Exact prior-free probabilistic inference on the heritability coefficient in a linear mixed model

Qianshun Cheng, Xu Gao and Ryan Martin

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607, USA
e-mail: qcheng5@uic.edu*;* xgao21@uic.edu*;* rgmartin@uic.edu

Abstract: Linear mixed-effect models with two variance components are often used when variability comes from two sources. In genetics applications, variation in observed traits can be attributed to biological and environmental effects, and the heritability coefficient is a fundamental quantity that measures the proportion of total variability due to the biological effect. We propose a new inferential model approach which yields exact prior-free probabilistic inference on the heritability coefficient. In particular we construct exact confidence intervals and demonstrate numerically our method's efficiency compared to that of existing methods.

AMS 2000 subject classifications: Primary 62F25; secondary 62J99.
Keywords and phrases: Differential equation, inferential model, plausibility, unbalanced design, variance components.

1. Introduction

Normal linear mixed effects models are useful in a variety of biological, physical, and social scientific applications with variability coming from multiple sources; see Khuri and Sahai (1985) and Searle, Casella and McCulloch (1992). In this paper, we focus on the case of two variance components, and the general model can be written as

$$Y = X\beta + Z\alpha + \varepsilon, \tag{1}$$

where Y is a n-vector of response variables, X and Z are design matrices for the fixed and random effects of dimension $n \times p$ and $n \times a$, respectively, β is a p-vector of unknown parameters, α is a normal random a-vector with mean 0 and covariance matrix $\sigma_\alpha^2 A$, and ε is a normal random n-vector with mean 0 and covariance matrix $\sigma_\varepsilon^2 I_n$. Here, $\sigma^2 = (\sigma_\alpha^2, \sigma_\varepsilon^2)$ is the pair of variance components. Assume A is known and X is of (full) rank $p < n$. The unknown parameters in this model are the p fixed-effect coefficients β and the two variance components σ^2, so the parameter space is $(p + 2)$-dimensional.

In biological applications, the quantities α and ε in (1) denote the genetic and environmental effects, respectively. Given that "a central question in biology is whether observed variation in a particular trait is due to environmental or biological factors" (Visscher, Hill and Wray, 2008), the heritability coefficient,

$\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$, which represents the proportion of phenotypic variance attributed to variation in genotypic values, is a fundamentally important quantity. Indeed, mixed-effect models and inference on the heritability coefficient has been applied recently in genome-wide association studies (Golan and Rosset, 2011; Yang *et al.*, 2010); see Section 5 for more on these applications.

Given the importance of the heritability coefficient, there are a number of methods available to construct confidence intervals for ρ or, equivalently, for the variance ratio $\psi = \sigma_\alpha^2/\sigma_\varepsilon^2$. When the design is balanced, Graybill (1976) and Searle, Casella and McCulloch (1992) give a confidence intervals for ψ and other quantities. When the design is possibly unbalanced, as we assume here, the problem is more challenging; in particular, exact ANOVA-based confidence intervals generally are not available. Wald (1940) gave intervals for ψ in the unbalanced case, and subsequent contributions include Harville and Fenech (1985), Fenech and Harville (1991), Lee and Seely (1996), and Burch and Iyer (1997). Bayesian (Gelman, 2006; Gelman *et al.*, 2004; Wolfinger and Kass, 2000) and fiducial methods (Cisewski and Hannig, 2012; E, Hannig and Iyer, 2008; Fisher, 1935) are also available.

The focus in this paper is exact prior-free probabilistic inference on the heritability coefficient based on the recently proposed inferential model (IM) framework. Martin and Liu (2013, 2015a,b) give a detailed account of the general framework, along with comparisons to other related approaches, including fiducial inference (Hannig, 2009, 2013), Dempster–Shafer inference (Dempster, 2008; Shafer, 1976), structural inference (Fraser, 1968), and confidence distributions (Xie and Singh, 2013). The IM approach is driven by the specification of an association between unobservable auxiliary variables, the observable data, and the unknown parameters. This association is followed up by using properly calibrated random sets to predict these auxiliary variables. Liu and Martin (2015) explain that it is this prediction of the auxiliary variables that distinguishes the IM approach. The IM output is a plausibility function that can be used for inference on the parameter of interest. In particular, the plausibility function yields confidence intervals with exact coverage. A brief overview of the general IM framework is given in Section 2.1. A key feature of the IM approach, and the motivation for this work, is that, besides being exact, its careful handling of uncertainty often leads to more efficient inference.

A key to the construction of an efficient IM is to reduce the dimension of the auxiliary variable as much as possible and, for the heritability coefficient, there are several dimension reduction steps. A first dimension reduction step eliminates the nuisance parameter β. These well-known calculations are reviewed in Section 2.2. The main technical contribution here is in Section 3.1, where we employ a differential equation-driven technique in the second dimension-reduction step, leading to a conditional IM for the heritability coefficient. In particular, this step allows us to reduce the dimension beyond that allowed by sufficiency. A predictive random set is introduced in Section 3.2, and in Section 3.3 we show that the corresponding plausibility function is properly calibrated and, therefore, the plausibility function-based confidence intervals are provably exact. Sections 4.2 and 4.3 illustrate the proposed method in simulated

and real data examples. The general message is that our proposed confidence intervals for the heritability coefficient are exact and tend to be more efficient compared to existing methods.

