Sections and local group meeting reports.

Inducement of population-level sparsity

Written by **Gilbert MacKenzie** on February 19th, 2023. Posted in Section and local group meeting reports.

The Northern Ireland local group of the RSS held an online meeting on Wednesday, February15th., 2023 at 1 pm (GMT), using MS Teams.

The speaker was Dr. Heather Battey, Department of Mathematics, Imperial College, London, UK.

Dr Battey explained that her talk was motivated by a situation where the number of nuisance parameters was large (high-dimensional) leading to the breakdown of classical maximum likelihood theory. The main idea was to avoid this situation by inducing sparsity in the Fisher Information matrix (orthogonalization) and, ideally, producing a function of the parameter(s) of interest to be used for inference. The benefits of such sparsity were deemed to be: (a) interpretative and (b) statistical (e.g., reduced bias and improved efficiency).

An important question was whether a general procedure could be developed? This was an open, and challenging task which might first be explored by reviewing some successful examples.

Example 1: *Parameter Orthogonalization*

If the parameters are (ψ, ϕ) where ψ is the parameter of interest and ϕ is the "nuisance", we should prefer an *interest-respecting reparametrization* $(\psi, \lambda(\psi, \phi))$ chosen to make λ orthogonal to ψ. In other words, to induce sparsity on $i_{\psi.\lambda}$. This, Heather explained, is operationalised by solving a system of partial differential equations (Cox & Reid, 1987) and recent work induces sparsity in the sense $i_{\psi.\lambda} \to 0$ as $n, p \to \infty$.

Example 2: *Covariance Models*

For a given (relevant) covariance model, with Σ) not obviously sparse in any domain, can a sparsity-inducing parametrization be deduced? By using, $\Sigma = \exp(L)$, where L is the matrix logarithm of Σ, and *assuming* the existence of a suitable sparsity scale (Biometrika, 2019). Heather provided a *proof-of-concept* example. The idea is to transform to a sparser covariance structure, via the matrix logarithm, perform inference there on the parameters of interest, and then transform back to the original scale. The details of the procedure are rather technical, involving transformations of the eigenvalues of the original covariance matrix and are covered in Battey (Bernoulli, 2017). Although, the proof-of -concept is clear, with many properties established, this remains in Heather's view an open problem.

Example 3: *Factorizable Transformations of data*

Here the goal is similar, it is inference on a treatment effect ψ in the presence of nuisance parameters, say, $\lambda_1, \ldots, \lambda_b$. It is sometimes possible to eliminate the b nuisance parameters by transforming the *observations* in such a way that the resulting likelihood function, $L(\psi, \lambda; x)$, factorizes as

$$L(\psi, \lambda; x) = L(\psi; x) \times L(\psi, \lambda; x)$$

a partial likelihood, with most of the information retained in $L(\psi; x)$. To illustrate, Heather presented Lindsay's (1980) example with b matched pairs of homozygotic twins in which one twin is treated (T) and the other is control where T_i and C_i are exponentially distributed with rates $\lambda_i \psi$ and λ_i/ψ, respectively. Then the marginal density of $S_i = T_i/C_i$ at s is $f_S(s) = \psi^2/(1 + \psi^2 s)^2$ which does not depend on λ_i. The S_i are independent and identically distributed and can be used for likelihood-based inference on ψ. The search for such transformations $S(\cdot)$ is based on sparsity induced by $\nabla_\lambda f_S(s)$.

Example 4: *Inference in high-dimensional linear regression*

Consider the p-dimensional regression model with coefficient vector β:

$$Y = X\beta + \epsilon = x_v \beta_v + X_{-v}\beta_{-v} + \epsilon \qquad (*)$$

where ϵ is an error term with mean zero and variance τ where β_{-v} is the nuisance parameter. Sparsity is sought by pre-multiplying (*) by an $n \times n$ matrix A^v designed to maximize the orthogonality between x_v and X_{-v}. In practice, exact orthogonalization is not achievable from this pre-multiplication strategy, so some bias is to be expected from the simple least squares regression of Y_v on the single column of the transformed x_v. Heather provided a simple LS estimator for $\tilde{\beta}_v$, the sparse estimator of β_v and discussed its properties.

Heather concluded by remarking that her talk attempted to synthesize four ideas through a unifying theme of sparsity inducement: a search for parameterizations or data transformations that yield zeros or near-zeros as components of key population-level objects.

Her talk, which was a *tour de force*, was received with acclaim. The select audience showed their appreciation in the usual way. This was a rich set of ideas. There followed a discussion of several issues. The connection with the Lasso was raised as were the difficulties surrounding the search for a unifying paradigm.

The Chair thanked the speaker for a very stimulating talk and concluded the meeting by thanking everyone for their attendance and support.

Key References

Battey, H.R. (2019). On sparsity scales and covariance matrix transformations. Biometrika, 106, 605-617.

Battey, H. S. (2017). Eigen structure of a new class of structured covariance and inverse covariance matrices. Bernoulli, 23, 3166-3177.

https://www.ma.imperial.ac.uk/~hbattey/