\gamma^{-1} + e^{\gamma c/\psi} e^{\gamma\beta} (c/\psi + \beta) \mathrm{Ei}\{-\gamma(c/\psi+\beta)\}],$$

so that, changing variables to $z = \gamma(c/\psi + \beta)$ gives

$$\begin{aligned} \mathrm{E}_1 &= \frac{1}{\psi^2 K}\int_0^\infty \gamma^2 f(\gamma)\Big\{\gamma^{-1}\int_0^\infty e^{-\gamma c/\psi}dc + e^{\gamma\beta}\psi\gamma^{-2}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big\}d\gamma \\ &= \frac{1}{\psi K}\int_0^\infty f(\gamma)\Big\{1 + e^{\gamma\beta}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big\}d\gamma. \end{aligned}$$

Now consider

$$\mathrm{E}_2 \triangleq \frac{1}{\psi^2}E\Big(\frac{C_i}{T_i\psi + C_i/\psi + \beta}\Big) = \frac{1}{\psi^2 K}\int_0^\infty \gamma^2 f(\gamma)\Big\{\int_0^\infty e^{-\gamma t\psi}\Big(\int_0^\infty \frac{ce^{-\gamma c/\psi}dc}{t\psi + c/\psi + \beta}\Big)dt\Big\}d\gamma.$$

The inner integral is

$$\psi^2[\gamma^{-1} + e^{\gamma t\psi}e^{\gamma\beta}(t\psi + \beta)\mathrm{Ei}\{-\gamma(t\psi + \beta)\}].$$

Integrating with respect to t and changing variables to $z = \gamma(t\psi + \beta)$ in the second term gives

$$\begin{aligned} \mathrm{E}_2 &= \frac{1}{K}\int_0^\infty \gamma^2 f(\gamma)\Big\{\gamma^{-1}\int_0^\infty e^{-\gamma t\psi}dt + e^{\gamma\beta}\psi^{-1}\gamma^{-2}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big\}d\gamma \\ &= \frac{1}{\psi K}\int_0^\infty f(\gamma)\Big\{1 + e^{\gamma\beta}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big\}d\gamma = \mathrm{E}_1. \end{aligned}$$

The demonstration is analogous for other $j \in \mathbb{N}$, the integrals being identical up to the ψ^2 term that arises from the same changes of variables used above.

A.4 Proof of consistency of the maximum likelihood estimator

By the argument following equation (7), it is required to show that the κ that equalizes the expectation of $C_i/\{\kappa^2(T_i\kappa + C_i/\kappa + \beta\}$ and $T_i/\{T_i\kappa + C_i/\kappa + \beta\}$ is $\kappa = \psi$, and that these expectations exist for any κ and ψ bounded away from zero.

Consider

$$\mathrm{I}_T \triangleq E\Big(\frac{T_i}{\kappa T_i + C_i/\kappa + \beta}\Big) = \frac{1}{K}\int_0^\infty \gamma^2 f(\gamma)\int_0^\infty e^{-\gamma c/\psi}\Big(\int_0^\infty \frac{te^{-\gamma t\psi}dt}{t\kappa + c/\kappa + \beta}\Big)dcd\gamma,$$

where, as before, K is the normalizing constant for the joint density function of T_i and C_i. Direct calculation shows that the innermost integral is

$$\frac{e^{\gamma(c/\kappa+\beta)\psi/\kappa}}{\kappa^2}\Big[\frac{e^{-\gamma(c/\kappa+\beta)\psi/\kappa}}{\gamma\psi/\kappa} + (c/\kappa + \beta)\mathrm{Ei}\{-\gamma(c/\kappa + \beta)\psi/\kappa\}\Big]. \tag{31}$$

Similarly,

$$\mathrm{I}_C \triangleq \frac{1}{\kappa^2}E\Big(\frac{C_i}{T_i\kappa + C_i/\kappa + \beta}\Big) = \frac{1}{\kappa^2 K}\int_0^\infty \gamma^2 f(\gamma)\int_0^\infty e^{-\gamma t\psi}\Big(\int_0^\infty \frac{ce^{-\gamma c/\psi}dc}{t\kappa + c/\kappa + \beta}\Big)dtd\gamma,$$

and the innermost integral is

$$\kappa^2 e^{\gamma(\kappa t+\beta)\kappa/\psi}\Big[\frac{e^{-\gamma(\kappa t+\beta)\kappa/\psi}}{\gamma\kappa/\psi} + (\kappa t + \beta)\mathrm{Ei}\{-\gamma(\kappa t + \beta)\kappa/\psi\}\Big]. \tag{32}$$