2. Preliminaries

2.1. Overview of inferential models

The goal of the IM framework is to get valid prior-free probabilistic inference. This is facilitated by first associating the observable data and unknown parameters to a set of unobservable auxiliary variables. For example, the marginal distribution of Y from (1) is

$$Y \sim \mathsf{N}_n(X\beta, \sigma_\varepsilon^2 I_n + \sigma_\alpha^2 Z A Z^\top), \tag{2}$$

which can be written in association form:

$$Y = X\beta + (\sigma_\varepsilon^2 I_n + \sigma_\alpha^2 Z A Z^\top)^{1/2} U, \quad U \sim \mathsf{N}_n(0, I_n). \tag{3}$$

That is, observable data Y and unknown parameters (β, σ^2) are associated with unobservable auxiliary variables U, in this case, a n-vector of independent standard normal random variables. Given this association, the basic IM approach is to introduce a predictive random set for U and combine with the observed value of Y to get a plausibility function. Roughly, this plausibility function assigns, to assertions about the parameter of interest, values in $[0,1]$ measuring the plausibility that the assertion is true. This plausibility function can be used to design exact frequentist tests or confidence regions.

Martin and Liu (2013) give a general description of this three-step IM construction and its properties; in Section 3.1 below we will flesh out these three steps, including choice of a predictive random set for the variance components problem at hand. The construction of exact plausibility intervals for the heritability coefficient is presented in Section 3.3, along with a theoretical justification of the claimed exactness.

Notice, in the association (3), that the auxiliary variable U is, in general, of higher dimension than that of the parameter (β, σ^2). There are two reasons for trying to reduce the auxiliary variable dimension. First, it is much easier to specify good predictive random sets when the auxiliary variable is of low dimension. Second, inference is more efficient when only a relatively small number of auxiliary variables require prediction. This dimension reduction is obtained via a conditioning operation, and Martin and Liu (2015a) give a general explanation of the gains as well as a novel technique for carrying out this reduction. We employ these techniques in Section 3.1 to get a dimension-reduced association for the heritability coefficient in the linear mixed model.

Another important challenge is when nuisance parameters are present. Our interest is in the heritability coefficient, a function of the full parameter (β, σ^2), so we need to marginalize over the nuisance parameters. The next section marginalizes over β using standard techniques; further marginalization will be carried out in Section 3.1.

2.2. Marginalizing out the fixed effect

Start with the linear mixed model (1). Following the setup in E, Hannig and Iyer (2008), let K be a $n \times (n-p)$ matrix such that $KK^\top = I_n - X(X^\top X)^{-1}X^\top$ and $K^\top K = I_{n-p}$. Next, let $B = (X^\top X)^{-1}X^\top$. Then $y \mapsto (K^\top y, By)$ is a one-to-one mapping. Moreover, the distribution of $K^\top Y$ depends on $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ only, and the distribution of BY depends on (β, σ^2), with β a location parameter. In particular, from (2), we get

$$K^\top Y \sim \mathsf{N}_{n-p}(0, \sigma_\varepsilon^2 I_{n-p} + \sigma_\alpha^2 G) \quad \text{and} \quad BY \sim \mathsf{N}_p(\beta, C_\sigma),$$

where $G = K^\top Z A Z^\top K$ and C_σ is a $p \times p$ covariance matrix of a known form that depends on $\sigma^2 = (\sigma_\alpha^2, \sigma_\varepsilon^2)$; its precise form is not important. From this point, the general theory in Martin and Liu (2015b) allows us to marginalize over β by simply deleting the BY component. Therefore, a marginal association for $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ is

$$K^\top Y = (\sigma_\varepsilon^2 I_{n-p} + \sigma_\alpha^2 G)^{1/2} U_2, \quad U_2 \sim \mathsf{N}_{n-p}(0, I_{n-p}).$$

This marginalization reduces the auxiliary variable dimension from n to $n-p$.

In the marginal association for $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ above, there are $n-p$ auxiliary variables but only two parameters. Classical results on sufficient statistics in mixed effects model that will facilitate further dimension reduction. For the matrix G defined above, let $\lambda_1 > \cdots > \lambda_L \geq 0$ denote the (distinct) ordered eigenvalues with multiplicities $r_1, \ldots, r_L$, respectively. Let $P = [P_1, \ldots, P_L]$ be a $(n-p) \times (n-p)$ orthogonal matrix such that $P^\top GP$ is diagonal with eigenvalues $\{\lambda_\ell : \ell = 1, \ldots, L\}$, in their multiplicities, on the diagonal. For P_ℓ, a $(n-p) \times r_\ell$ matrix, define

$$S_\ell = Y^\top K P_\ell P_\ell^\top K^\top Y, \quad \ell = 1, \ldots, L.$$

Olsen, Seely and Birkes (1976) showed that $(S_1, \ldots, S_L)$ is a minimal sufficient statistic for $(\sigma_\alpha^2, \sigma_\varepsilon^2)$. Moreover, the distribution of $(S_1, \ldots, S_L)$ is determined by

$$S_\ell = (\lambda_\ell \sigma_\alpha^2 + \sigma_\varepsilon^2) V_\ell, \quad V_\ell \sim \mathsf{ChiSq}(r_\ell), \quad \text{independent}, \quad \ell = 1, \ldots, L. \quad (4)$$

This reduces the auxiliary variable dimension from $n-p$ to L. We take (4) as our "baseline association."

Even in this reduced baseline association, there are L auxiliary variables but only two parameters, which means there is room for even further dimension reduction. The next section shows how to reduce to a scalar auxiliary variable when the parameter of interest is the scalar heritability coefficient.

