On the role of parametrisation in models with a misspecified nuisance component

Heather Battey
Department of Mathematics, Imperial College London

ISNPS, June 26, 2024

Statistical models

Provisional base for scientific understanding.

Aspects of core scientific interest are often relatively securely specified in terms of a small number of interpretable parameters.

Other aspects may be chosen somewhat arbitrarily, typically on the basis of mathematical convenience.

Prospect of misspecification in the nuisance component is high.

A motivating example

Treatment effects and block effects

Matched pairs. One individual from each pair chosen at random to receive a treatment, the other is the control.

Outcomes (Y_{i1}, Y_{i0}) exponentially distributed of rates $\gamma_i\psi$ and γ_i/ψ.

Treatment param. ψ; *n* nuisance parameters $\gamma_1, \ldots, \gamma_n$ encapsulate arbitrary dependence on (perhaps unmeasured) covariates. Semipara.

Treating pair effects as fixed

If $\gamma_1, \ldots, \gamma_n$ are treated as fixed arbitrary constants they can be eliminated from the analysis by taking ratios $Z_i = Y_{i1}/Y_{i0}$.

Density function: $f_Z(z;\psi) = \frac{\psi^2}{(1+\psi^2 z)^2}$ no dependence on γ_i.

This is what I would do, but . . .

Treating pair effects as random

An alternative approach treats pair-specific parameters as random.

Convenient choice of distribution in previous example, e.g. gamma.

A lengthy calculation (Battey and Cox, 2020) shows that $\hat{\psi}$ is consistent in spite of arbitrary misspecification of the distribution for $(\gamma_i)_{i=1}^n$.

Questions

How sensitive is this conclusion to the formulation of the model?

- Can the response distribution be changed?
- Can the parametrisation be changed?
- Can the assumed distribution on $\gamma_1, \ldots, \gamma_n$ be changed?

Is there identifiable structure in the model that led to consistency of the MLE for the interest parameter?

Results for general models specialised to particular cases.

Motivation and scope

Provide insight into the structure of inference at the population level under misspecification.

Reference to the unsolved nature of the question being addressed, e.g.:

> *[...] if the model is misspecified there is no guarantee that the estimator will be consistent or even close to the true value.*

Evans and Didelez (2024+). Parametrising and simulating from causal models (with discussion). *J. R. Statist. Soc. B*, to appear.

Misspecification: early history

Cox (1961, 1962)

Parametric model to be fitted: $m'(y;\theta)$, $y = (y_1, \ldots, y_n)$

True model $m(y)$.

MLE $\hat{\theta} \to_p \theta_m^0$, where θ_m^0 solves

$$\mathbb{E}_m[\nabla_\theta \log m'(Y;\theta)]_{\theta=\theta_m^0} = 0.$$

Equivalently θ_m^0 minimizes with respect to θ the Kullback-Leibler divergence

$$\int m(y) \log\left\{ \frac{m(y)}{m'(y;\theta)} \right\} dy.$$

Asymptotic variance given by the "sandwich formula".

Tests of separate families (assumes one is correct).

Elaboration and development

Huber (1967): regularity conditions etc.

White (1982): identifiability of θ_m^0; tests of misspec.

The inferential target

This talk: the inferential target has a stable interpretation in the true model and the fitted model. Other aspects are needed to complete the specification: these might be misspecified.

Formalisation

Battey, H. S. and Reid, N. (2024). On the role of parametrisation in models with a misspecified nuisance component. *arXiv:2402.05708*

Misspecified nuisance component

True density function m for outcomes parametrised in terms of an interest parameter β with true value β^*.

Assumed model m': same interpretable interest parameter β; misspecified in other ways; notional nuisance parameter α.

Log-lik for fitted model: $\ell(\beta, \alpha) = \log m'(y; \beta, \alpha)$; $(\beta, \alpha) \in \mathcal{B} \times \mathcal{A}$. Maximisation: $(\hat{\beta}, \hat{\alpha}) \to_p (\beta_m^0, \alpha_m^0)$ solves

$$\mathbb{E}_m[\nabla_{(\beta,\alpha)} \ell(\beta_m^0, \alpha_m^0)] = 0.$$

Misspecification: no value of $\alpha \in \mathcal{A}$ gives back m.

The inferential target

Our inferential target is β^*, not (β_m^0, α_m^0) to which the classical literature applies.

What structure implies that $\beta^* = \beta_m^0$ for any m?

Ideally would like to be able to check this without knowledge of m or α_m^0.

Parameter *m*-orthogonality

Definition. *Let* $\nabla^2_{\beta\alpha}\ell(\beta,\alpha)$ *denote the cross-partial derivative of the log-likelihood function. The parameter* β *is said to be m-orthogonal to the notional parameter* α *if* $\mathbb{E}_m[\nabla^2_{\beta\alpha}\ell(\beta,\alpha)] = 0$.

Notation. $\mathcal{B} \perp_m \mathcal{A}$: global *m*-orthogonality;
$\mathcal{B} \perp_m \alpha$ local *m*-orthogonality at α for any β;
$\beta \perp_m \mathcal{A}$ and local *m*-orthogonality at β for any α.

Geometrically. The stronger property $\nabla^2_{\beta\alpha}\ell(\beta,\alpha) = 0$ is an absence of torsion; the true model *m* trivially plays no role.
$\mathbb{E}_m[\nabla^2_{\beta\alpha}\ell(\beta,\alpha)] = 0 \Rightarrow$ any torsion is not systematic when data are generated from *m*.

