Statistical Science
2011, Vol. 26, No. 2, 296–298
DOI: 10.1214/11-STS349REJ
Main article DOI: 10.1214/10-STS349

Rejoinder

Carl Morris

I thank all four discussants for their valuable insights. Before responding to their specific comments, let me help clarify to readers that adjustment for density (or likelihood, if appropriate) maximization is a method for approximation and not a stand-alone procedure for inference. The favorable frequency properties of the ADM-SHP procedure rely particularly on the flat prior chosen for the random effects variance A. After that, the responses of Partha Lahiri and Santanu Pramanik and of Claudio Fuentes and George Casella are addressed.

Because shrinkage factors B_j are constrained to $[0, 1]$, a Beta distribution ostensibly serves as a better approximation to the likelihood function or posterior density of B_j than does a Normal distribution. MLE and ADM methods are fitted based on computing two derivatives, and they agree exactly when a Normal density is chosen to approximate a likelihood function (or a posterior with a flat prior). However, as Lahiri and Pramanik's Figure 2 shows, sometimes no Normal distribution can closely approximate the distribution of a shrinkage factor B_j and then the MLE will yield misleading inferences unless it can be liberated from its usual Normal approximation.

ADM (Morris, 1988) was designed to approximate a given (one-dimensional) distribution with any chosen Pearson family, perhaps with a Normal distribution if for MLE purposes, or a Beta for shrinkage factors, or a Gamma, an Inverted Gamma, an F, a t, or a *Skew-t* distribution for other situations. ADM does not alter a posterior density or a likelihood function. The new curve that the ADM creates via multiplication by the "adjustment" (A, in this paper) has no meaning other than to provide a mode in the interior of the parameter space that one believes will lie closer to the mean of the actual density, or likelihood.

Carl Morris is Professor of Statistics, Department of Statistics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138, USA (e-mail: morris@stat.harvard.edu). Ruoxi Tang obtained his Ph.D. from Harvard's Statistics Department and is with a New York investment firm, Bloomberg L.P., 731 Lexington Ave., New York, New York 10022, USA.

The statistical properties of the ADM approximation depend crucially on the corresponding properties of the procedure it approximates. While ADM can be used to approximate various Bayes procedures, for proper and for improper priors, that is not the goal in this paper. Rather, the objective is to provide estimates of shrinkage factors via calculations similar to those of MLE procedures that improve on the MLE for resulting inferences about random effects. The flat prior on A was chosen neither for Bayesian reasons nor for subjective reasons, but because it leads to Stein's harmonic prior (SHP) on the Level-I parameter vector θ and yields formal Bayes point estimators of the random effects with verified and dominant mean squared error risks in the frequency sense. The paper provides additional strong evidence that the formal Bayes posterior intervals, whether computed exactly or as approximated by ADM, meet (or nearly meet) their nominal (95% in the paper) confidence coverage rates in the equal variance two-level Normal model, whatever be the unknown between groups parameters β, A.

Crucially, the conditional Level-II mean and variance of each random effect θ_j depends linearly on B_j and nonlinearly on A. For that reason ADM, which is designed to approximate a mean, starts in this application by approximating B_j with a Beta distribution, rather than applying ADM directly to A (perhaps with an approximating F or a Gamma distribution). By good fortune this turns out to be equivalent to setting $\hat{A} = \operatorname{argmax}(AL(A))$ with $L(A)$ the likelihood function [or perhaps a REML version of $L(A)$ if $r \geq 1$] so that A legitimately can be viewed as a likelihood "adjustment." However, this adjustment actually arises as a principled choice based on three considerations: (a) the established frequency properties of formal Bayes procedures that stem from SHP; (b) the ADM approximation that uses a Beta distribution, for which the adjustment is $B_j(1 - B_j)$; and (c) that the shrinkage factor B_j enters linearly in the first two Level-II moments, given (β, A), of θ_j.

Perhaps other confidence interval shrinkage procedures for the Normal two-level model have been proven to do as well by frequency standards as the procedures based on SHP and its ADM-SHP hybrid here. We know from Figures 6 and 7 that coverage rates