3. Inferential model for heritability

3.1. Association

For the moment, it will be convenient to work with the variance ratio, $\psi = \sigma_\alpha^2/\sigma_\varepsilon^2$. Since $\psi = \rho/(1 - \rho)$ is a one-to-one function of ρ, the two parametriza-

tions are equivalent. Rewrite the baseline association (4) as

$$S_\ell = \sigma_\varepsilon^2(\lambda_\ell\psi + 1)V_\ell, \quad V_\ell \sim \mathsf{ChiSq}(r_\ell), \quad \text{independent}, \quad \ell = 1, \ldots, L. \quad (5)$$

If we make the following transformations,

$$X_\ell = (S_\ell/r_\ell)/(S_L/r_L), \quad \ell = 1, \ldots, L-1, \qquad X_L = S_L,$$
$$U_\ell = (V_\ell/r_\ell)/(V_L/r_L), \quad \ell = 1, \ldots, L-1, \qquad U_L = V_L.$$

then the association (5) becomes

$$X_\ell = \frac{\lambda_\ell\psi + 1}{\lambda_L\psi + 1}U_\ell, \quad \ell = 1, \ldots, L-1, \qquad X_L = \sigma_\varepsilon^2(\lambda_L\psi + 1)U_L.$$

Since for every (X, U, ψ), there exists a σ_ε^2 that solves the right-most equation, it follows from the general theory in Martin and Liu (2015b) that a marginal association for ψ is obtained by deleting the component above involving σ_ε^2. In particular, a marginal association for ψ is

$$X_\ell = \frac{\lambda_\ell\psi + 1}{\lambda_L\psi + 1}U_\ell, \quad \ell = 1, \ldots, L-1.$$

If we write

$$f_\ell(\rho) = \frac{1 + \rho(\lambda_\ell - 1)}{1 + \rho(\lambda_L - 1)}, \quad \ell = 1, \ldots, L-1,$$

then we get a marginal association for ρ of the form

$$X_\ell = f_\ell(\rho)U_\ell, \quad \ell = 1, \ldots, L-1. \quad (6)$$

Marginalization reduces the auxiliary variable dimension by 1. Further dimension reduction will be considered next. Note that the new auxiliary variable U is a multivariate F-distributed random vector (e.g., Amos and Bulgren, 1972).

Here we construct a local conditional IM for ρ as described in Martin and Liu (2015a). Select a fixed value ρ_0; more on this choice in Section 3.3. To reduce the dimension of the auxiliary variable U, in (6), from $L - 1$ to 1, we construct two pairs of functions—one pair, (T, H_0), on the sample space, and the other pair, (τ, η_0), on the auxiliary variable space. We insist that $x \mapsto (T(x), H_0(x))$ and $u \mapsto (\tau(u), \eta_0(u))$ are both one-to-one, and $H_0 = H_{\rho_0}$ and $\eta_0 = \eta_{\rho_0}$ are allowed to depend on the selected ρ_0. The goal is to find η_0 which is, in a certain sense, not sensitive to changes in the auxiliary variable.

Write the association (6) so that u is a function of x and ρ, i.e.,

$$u_\ell(x, \rho) = x_\ell/f_\ell(\rho), \quad \ell = 1, \ldots, L-1.$$

We want to choose the function η_0 such that the partial derivative of $\eta_0(u(x, \rho))$ with respect to ρ vanishes at $\rho = \rho_0$. By the chain rule, we have

$$\frac{\partial\eta_0(u(x, \rho))}{\partial\rho} = \frac{\partial\eta_0(u)}{\partial u}\bigg|_{u=u(x,\rho)}\frac{\partial u(x, \rho)}{\partial\rho},$$

so our goal is to find η_0 to solve the following partial differential equation:

$$\frac{\partial \eta_0(u)}{\partial u}\bigg|_{u=u(x,\rho)} \frac{\partial u(x,\rho)}{\partial \rho} = 0, \quad \text{at } \rho = \rho_0;$$

here $\partial u/\partial \rho$ is a $(L-1) \times 1$ vector and $\partial \eta_0/\partial u$ is a $(L-2) \times (L-1)$ matrix of rank $L-2$. Toward solving this partial differential equation, first we get that the partial derivative of $u_\ell(x,\rho)$ with respect to ρ satisfies

$$\frac{\partial u_\ell(x,\rho)}{\partial \rho} = -\frac{f_\ell'(\rho)}{f_\ell(\rho)^2} x_\ell = -g(\rho)\, u_\ell(x,\rho),$$

where $g(\rho) = \{\frac{\partial}{\partial \rho} \log f_1(\rho), \ldots, \frac{\partial}{\partial \rho} \log f_{L-1}(\rho)\}^\top$. This simplifies the relevant partial differential equation to the following:

$$\frac{\partial \eta_0(u)}{\partial u}\bigg|_{u=u(x,\rho)} \text{diag}\{u(x,\rho)\}\, g(\rho) = 0, \quad \text{at } \rho = \rho_0, \tag{7}$$

where $\text{diag}(a)$ is a diagonal matrix constructed from a vector a. The method of characteristics (e.g., Polyanin, Zaitsev and Moussiaux, 2002) for solving partial differential equations identifies a logarithmic function of the form

$$\eta_0(u)^\top = (\log u_1, \ldots, \log u_{L-1}) M_0^\top, \tag{8}$$

where $M_0 = M_{\rho_0}$ is a $(L-2) \times (L-1)$ matrix with rows orthogonal to $g(\rho)$ at $\rho = \rho_0$. For example, since the matrix that projects to the orthogonal complement of the column space of $g(\rho_0)$ has rank $L-2$, we can take M_0 to be a matrix whose $L-2$ rows form a basis for that space. For M_0 defined in this way, it is easy to check that η_0 in (8) is indeed a solution to the partial differential equation (7).

