

Robust Importance Sampling for EM-Based NLME Estimation Algorithms

Robert H. Leary and Michael R. Dunlavey

Certara/Pharsight® Corporation

Pharsight

Introduction and Overview

For EM-based NLME estimation algorithms such as QRPEM in Phoenix NLME, IMP in NONMEM, and MCPPEM in S-ADAPT, a Monte Carlo or quasi-Monte Carlo integration to estimate moments of a target empirical Bayesian (EB) posterior density for each individual must be performed at each major iteration. Since it usually is not possible to sample directly from the EB density, the samples are taken from a (usually Gaussian) importance sampling density $N(\mu, \lambda C)$, where the mean μ of the ISD is taken to be the mean or mode of the EB density as estimated on the previous iteration. The covariance λC of the ISD is simply a scaled version of the estimated covariance C of the EB density with a scale factor $\lambda > 1$ to widen the tails of the ISD relative to the target EB density. Thus the ISD is tuned to mimic the target density version fairly closely in the modal region, but with somewhat wider tails (See Fig. 1 below).

The scaling is critical to the success of the overall EM algorithm, and scale factors λ that are too high or too low will adversely affect performance and estimation accuracy to the point where a mis-scaled attempt may simply fail. The window of workable scale factors may be narrow, not knowable in advance, and the standard default factor may not give good results.