Changing variables to $z = (c/\kappa + \beta)$ in (31) and $s = (\kappa t + \beta)$ in (32) shows that

$$\begin{aligned} \mathrm{I}_T &= \frac{1}{\kappa K}\int_0^\infty \gamma^2 f(\gamma)e^{\gamma\beta\kappa/\psi}\int_\beta^\infty e^{-\gamma z\kappa/\psi}e^{\gamma z\psi/\kappa}\Big[\frac{e^{-\gamma z\psi/\kappa}}{\gamma\psi/\kappa} + z\mathrm{Ei}(-\gamma z\psi/\kappa)\Big]dzd\gamma \\ \mathrm{I}_C &= \frac{1}{\kappa K}\int_0^\infty \gamma^2 f(\gamma)e^{\gamma\beta\psi/\kappa}\int_\beta^\infty e^{-\gamma s\psi/\kappa}e^{\gamma s\kappa/\psi}\Big[\frac{e^{-\gamma s\kappa/\psi}}{\gamma\kappa/\psi} + s\mathrm{Ei}(-\gamma s\kappa/\psi)\Big]dsd\gamma. \end{aligned}$$

If both these integrals exist, the limit of $\hat{\psi}$ is the unique value of κ that sets $\mathrm{I}_T = \mathrm{I}_C$, i.e. $\kappa = \psi$. Since the exponential integral $\mathrm{Ei}(-x)$ is negative for $x > 0$, I_T and I_C are both upper bounded by $(\kappa K)^{-1}\int_0^\infty f(\gamma)d\gamma = (\kappa K)^{-1} < \infty$ for all κ bounded away from zero. This justifies the previous use of the a strong law of large numbers. Thus $\hat{\psi}$ converges almost surely to ψ.

A.5 Derivation of equation (10)

The squared derivative with respect to ψ of the likelihood contribution from the ith pair is

$$\ell_{i,\psi}^2 = \frac{(\alpha+2)^2(t-c/\psi^2)^2}{(t\psi+c/\psi+\beta)^2}.$$

Taking expectations, $E(\ell_{i,\psi}^2) = \mathrm{T}_1 + \mathrm{T}_2 - \mathrm{T}_3$, where

$$\begin{aligned}
\mathrm{T}_1 &= (\alpha+2)^2\int_0^\infty \gamma^2 f(\gamma)\int_0^\infty t^2 e^{-\gamma t\psi}\int_0^\infty \frac{e^{-\gamma c/\psi}dc}{(t\psi+c/\psi+\beta)^2}dtd\gamma,\\
\mathrm{T}_2 &= \frac{(\alpha+2)^2}{\psi^4}\int_0^\infty \gamma^2 f(\gamma)\int_0^\infty c^2 e^{-\gamma c/\psi}\int_0^\infty \frac{e^{-\gamma t\psi}dt}{(t\psi+c/\psi+\beta)^2}dcd\gamma,\\
\mathrm{T}_3 &= \frac{2(\alpha+2)^2}{\psi^2}\int_0^\infty \gamma^2 f(\gamma)\int_0^\infty t e^{-\gamma t\psi}\int_0^\infty \frac{ce^{-\gamma c/\psi}dc}{(t\psi+c/\psi+\beta)^2}dtd\gamma.
\end{aligned}$$

Consider T_1. A change of variables to $z = (t\psi + c/\psi + \beta)$ leads to

$$\int_0^\infty \frac{e^{-\gamma c/\psi}dc}{(t\psi+c/\psi+\beta)^2} = \psi e^{\gamma\beta}e^{\gamma t\psi}\Big[\frac{e^{-\gamma(t\psi+\beta)}}{t\psi+\beta} + \gamma\mathrm{Ei}\{-\gamma(t\psi+\beta)\}\Big].$$

The term $e^{\gamma t\psi}$ cancels with $e^{-\gamma t\psi}$ so that the relevant integrals with respect to t are

$$\begin{aligned}
\int_0^\infty \frac{t^2 e^{-\gamma(t\psi+\beta)}}{t\psi+\beta}dt &= \psi^{-3}[\{\gamma^{-2} - (\beta/\gamma)\}e^{-\gamma\beta} - \beta^2\mathrm{Ei}(-\gamma\beta)]\\
\gamma\int_0^\infty t^2\mathrm{Ei}\{-\gamma(t\psi+\beta)\} &= \psi^{-3}\Big[\gamma^{-2}\int_{\gamma\beta}^\infty z^2\mathrm{Ei}(-z)dz\\
&\quad - \beta^2\int_{\gamma\beta}^\infty \mathrm{Ei}(-z)dz - 2\beta\gamma^{-1}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big].
\end{aligned}$$