We have completed specification of the η_0 function; it remains to specify τ and (T, H_0). The easiest to specify next is $H_0(x)$, the value of $\eta(u(x,\rho_0))$, as a function of x:

$$H_0(x)^\top = \left(\log \frac{x_1}{f_1(\rho_0)}, \ldots, \log \frac{x_{L-1}}{f_{L-1}(\rho_0)}\right) M_0^\top.$$

As we describe below, the goal is to condition on the observed value of $H_0(X)$.

Next, we define T and τ to supplement H_0 and η_0, respectively. In particular, take a $(L-1)$-vector $w(\rho)$ which is not orthogonal to $g(\rho)$ at $\rho = \rho_0$. It is easy to check that the entries in $g(\rho)$ are strictly positive for all ρ. Therefore, we can take $w(\rho)$ independent of ρ; for example, $w(\rho) \equiv 1_{L-1}$ is not orthogonal to $g(\rho)$. Now set

$$\begin{pmatrix} \tau(u) \\ \eta_0(u) \end{pmatrix} = \begin{pmatrix} 1_{L-1}^\top \\ M_0 \end{pmatrix} \begin{pmatrix} \log u_1 \\ \vdots \\ \log u_{L-1} \end{pmatrix}. \tag{9}$$

This is a log-linear transformation, and the linear part is non-singular, so this is a one-to-one mapping. Finally, we take T as

$$T(x) = \sum_{\ell=1}^{L-1} \log x_\ell.$$

Since $(T(x), H_0(x))$ is log-linear, just like $(\tau(u), \eta_0(u))$, it is also one-to-one.

We can now write the conditional association for ρ. For the given ρ_0, the mapping $x \mapsto (T(x), H_0(x))$ describes a split of our previous association (6) into two pieces:

$$\sum_{\ell=1}^{L-1} \log X_\ell = \sum_{\ell=1}^{L-1} \log f_\ell(\rho) + \sum_{\ell=1}^{L-1} \log U_\ell \quad \text{and} \quad H_0(X) = \eta_0(U).$$

The first piece carries direct information about ρ. The second piece plays a conditioning role, correcting for the fact that some information was lost in reducing the $(L-1)$-dimensional X to a one-dimensional $T(X)$. To complete the specification of the conditional association, write $\varphi(\rho) = \sum_{\ell=1}^{L-1} \log f_\ell(\rho)$ and $V = \sum_{\ell=1}^{L-1} \log U_\ell$. Then we have

$$T(X) = \varphi(\rho) + V, \quad V \sim \mathsf{P}_{V|h_0, \rho_0}, \tag{10}$$

where $\mathsf{P}_{V|h_0, \rho_0}$ is the conditional distribution of $\tau(U)$, given that $\eta_0(U)$ equals to the observed value h_0 of $H_0(X)$. To summarize, (10) completes the association step that describes the connection between observable data, unknown parameter of interest, and unobservable auxiliary variables. Of particular interest is that this association involves only a one-dimensional auxiliary variable compared to the association (5) obtained from the minimal sufficient statistics that involves an L-dimensional auxiliary variable. This dimension reduction will come in handy for the choice of predictive random set in the following section. The price we paid for this dimension reduction was the choice of a particular localization point ρ_0. In Section 3.3 we employ a trick to side-step this issue when the goal is, as in this paper, to construct confidence intervals.

3.2. Predictive random sets

Having reduced the auxiliary variable to a scalar in (10), the choice of an efficient predictive random set is now relatively simple. Though there is an available theory of optimal predictive random sets (Martin and Liu, 2013), here we opt for simplicity; in particular, we propose a default predictive random set that is theoretically sound and computational and intuitively simple.

Consider the following version of what Martin and Liu (2013) call the "default" predictive random set:

$$\mathcal{S} = \{v : |v - \mu_0| \leq |V - \mu_0|\}, \quad V \sim \mathsf{P}_{V|h_0, \rho_0}. \tag{11}$$

This $\mathcal{S}$, with distribution $\mathsf{P}_{\mathcal{S}|h_0, \rho_0}$, is a random interval, centered at the mean μ_0 of the conditional distribution $\mathsf{P}_{V|h_0, \rho_0}$. One key feature of this predictive random set is that it is nested, i.e., for any two distinct realizations of $\mathcal{S}$, one is a subset of the other. The second key feature is a calibration of the predictive random set distribution with the underlying distribution of the auxiliary variable. Following Martin (2014), define the contour function

$$\gamma_{\mathcal{S}}(v) = \mathsf{P}_{\mathcal{S}|h_0, \rho_0}(\mathcal{S} \ni v),$$

which represents the probability that the predictive random set contains the value v of the auxiliary variable. We shall require that

$$\mathsf{P}_{V|h_0,\rho_0}\{\gamma_{\mathcal{S}}(V) \geq 1 - \alpha\} \leq \alpha, \quad \forall\, \alpha \in (0,1). \tag{12}$$

For the default predictive random set in (11), it is easy to check that

$$\gamma_{\mathcal{S}}(v) = \mathsf{P}_{V|h_0,\rho_0}\{|V - \mu_0| \geq |v - \mu_0|\} = 1 - F_{h_0,\rho_0}(|v - \mu_0|),$$

where F_{h_0,ρ_0} is the distribution function of $|V - \mu_0|$ for $V \sim \mathsf{P}_{V|h_0,\rho_0}$. From the construction above, it is clear that it is a continuous distribution. Then, $|V - \mu_0|$ is a continuous random variable, so $\gamma_{\mathcal{S}}(V)$ is uniformly distributed on $(0,1)$. Therefore, (12) holds for the default predictive random set $\mathcal{S}$. Results on optimal predictive random sets are available (Martin and Liu, 2013), but here, again, our focus is on simplicity. See Section 5.