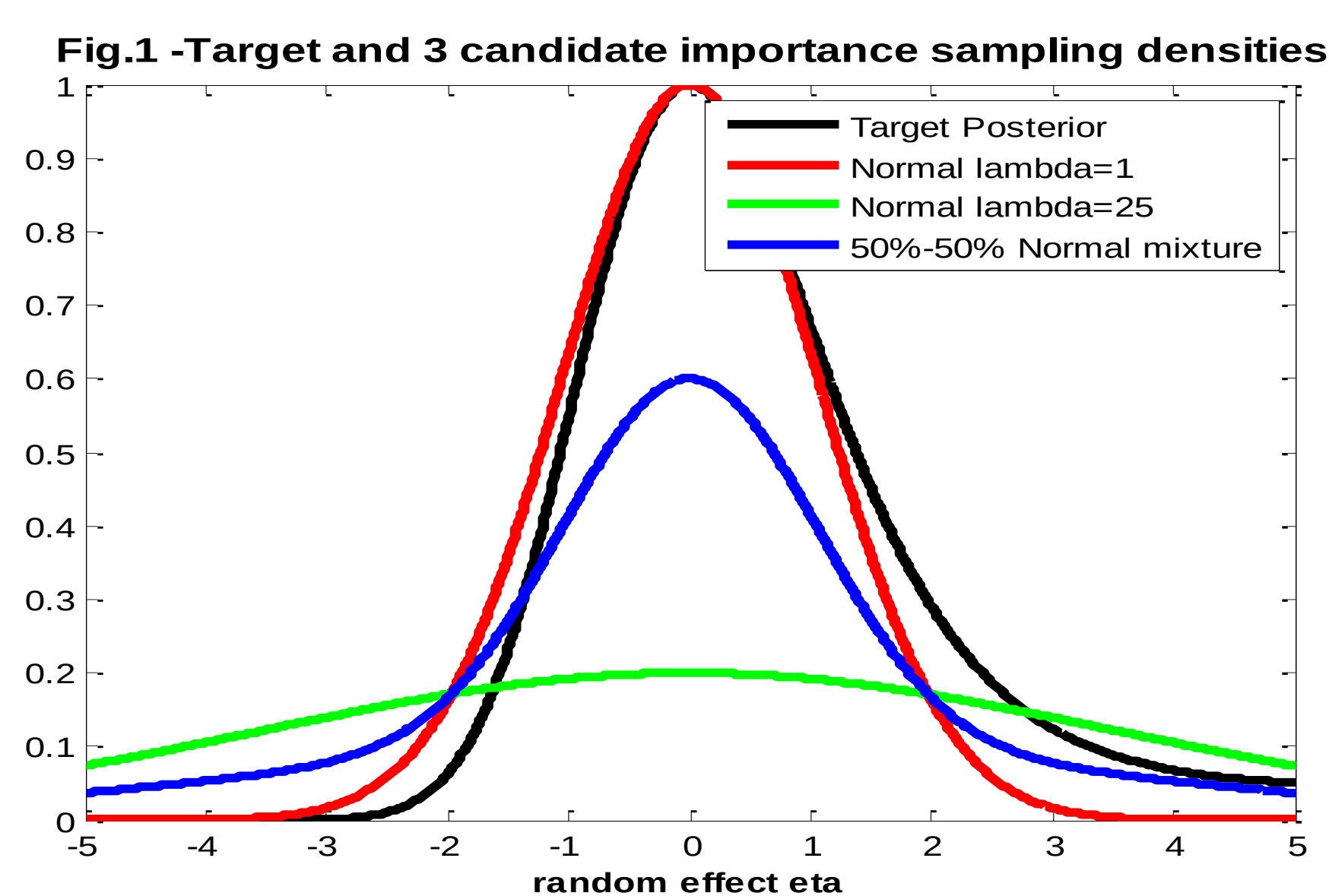
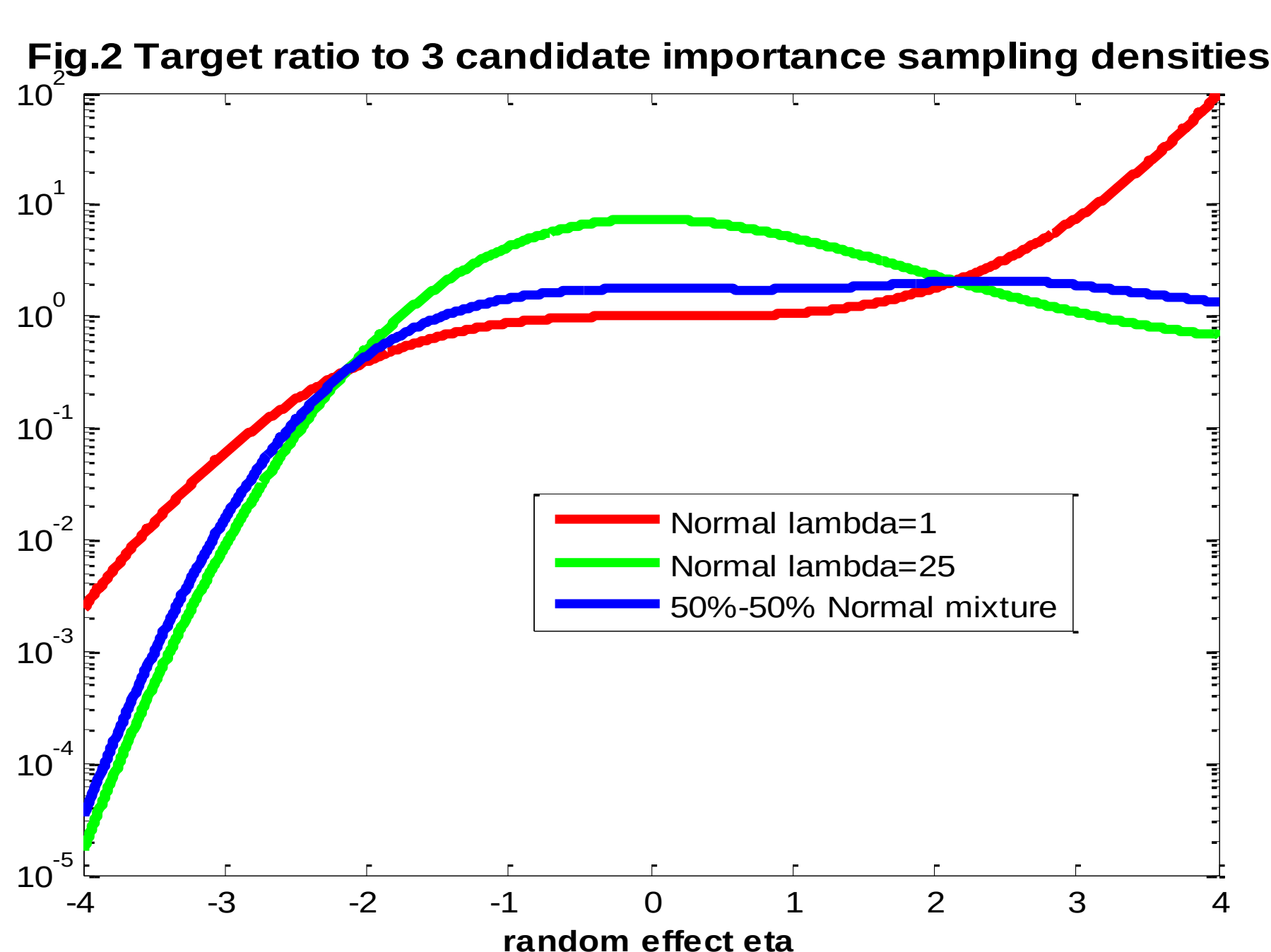