Relevance of parameter *m*-orthogonality

Immediate: $\beta^* = \beta_m^0 \iff \underbrace{\mathbb{E}_m[\nabla_\beta \ell(\beta^*, \alpha_m^0)] = 0}_{C0}$.

But we don't know α_m^0 or m.

A first general result:
$\mathbb{E}_m[\nabla_\beta \ell(\beta^*, \alpha_m^0)] = 0$ is equivalent to
$\underbrace{(\mathbb{E}_m[\nabla_\beta \ell(\beta^*, \alpha)] = 0 \ \forall \alpha \in \mathcal{A})}_{C1.1}$ if and only if $\underbrace{\beta^* \perp_m \mathcal{A}}_{C1.2}$.

$C1.1$ and $C1.2$ imply $C0$ and therefore consistency of $\hat{\beta}$.
In some classes of models $C1.1$ & $C1.2$ can be guaranteed for any m through suitable parametrisation.

A weaker requirement

Suppose $C1.2$ fails, i.e. $\beta^* \not\perp_m \mathcal{A}$.

Second result: $(i^{\beta\beta} g_\beta + i^{\beta\alpha} g_\alpha = 0 \ \forall \alpha \in \mathcal{A}) \implies \beta^* = \beta^0_m$.
Scalar case: $(i_{\alpha\alpha} g_\beta + i_{\beta\alpha} g_\alpha = 0 \ \forall \alpha \in \mathcal{A}) \implies \beta^* = \beta^0_m$

where $i = i(\beta^*, \alpha) = \mathbb{E}_m[-\nabla^2_{(\beta,\alpha)} \ell(\beta^*, \alpha)]$

$$\begin{pmatrix} i^{\beta\beta} & i^{\beta\alpha} \\ i^{\alpha\beta} & i^{\alpha\alpha} \end{pmatrix} = \begin{pmatrix} i_{\beta\beta} & i_{\beta\alpha} \\ i_{\alpha\beta} & i_{\alpha\alpha} \end{pmatrix}^{-1}.$$

$$\begin{aligned} g_\beta := g_\beta(\beta^*, \alpha) &= \mathbb{E}_m[\nabla_\beta \ell(\beta^*, \alpha)], \\ g_\alpha := g_\alpha(\beta^*, \alpha) &= \mathbb{E}_m[\nabla_\alpha \ell(\beta^*, \alpha)]. \end{aligned}$$

Examples

A class of natural examples involve misspecified random effects distributions.

There are also many examples where parameter *m*-orthogonality arises due to a parameter cut: these are relatively easy cases, although the conclusion may not have been obvious without having the structure made clear.

Examples of neither form. . .

Matched comparison and balanced two-group problems

Outcomes $(Y_{i1}, Y_{i0})_{i=1}^{n}$ on treated and untreated individuals. Distributions parametrised by treatment effect ψ and pair-effect γ_i (drop subscript) belong to a transformation model under the action of $g = g_\psi \in G$.

ψ-**symmetric parametrization**: density functions f_1, f_0 related by

$$f_U(u;\gamma)du = f_1(g\,u; g\gamma)d(g\,u) = f_0(g^{-1}u; g^{-1}\gamma)d(g^{-1}u)$$

for some $g \in G$. In other words, $g^{-1}Y_1 \overset{d}{=} U$ and $g\,Y_0 \overset{d}{=} U$, where the distribution of U is of "standardised form".

For any assumed mixing distribution over γ parametrised by $\alpha \in \mathcal{A}$, and for any true mixing distribution, conditions $C1.1$ and $C1.2$ are satisfied.

Some less abstract applications

$(Y_{i1}, Y_{i0})_{i=1}^{n}$ exponentially distributed. Symmetric parametrisation $\gamma_i\psi$ and γ_i/ψ. Treat γ_i as random with any parametric distribution. Regardless of the assumed and true random effects distribution, $\hat{\psi} \rightarrow_p \psi^*$.

Extends to the symmetric parametrisation of Weibull and gamma, with treatment effect multiplicative on the rate.

In location families the symmetric parametrisation is in the means $\gamma_i + \psi$, and $\gamma_i - \psi$.

(Other relevant groups? Theory OK for rotation models but this class of examples seems a bit contrived (??))

Unbalanced two-group problems

Let $(Y_{ij1})_{i=1}^{r_{j1}}$ and $(Y_{ij0})_{i=1}^{r_{j0}}$ be observations within the jth stratum for treated and untreated individuals respectively. Unbalanced $r_{j1} \neq r_{j0}$.

Reduce by sufficiency within treatment groups and strata: S_{j1} and S_{j0}.

Stratified two-group Poisson problem with unbalanced strata

Conditionally on γ_j, Y_{ij1} and Y_{ij0} are Poisson distributed counts of rates $\gamma_j \psi^*$ and γ_j / ψ^*, then S_{j1} and S_{j0} are sums of these counts and are Poisson distributed of rates $r_{j1} \gamma_j \psi^*$ and $r_{j0} \gamma_j / \psi^*$.

r_{j1} and r_{j0} reflect the number of patients at risk in each group; $(\gamma_j)_{j=1}^m$ is a stratum-specific nuisance parameter.

Fitted model treats $(\gamma_j)_{j=1}^m$ as gamma distributed: $\hat{\psi} \to_p \psi^*$ under any random effects distribution with the same mean.

Evans and Didelez (2024) JRSSB discussion paper

Marginal structural model in a 'frugal parametrisation'.

Nuisance parameters enter through the propensity score.

E&D model has a parameter cut, implying parameter *m*-orthogonality when the propensity score is misspecified.

The proof of E&D's main theorem implicitly establishes the remaining condition. See HB discussion of E&D.

The end