3.3. Plausibility intervals

Here we combine the association in Section 3.1 with the predictive random set described above to produce a plausibility function for inference about ρ. In general, a plausibility function is a data-dependent mapping that assigns, to each assertion about the parameter, a value in $[0,1]$, with the interpretation that small values suggest the assertion is not likely to be true, given the data; see Martin and Liu (2013). For simplicity, we focus only on the collection of singleton assertions, i.e., $\{\rho = r\}$ for $r \in [0,1]$. These are also the relevant assertions for constructing interval estimates based on the IM output.

Let $X = x$ be the observations in (6). The association step in Section 3.1 yields a data-dependent collection of sets indexed by the auxiliary variable. In particular, write $R_x(v) = \{\rho : T(x) = \varphi(\rho) + v\}$, a set-valued function of v. These sets are combined with the predictive random set in Section 3.2 to get an enlarged x-dependent random set:

$$R_x(\mathcal{S}) = \bigcup_{v \in \mathcal{S}} R_x(v). \tag{13}$$

Now, for a given assertion $\{\rho = r\}$, we compute the plausibility function,

$$\mathsf{pl}_{x|h_0,\rho_0}(r) = \mathsf{P}_{\mathcal{S}|h_0,\rho_0}\{R_x(\mathcal{S}) \ni r\},$$

the probability that the random set $R_x(\mathcal{S})$ contains the asserted value r of ρ. A simple calculation shows that, in this case with singleton assertions, we have

$$\mathsf{pl}_{x|h_0,\rho_0}(r) = \gamma_{\mathcal{S}}\big(T(x) - \varphi(r)\big) = 1 - F_{h_0,\rho_0}(|T(x) - \varphi(r) - \mu_0|),$$

where F_{h_0,ρ_0} is defined in Section 3.2. The above display shows that the plausibility function can be expressed directly in terms of the distribution of the predictive random set, without needing to go through the construction of $R_x(\mathcal{S})$ as in (13).

We pause here to answer a question that was left open from Section 3.1, namely, how to choose the localization point ρ_0. Following Martin and Liu (2015a), we propose here to choose ρ_0 to match the value of ρ specified by the singleton assertion. That is, we propose to let the localization point depend on the assertion. All the elements in the plausibility function above with a 0 subscript, denoting dependence on ρ_0, are changed in an obvious way to get a new plausibility function

$$\mathsf{pl}_{x|h_\rho,\rho}(\rho) = 1 - F_{h_\rho,\rho}(|T(x) - \varphi(\rho) - \mu_\rho|). \tag{14}$$

We treat this as a function of ρ to be used for inference. In particular, we can construct a $100(1-\alpha)\%$ plausibility interval for ρ as follows:

$$\Pi_\alpha(x) = \{\rho : \mathsf{pl}_{x|h_\rho,\rho}(\rho) > \alpha\}. \tag{15}$$

The plausibility function, and the corresponding plausibility region, are easy to compute, as we describe in Section 4.1. Moreover, the calibration (12) of the predictive random set leads to exact plausibility function-based confidence intervals, as we now show.

We need some notation for the sampling distribution of X, given all the relevant parameters. Recall that the distribution of X actually depends on $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ or, equivalently, $(\rho, \sigma_\varepsilon^2)$. The error variance σ_ε^2 is a nuisance parameter, but σ_ε^2 still appears in the sampling model for X. We write this sampling distribution as $\mathsf{P}_{X|\rho,\sigma_\varepsilon^2}$.

Theorem 1. *Take the association* (10) *and the default predictive random set* $\mathcal{S}$ *in* (11). *Then for any* ρ, *any value* h_ρ *of* H_ρ, *and any* σ_ε^2, *the plausibility function satisfies*

$$\mathsf{P}_{X|\rho,\sigma_\varepsilon^2}\{\mathsf{pl}_{X|h_\rho,\rho}(\rho) \leq \alpha \mid H_\rho(X) = h_\rho\} = \alpha, \quad \forall\, \alpha \in (0,1). \tag{16}$$

Proof. For given $(\sigma_\varepsilon^2, \rho)$, if h_ρ is the value of $H_\rho(X)$, then it follows from the conditional distribution construction that the plausibility function in (14), as a function of $T(X)$ with $X \sim \mathsf{P}_{X|\rho,\sigma_\varepsilon^2}$, is $\mathsf{Unif}(0,1)$. Then the equality in (16) follows immediately. □

Averaging the left-hand side of (16) over h_ρ, with respect to the distribution of $H_\rho(X)$, and using iterated expectation gives the following unconditional version of Theorem 1.

Corollary 1. *Under the conditions of Theorem 1, for any* $(\rho, \sigma_\varepsilon^2)$,

$$\mathsf{P}_{X|\rho,\sigma_\varepsilon^2}\{\mathsf{pl}_{X|H_\rho(X),\rho}(\rho) \leq \alpha\} = \alpha, \quad \forall\, \alpha \in (0,1).$$

Since we have proper calibration of the plausibility function, both conditionally and unconditionally, coverage probability results for the plausibility interval (15) are also available. This justifies our choice to call $\Pi_\alpha(x)$ a $100(1-\alpha)\%$ plausibility interval, i.e., the frequentist coverage probability of Π_α is exactly $1 - \alpha$.