Integration by parts shows that $\int z^{b-1}\mathrm{Ei}(-z)dz = b^{-1}\{z^b\mathrm{Ei}(-z) + \Gamma(b,z)\}$ and there is the recursive formula $\Gamma(b+1,z) = b\Gamma(b,z) + z^b e^{-z}$ so that $\Gamma(1,\gamma\beta) = e^{-\gamma\beta}$ and

$$\begin{aligned}
\int_{\gamma\beta}^\infty z^2\mathrm{Ei}(-z)dz &= -\frac{1}{3}[(\gamma\beta)^3\mathrm{Ei}(-\gamma\beta) + \{(\gamma\beta)^2 + 2(\gamma\beta) + 2\}e^{-\gamma\beta}]\\
\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz &= -\frac{1}{2}[(\gamma\beta)^2\mathrm{Ei}(-\gamma\beta) + (\gamma\beta+1)e^{-\gamma\beta}]\\
\int_{\gamma\beta}^\infty \mathrm{Ei}(-z)dz &= -\{\gamma\beta\mathrm{Ei}(-\gamma\beta) + e^{-\gamma\beta}\}.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
\mathrm{T}_1 &= \frac{(\alpha+2)^2}{\psi^2}\Big[\frac{1}{3} - \frac{2\beta}{3}E(\gamma_i) - \frac{\beta^2}{3}E(\gamma_i^2)\\
&\quad - \beta^2\int_0^\infty \gamma^2 f(\gamma)e^{\gamma\beta}\mathrm{Ei}(-\gamma\beta)d\gamma - \frac{\beta^3}{3}\int_0^\infty \gamma^3 f(\gamma)e^{\gamma\beta}\mathrm{Ei}(-\gamma\beta)d\gamma\Big]
\end{aligned}$$

and an analogous calculation shows that $\mathrm{T}_2 = \mathrm{T}_1$.

Consider T_3. The inner integrals are

$$\begin{aligned}\int_0^\infty \frac{ce^{-\gamma c/\psi}}{(t\psi + c/\psi + \beta)^2} dc &= -\psi^2 e^{\gamma t\psi} e^{\gamma\beta}\Big[\mathrm{Ei}\{-\gamma(t\psi+\beta)\} \\ &+ e^{-\gamma(t\psi+\beta)} + \gamma(t\psi+\beta)\mathrm{Ei}\{-\gamma(t\psi+\beta)\}\Big],\end{aligned}$$

and

$$\begin{aligned}\int_0^\infty te^{-\gamma t\psi}\int_0^\infty \frac{ce^{-\gamma c/\psi}dc}{(t\psi+c/\psi+\beta)^2}dt = -e^{\gamma\beta}\Big[\gamma^{-2}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz - \beta\gamma^{-1}\int_{\gamma\beta}^\infty \mathrm{Ei}(-z)dz \\ + \gamma^{-2}e^{-\gamma\beta} + \gamma^{-2}\int_{\gamma\beta}^\infty z^2\mathrm{Ei}(-z)dz - \gamma^{-1}\beta\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dz\Big]\end{aligned}$$

Using the previous expression for the integrals of the $\mathrm{Ei}(z)$ functions, we obtain

$$\begin{aligned}\mathrm{T}_3 &= \frac{(\alpha+2)^2}{\psi^2}\Big[\frac{1}{3} - \frac{2\beta}{3}E(\gamma_i) - \frac{\beta^2}{3}E(\gamma_i^2) \\ &- \beta^2\int_0^\infty \gamma^2 f(\gamma)e^{\gamma\beta}\mathrm{Ei}(-\gamma\beta)d\gamma - \frac{\beta^3}{3}\int_0^\infty \gamma^3 f(\gamma)e^{\gamma\beta}\mathrm{Ei}(-\gamma\beta)d\gamma\Big].\end{aligned}$$

Thus $\mathrm{T}_1 = \mathrm{T}_2 = \mathrm{T}_3$ and $E\{\ell_{i,\psi}^2\} = \mathrm{T}_1$. In the correctly specified case this is the Fisher information. On replacing $f(\gamma)$ by $\beta^\alpha\gamma^{\alpha-1}e^{-\gamma\beta}/\Gamma(\alpha)$ in the expression for T_1 we obtain the result from section 3.2, namely $2(\alpha+2)/\psi^2(\alpha+3)$.

