Errors and uncertainty in variables – When to worry and when to Bayes?

Stefanie Muff

The existence of uncertainty and measurement error (ME) in medical, environmental or ecological variables is an undeniable fact, and most researchers are aware of it. The sources of ME are manifold and imply much more than just instrumental imprecision in the measurement of physical variables, but may include for instance biases due to preferential sampling, incomplete observations or misclassification. Despite this, an often preferred way to dealing with the presence of ME is to ignore it, especially in covariates of regression models, although this may lead to seriously biased parameter estimates and confidence intervals in statistical modelling procedures (Fuller, 1987; Carroll et al., 2006). Moreover, the bias in the effect estimates is often, but not always, towards the Null (Muff and Keller, 2015).

In the presence of ME, a researcher may concerned with the question whether there is a reason to worry. In simple cases, analytical formulas to estimate the bias in the parameter estimates exist, so that the effect of the error can be assessed. Otherwise it may be useful to carry out a small simulation study. If the researcher concludes that the bias cannot be tolerated, it is crucial to formulate an error model.

Bayesian methods for error modelling have proven to be particularly useful (Gustafson, 2005), namely because it is straightforward to include prior knowledge about the error in terms of prior distributions. The posterior distribution can then be estimated by Markov chain Monte Carlo (MCMC) methods, or, if certain modelling assumptions are fulfilled, by integrated nested Laplace approximations (INLA Rue et al., 2009; Muff et al., 2015). In summary, the intention of this talk is to raise awareness for the impacts of measurement error and uncertainty in the data, but also to provide guidance on what to do in such cases.

References

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (2 ed.). Boca Raton: Chapman & Hall.

Fuller, W. A. (1987). *Measurement Error Models.* New York: John Wiley & Sons.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statistical Science 20*, 111–140.

Muff, S. and L. F. Keller (2015). Reverse attenuation in interaction terms due to covariate error. *Biometrical Journal 57*, 1068–1083.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 64*, 231–252.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological) 71*, 319–392.