Corollary 2. *The coverage probability of* $\Pi_\alpha(X)$ *in* (15) *is exactly* $1 - \alpha$.

4. Numerical results

4.1. Implementation

Evaluation of the plausibility function in (14) requires the distribution function $F_{h_\rho,\rho}$ of $|V - \mu_\rho|$ corresponding to the conditional distribution $\mathsf{P}_{V|h_\rho,\rho}$ of $V = \tau(U)$, given $\eta_\rho(U) = H_\rho(X)$. This conditional distribution is not of a convenient form, so numerical methods are needed. For ρ fixed, since the transformation (9) from U to $(\tau(U), \eta_\rho(U))$ is of a log-linear form, and the density function of U can be written in closed-form, we can evaluate the joint density for $(\tau(U), \eta_\rho(U))$ and, hence, the conditional density of $V = \tau(U)$. Numerical integration is used to evaluate the normalizing constant, the mean μ_ρ, and the distribution function $F_{h_\rho,\rho}$. R code is available at www.math.uic.edu/~rgmartin.

4.2. Simulation results

In this section, we consider a standard one-way random effects model, i.e.,

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n_i,$$

where $\alpha_1, \ldots, \alpha_a$ are independent with common distribution $\mathsf{N}(0, \sigma_\alpha^2)$, and the ε_{ij}s are independent with common distribution $\mathsf{N}(0, \sigma_\varepsilon^2)$; the α_is and ε_{ij}s are also mutually independent. Our goal is to compare the proposed IM-based plausibility intervals for ρ with the confidence intervals based on several competing methods. Of course, the properties of the various intervals depend on the design, in this case, the within-group sample sizes $n_1, \ldots, n_a$, and the values of $(\sigma_\alpha^2, \sigma_\varepsilon^2)$. Our focus here is on cases with small sample sizes, namely, where the total sample size $n = n_1 + \cdots + n_a$ is fixed at 15. The three design patterns $(n_1, \ldots, n_a)$ considered are: $(1, 1, 1, 1, 1, 10)$, $(2, 4, 4, 5)$, and $(2, 3, 10)$. The nine $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ pairs considered are: $(0.05, 10)$, $(0.1, 10)$, $(0.5, 10)$, $(1, 10)$, $(0.5, 2)$, $(1, 1)$, $(2, 0.5)$, $(5, 0.2)$, and $(10, 0.1)$. Without loss of generality, we set $\mu = 0$.

For each design pattern and pair of $(\sigma_\alpha^2, \sigma_\varepsilon^2)$, 1000 independent data sets were generated and 95% two-sided interval estimates for ρ were computed based on the exact method of Burch and Iyer (1997), the fiducial method of E, Hannig and Iyer (2008), and the proposed IM method. Empirical coverage probabilities and average length of the confidence interval under each setting were compared to investigate the performance of each method. Besides these three methods, we also implemented Bayesian and profile likelihood approaches. The three aforementioned methods all gave better intervals than the Bayesian method, and the profile likelihood method was very unstable with small sample sizes, often having very high coverage with very wide intervals or very low coverage with very narrow intervals. So, these results are not reported.

A summary of the simulation results is displayed in Figure 1. Panel (a) displays the coverage probabilities, and Panel (b) displays the relative length difference, which is defined as the length of the particular interval minus the length of the IM interval, scaled by the length of the IM interval. As we expect from Corollary 2, the IM plausibility intervals have coverage at the nominal 95% level. We also see that the IM intervals tend to be shorter than the fiducial and
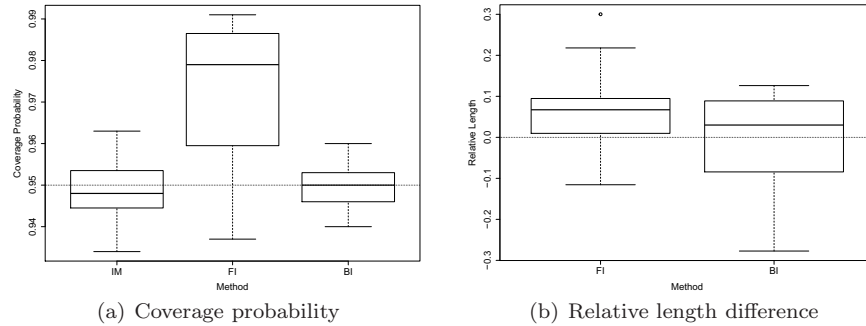
(a) Coverage probability   (b) Relative length difference

FIG 1. *Simulation results from Section 4.2. BI corresponds to the exact method of Burch and Iyer (1997), and FI corresponds to the fiducial method of E, Hannig and Iyer (2008).*

Burch–Iyer confidence intervals. The fiducial intervals have coverage probability exceeding the nominal 95% level, but this comes at the expense of longer intervals on average. Overall, the proposed IM-based method performs quite well compared to these existing methods. We also replicated the simulation study in E, Hannig and Iyer (2008), which involves larger sample sizes and a broader range of imbalance, and the relative comparisons between these three methods are the same as here.