For the calculation of $E(\ell_{i,\psi\psi})$, it is required to calculate the expectations of the terms $\mathrm{I}_1 - \mathrm{I}_4$ in equation (27), under misspecification. It is clear from their expressions that these expectations are related to the above calculations in the following way: $E(\mathrm{I}_4) = (\alpha+2)^{-2}\mathrm{T}_1$, $E(\mathrm{I}_2) = (\alpha+2)^{-2}\mathrm{T}_2$ and $E(\mathrm{I}_1) = (\alpha+2)^{-2}\mathrm{T}_3$, but $\mathrm{T}_1 = \mathrm{T}_2 = \mathrm{T}_3$ so that

$$E(\ell_{i,\psi\psi}) = -(\alpha+2)(E\mathrm{I}_1 - E\mathrm{I}_2 + E\mathrm{I}_3 - E\mathrm{I}_4) = (\alpha+2)^{-1}\mathrm{T}_1 - (\alpha+2)E\mathrm{I}_3.$$

The missing expectation is

$$\begin{aligned}Q \triangleq E(\mathrm{I}_3) &= \frac{2}{\psi^2}\Big\{1 + \int_0^\infty f(\gamma)e^{\gamma\beta}\int_{\gamma\beta}^\infty z\mathrm{Ei}(-z)dzd\gamma\Big\} \\ &= \frac{1}{\psi^2}\Big\{1 + \beta^2\int_0^\infty \gamma^2 f(\gamma)e^{\gamma\beta}\mathrm{Ei}(-\gamma\beta)d\gamma - \beta E(\gamma_i)\Big\}.\end{aligned}$$

On writing $R = (\alpha+2)^{-2}\mathrm{T}_1$, it follows that

$$\{E(\ell_{i,\psi\psi})\}^{-2}E(\ell_{i,\psi}^2) = \frac{\mathrm{T}_1}{(\alpha+2)^2\{\mathrm{T}_1(\alpha+2)^{-2} - Q\}^2} = \frac{R}{(R-Q)^2}.$$

Under the correct specification of the gamma random effects model we also obtain

$$E(\ell_{i,\psi\psi}) = (\alpha+2)^{-1}\mathrm{T}_1 - (\alpha+2)Q = 2\psi^{-2}\{(\alpha+3)^{-1} - 1\} = -\frac{2(\alpha+2)}{\psi^2(\alpha+3)},$$

as expected.

A.6 Derivation of equation (12)

The Laplace transform of the density of S_i at z is

$$E(e^{-zS_i}) = \frac{\rho_i^2 - \Delta^2}{(\rho_i + \Delta + z)(\rho_i - \Delta + z)}$$

and the density function of each S_i at s is

$$\begin{aligned} f_{S_i}(s) &= \text{Res}\left\{\frac{\exp(zs)(\rho_i^2-\Delta^2)}{(\rho_i+\Delta+z)(\rho_i-\Delta+z)}, -(\rho_i+\Delta)\right\} \\ &+ \text{Res}\left\{\frac{\exp(zs)(\rho_i^2-\Delta^2)}{(\rho_i+\Delta+z)(\rho_i-\Delta+z)}, -(\rho_i-\Delta)\right\} \\ &= (\rho_i^2-\Delta^2)\{e^{-(\rho_i-\Delta)s}-e^{-(\rho_i+\Delta)s}\}/2\Delta, \end{aligned}$$

where for a function $g(z)$ $z \in \mathbb{C}$, $\text{Res}\{g, a\}$ denotes the residue of g at $z = a$.

REFERENCES

Amari, S-I. (1982). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika*, 69, 1–17.

Amari, S-I. and Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family, *Ann. Inst. Statist. Math.*, 35, 1–24.

Amari, S-I. (1983). *Differential geometry of statistical inference*, Springer, Berlin.

Anderson, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. R. Statist. Soc. B*, 32, 283–301.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A*, 160, 268–82.

Barndorff-Nielsen, O. E., Cox, D. R. and Reid, N. M. (1986). The role of differential geometry in statistical theory. *International Statistical Review*, 54, 83–96.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall, London.

Cox, D. R. (1958). Two further applications of a model for binary regression.*Biometrika*, 45, 562–65.

Cox, D. R. and Reid, N. M. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, 49, 1–39.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Kartsonaki, C. and Cox, D. R. (2016). Some matched comparisons of two distributions of survival time. *Biometrika*, 103, 219–24.

Kumon, M. and Amari, S-I. (1983). Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference. *Proc. Roy. Soc. London Ser. A*, 387, 429–458.

Kumon, M. and Amari, S-I. (1984). Estimation of a structural parameter in the presence of a large number of nuisance parameters. *Biometrika*, 71, 445–59.

Lindsay, B. G. (1980). Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Phil. Trans. R. Soc. Lond.*, 296, 639–65.

Lindsay, B. G. (1985). Using empirical partially Bayes inference for increased efficiency. *Ann. Statist.*, 13, 914–31.

Olver, F. W. J. (1974). *Introduction to Asymptotics and Special Functions.* Academic Press, New York.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. In *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 197–206.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *J. R. Statist. Soc. B*, 58, 267–88.

Whittaker, E.T. and Watson, G.N. (1927). *A Course of Modern Analysis.* (1965 reprint of the fourth edition). Cambridge University Press, London.

Yates, F. (1935). Complex experiments, *J. R. Statist. Soc., B* (with discussion), 2, 181–223.

Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.*, 26, 424–55.