TMB

Adam Howes, Imperial College London

Consider unobserved latent random effects $\mathbf{x} \in \mathbb{R}^n$ and parameters $\boldsymbol{\theta} \in \mathbb{R}^m$.[1] Let $\ell(\mathbf{x}, \boldsymbol{\theta}) \triangleq -\log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta})$ be the negative joint log-likelihood. In `TMB`, the user writes C++ code to evaluate this negative log-likelihood function ℓ. A standard maximum likelihood approach is to optimise

$$L_\ell(\boldsymbol{\theta}) \triangleq \int_{\mathbb{R}^n} p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}) \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^n} \exp(-\ell(\mathbf{x}, \boldsymbol{\theta})) \mathrm{d}\mathbf{x} \tag{1}$$

with respect to $\boldsymbol{\theta}$ to find the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}$. Taking a superficially more Bayesian approach than above, instead of ℓ, the user may instead write a function to evaluate the negative joint penalised log-likelihood given by

$$f(\mathbf{x}, \boldsymbol{\theta}) \triangleq -\log p(\mathbf{y} \,|\, \mathbf{x}, \boldsymbol{\theta}) p(\mathbf{x}, \boldsymbol{\theta}) = \ell(\mathbf{x}, \boldsymbol{\theta}) - \log p(\mathbf{x}, \boldsymbol{\theta}), \tag{2}$$

equivalent up to an additive constant to the negative log-posterior.

$$f(\mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}) - C, \tag{3}$$

where $C = \log p(\mathbf{y})$ is the log evidence. Using f in place of ℓ, then the penalised likelihood is proportional to the posterior marginal of $\boldsymbol{\theta}$

$$L_f(\boldsymbol{\theta}) \triangleq \int_{\mathbb{R}^n} \exp(-f(\mathbf{x}, \boldsymbol{\theta})) \mathrm{d}\mathbf{x} \propto \int_{\mathbb{R}^n} p(\mathbf{x}, \boldsymbol{\theta} \,|\, \mathbf{y}) \mathrm{d}\mathbf{x} = p(\boldsymbol{\theta} \,|\, \mathbf{y}). \tag{4}$$

Integrating out the random effects directly, as in Equation 4 above, is usually intractable because $\mathbf{x}$ is high-dimensional, so Kristensen (2016, Equation 3) use a Laplace approximation $L_f^\star(\boldsymbol{\theta})$ based instead upon integrating out a Gaussian approximation to the random effects. This Laplace approximation is analogous to the INLA approximation $\tilde{p}(\boldsymbol{\theta} \,|\, \mathbf{y})$.

$$f''_{\mathbf{x}\mathbf{x}}(\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = -\frac{\partial^2}{\partial \mathbf{x}^2} \log p(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}) \Big|_{\mathbf{x} = \hat{\boldsymbol{\mu}}(\boldsymbol{\theta})} = -\frac{\partial^2}{\partial \mathbf{x}^2} \log p(\mathbf{x} \,|\, \boldsymbol{\theta}, \mathbf{y}) \Big|_{\mathbf{x} = \hat{\boldsymbol{\mu}}(\boldsymbol{\theta})} = \hat{\boldsymbol{Q}}(\boldsymbol{\theta}).$$

Inference proceeds by optimising $L_f^\star(\boldsymbol{\theta})$ via minimisation of

$$-\log L_f^\star(\boldsymbol{\theta}) \propto \frac{1}{2} \log \det(\hat{\boldsymbol{Q}}(\boldsymbol{\theta})) + f(\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \boldsymbol{\theta}), \tag{5}$$

where $\propto$ is used to mean proportional up to an additive constant. The parameters of the Gaussian approximation, are found in terms of f via $\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \arg\min_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\theta})$ and $\hat{\boldsymbol{Q}}(\boldsymbol{\theta}) = f''_{\mathbf{x}\mathbf{x}}(\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \boldsymbol{\theta})$ and must be recomputed for each value of $\boldsymbol{\theta}$. Obtaining $\hat{\boldsymbol{\mu}}(\boldsymbol{\theta})$ is known as the inner optimisation step.

[1] Kristensen (2016) use the notation u for random effects and θ for parameters. We aim for consistency with Section **??**.