4.3. Real-data analysis

Example 1. Equal numbers of subjects are tested under each standard and test preparations and a blank dose under a $(2K + 1)$-point symmetrical slope-ratio assay. The response, on logarithmic scale, is assumed to depend linearly on the dose level. A modified balanced incomplete block design with $2K' + 1$ ($K' < K$) block size is introduced by Das and Kulkarni (1966). The ith dose levels for standard and test preparations are represented by s_i and t_i, $i = 1, \ldots, K$. Under this design, the dose will be equally spaced and listed in ascending order. A balanced incomplete block design with K doses of the standard preparation inside K' blocks is constructed and used as the basic design. Then a modified design is constructed by adding a blank dose and K' doses of the test preparation into every block, under the rule that dose t_i should accompany s_i in every blocks. The model developed by Das and Kulkarni can be written as

$$y_{ijm} = \mu + \beta_j x_{ij} + \alpha_m + \varepsilon_{ijm}, \quad i \in \{s, t, c\}, \quad j = 1, \ldots, k, \quad m = 1, \ldots, b$$

where y_{sjm}, y_{tjm}, y_{cjm} represent observation response in mth block for jth dose of standard preparation, test preparation and blank dose; x_{sj} and x_{tj} represent jth dose level for standard and test preparation; x_{cj} is zero by default, α_m denotes mth block effect; and ε_{ijm} denotes independent random errors with common distribution $\mathsf{N}(0, \sigma_\varepsilon^2)$. We consider random block effects and assume that α_m are independent with common distribution $\mathsf{N}(0, \sigma_\alpha^2)$. Independence of α_m and ε_{ijm} is also assumed.

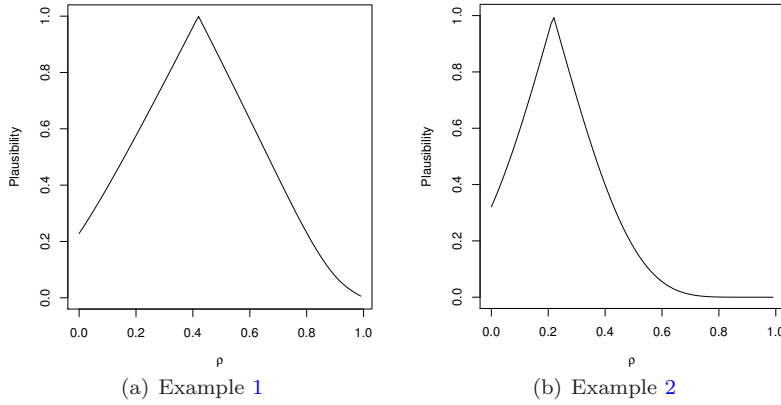(a) Example 1   (b) Example 2

Fɪɢ 2. *Plausibility functions for ρ in the two examples.*

Tᴀʙʟᴇ 1
Upper bounds on the interval estimates for ρ (lower bounds are all zero) based on the three methods in the two real-data examples

	Example 1		Example 2	
Method	90%	95%	90%	95%
Burch–Iyer	0.913	0.956	0.567	0.615
Fiducial	0.916	0.957	0.466	0.530
IM	0.881	0.924	0.554	0.597

We analyze data coming from a nine-point slope-ratio assay on riboflavin content of yeast, with two replications in each dose; see Table 2 in E, Hannig and Iyer (2008) for the design and data. For this design, we have $L = 3$ distinct eigenvalues, namely, 4.55, 1, and 0, with multiplicities 1, 1, and 10, respectively. A plot of the plausibility function for ρ is shown in Figure 2(a). The function exceeds 0.2 at $\rho = 0$, which implies that 90% and 95% plausibility intervals include zero. The left panel of Table 1 shows the 90% and 95% interval estimates for ρ based on the Burch–Iyer, fiducial, and IM methods. In this case, the IM intervals are provably exact and shorter.

Example 2. Harville and Fenech (1985) analyzed data on birth weights of lambs. These data consist of the weights information at the birth of 62 single-birth male lambs, and were collected from three selection lines and two control lines. Each lamb was the offspring of one of the 23 rams and each lamb had a distinct dam. Age of the dam was also recorded and separated into three categories, numbered 1 (1–2 years), 2 (2–3 years), and 3 (over 3 years). A linear mixed model for these data is

$$y_{ijkl} = \mu + \beta_i + \pi_j + \alpha_{jk} + \varepsilon_{ijkl},$$

where y_{ijkl} represents the weight of the lth offspring of the kth sire in the jth population lines and of a dam in the ith age category; β_i represents ith level

age effect; π_j represents the jth line effects; α_{jk} denotes random sire effects and are assumed to be independently distributed as $\mathsf{N}(0, \sigma_\alpha^2)$; and random errors denoted by ε_{ijkl} is supposed to be independently distributed as $\mathsf{N}(0, \sigma_\varepsilon^2)$. Furthermore, the α_{jk}s and ε_{ijkl}s are assumed to be independent. In this case, $L = 18$, $\lambda_1 = 5.09$, $\lambda_L = 0$ and $r_L = 37$; all non-zero eigenvalues have multiplicity 1 except $\lambda_8 = 2$ with multiplicity 2. A plot of the plausibility function for ρ is shown in Figure 2(b). As in the previous example, the plausibility function is positive at $\rho = 0$, which means that plausibility intervals with any reasonable level will contain $\rho = 0$. We also used each of the three methods considered above to compute 90% and 95% interval estimates for ρ. The results are shown in the right panel of Table 1. In this case, IM gives a shorter interval compared to Burch–Iyer. The fiducial interval, however, is shorter than both exact intervals. We expect the IM interval to be most efficient, so we explore the relative performance a bit further by simulating 1000 independent data sets from the fitted model in this case, i.e., with $\hat{\sigma}_\alpha^2 = 0.767$ and $\hat{\sigma}_\varepsilon^2 = 2.763$ as the true values. In these simulations, the fiducial and IM coverage probabilities were 0.944 and 0.954, respectively, both within an acceptable range of the nominal level, but the average lengths of the intervals are 0.488 and 0.456. That is, the IM intervals tend to be shorter than the fiducial intervals in problems similar to this example, as we would expect.

5. Concluding remarks

The IM method proposed here gives exact confidence intervals for the heritability coefficient ρ, as well as the variance ratio ψ, and numerical results suggest increased efficiency compared to existing methods. In fact, we found here that our IM approach was more efficient than fiducial for inference on ρ, and a key reason for this is that the IM can be carefully tailored for efficient marginal inference on ρ, while the fiducial approach applies to the full parameter $(\sigma_\alpha^2, \sigma_\varepsilon^2)$ and the marginalization to ρ via integration is out of the statistician's hands. A question is if these same techniques can be employed for exact prior-free probabilistic inference on other quantities related to the variance components $(\sigma_\alpha^2, \sigma_\varepsilon^2)$. It is well-known that, for the unbalanced design case, exact marginalization is challenging. In the IM context, this means that the association is not "regular" in the sense of Martin and Liu (2015b). Therefore, some more sophisticated tools are needed for exact inference on, say, the individual variance components σ_α^2 and σ_ε^2. This application provides clear motivation for our ongoing investigations into more general IM-based marginalization strategies.

Another important question is if the techniques presented herein can be applied in more complex and high-dimensional mixed-effect models. In genome-wide association studies, for example, the dimensions of the problem are extremely large. We expect that, conceptually, the techniques described here will carry over to the more complex scenario. However, there will be computational challenges to overcome, as with all approaches (Kang *et al.*, 2010; Zhou and Stephens, 2014). This, along with the incorporation of optimal predictive random sets (e.g. Martin and Liu, 2013) is a focus of ongoing research.

Acknowledgements

References

AMOS, D. E. and BULGREN, W. G. (1972). Computation of a multivariate F distribution. *Math. Comp.* **26** 255–264. MR0298881

BURCH, B. D. and IYER, H. K. (1997). Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model. *Biometrics* **53** 1318–1333. MR1614382

CISEWSKI, J. and HANNIG, J. (2012). Generalized fiducial inference for normal linear mixed models. *Ann. Statist.* **40** 2102–2127. MR3059078

DAS, M. N. and KULKARNI, G. A. (1966). Incomplete block designs for bioassays. *Biometrics* **22** 706–729. MR0212959

DEMPSTER, A. P. (2008). The Dempster–Shafer calculus for statisticians. *Internat. J. Approx. Reason.* **48** 365–377. MR2419025

E, L., HANNIG, J. and IYER, H. (2008). Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *J. Amer. Statist. Assoc.* **103** 854–865. MR2524335

FENECH, A. P. and HARVILLE, D. A. (1991). Exact confidence sets for variance components in unbalanced mixed linear models. *Ann. Statist.* **19** 1771–1785. MR1135148

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics* **6** 391–398.

FRASER, D. A. S. (1968). *The Structure of Inference.* John Wiley & Sons Inc., New York. MR0235643

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533. MR2221284

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, Second ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

GOLAN, D. and ROSSET, S. (2011). Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics* **27** 317–323.

GRAYBILL, F. A. (1976). *Theory and Application of the Linear Model.* Duxbury Press, North Scituate, Mass. MR0455218

HANNIG, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19** 491–544. MR2514173

HANNIG, J. (2013). Generalized fiducial inference via discretization. *Statist. Sinica* **23** 489–514. MR3086644

HARVILLE, D. A. and FENECH, A. P. (1985). Confidence intervals for variance ratio, or for heritability, in an unbalanced mixed linear model. *Biometrics* **41** 137–152. MR793438

3076 Q. Cheng et al.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-y., Freimer, N. B., Sabatti, C. and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42** 348–354.

Khuri, A. I. and Sahai, H. (1985). Variance components analysis: A selective literature survey. *Internat. Statist. Rev.* **53** 279–300. MR967214

Lee, Y. and Seely, J. (1996). Computing the Wald interval for a variance ratio. *Biometrics* **52** 1486–1491.

Liu, C. and Martin, R. (2015). Frameworks for prior-free posterior probabilistic inference. *WIREs Comput. Stat.* **7** 77–85.

Martin, R. (2014). Random sets and exact confidence regions. *Sankhyā A* **76** 288–304.

Martin, R. and Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.* **108** 301–313.

Martin, R. and Liu, C. (2015a). Conditional inferential models: combining information for prior-free probabilistic inference. *J. R. Stat. Soc. Ser. B* **77** 195–217.

Martin, R. and Liu, C. (2015b). Marginal inferential models: Prior-free probabilistic inference on interest parameters. *J. Amer. Statist. Assoc.*, to appear; arXiv:1306.3092.

Olsen, A., Seely, J. and Birkes, D. (1976). Invariant quadratic unbiased estimation for two variance components. *Ann. Statist.* **4** 878–890. MR0418345

Polyanin, A. D., Zaitsev, V. F. and Moussiaux, A. (2002). *Handbook of first order partial differential equations. Differential and Integral Equations and Their Applications* **1**. Taylor & Francis Ltd., London. MR1920177

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance components. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.* John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. MR1190470

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, N.J. MR0464340

Visscher, P. M., Hill, W. G. and Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **9** 255–266.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. *Ann. Math. Statistics* **11** 96–100. MR0001502

Wolfinger, R. D. and Kass, R. E. (2000). Nonconjugate Bayesian analysis of variance component models. *Biometrics* **56** 768–774.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution of a parameter—a review. *Int. Statist. Rev.* **81** 3–39.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42** 565–569.

Zhou, X. and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11** 407–409.