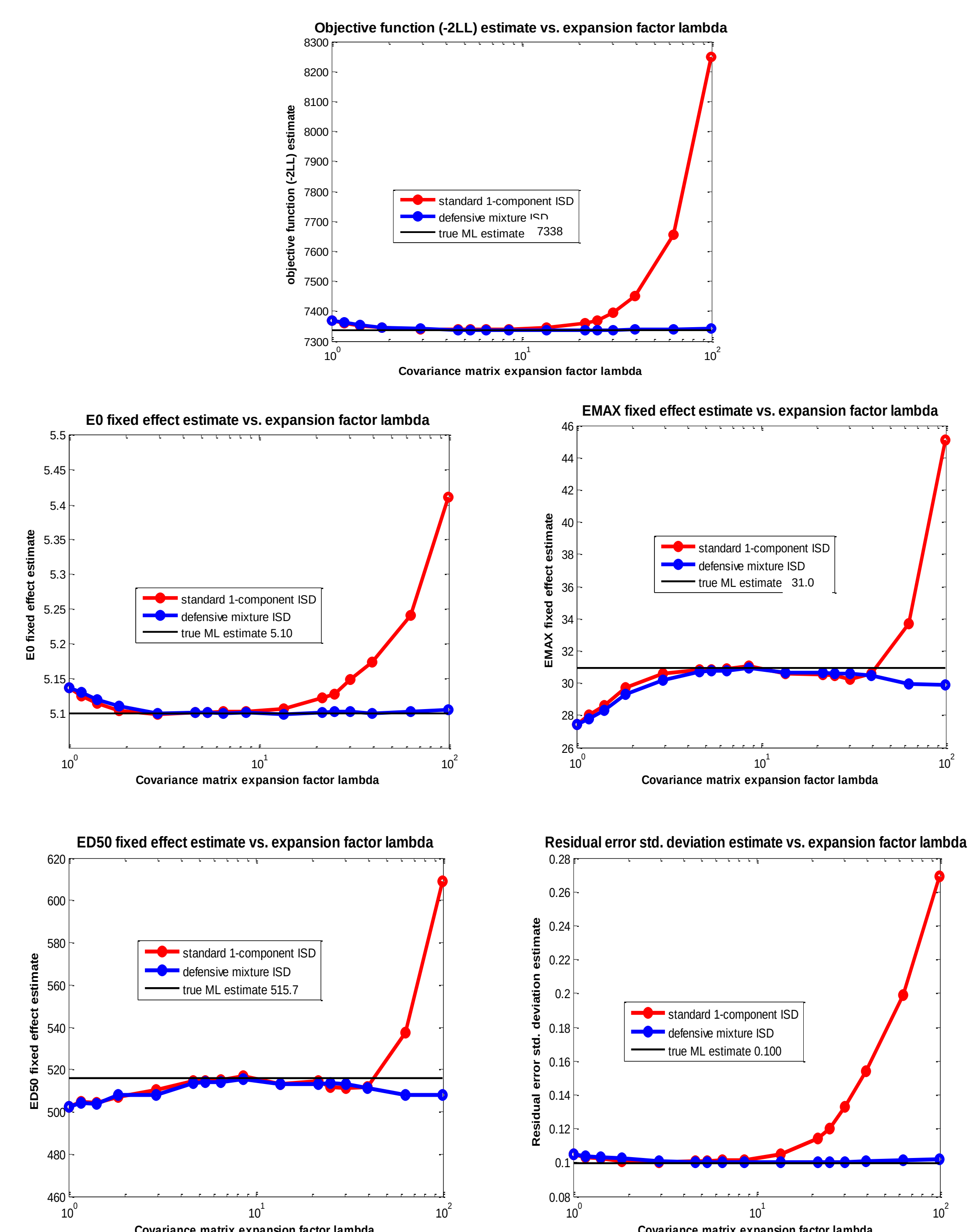
Figure 1 above shows a sample 1D target EB density that is skewed to the right with a slowly decaying right tail. The best approximating normal density ($\lambda=1$) in the central region fails to cover the right tail and will be a poor or completely unworkable ISD. A widened ($\lambda=25$) normal density provides adequate tail coverage but at the cost of greatly lower efficiency in the central region. But a 50%-50% mixture of the $\lambda=1$ and $\lambda=25$ cases provides adequate tail coverage and reasonably good central region behavior.

Figure 2 below shows the critical ratio of the target EB to importance sampling density as a function of eta, the random effect (shown here as a 1D variable for simplicity, but more generally this is multi-dimensional). The $\lambda=1$ case leads to a large and increasing ratio in the right tail, a classic indication of a poor ISD that will almost certainly fail. The $\lambda=25$ case captures the right tail acceptably but at the cost of greatly under-sampling the center. The 50%-50% mixture is a much more reasonable compromise.



Defensive Mixture ISD's

Hesterberg [1] proposes the use of an importance sampling density that is a 'defensive mixture' of Gaussians with differing tail coverage characteristics, for example, $0.5 \cdot N(\mu, C) + 0.5 \cdot N(\mu, \lambda C)$ where λ is a scale factor. Owen and Zhou [2] suggest the addition of control variates to improve accuracy and avoid pathologies in the mixture sampling case. We introduced both techniques into QRPEM and compared both techniques with the traditional single Gaussian ISD on a variety of models. Here we report representative comparative results on a single test model without control variates, which are still under evaluation but appear promising. The test model is a rather difficult sigmoidal EMAX model adapted from [3] with log normally distributed parameters E_0 , $EMAX$, and ED_{50} (for example, $E_0 = tvE_0 \cdot \exp(\eta E_0)$, where tvE_0 is a fixed effect and ηE_0 a random effect). The model includes a proportional residual error model, and a fixed Hill exponent of 3.0. Previously we have used this same model to demonstrate the rather impressive importance improvements of quasi-random relative to Monte Carlo sampling [4]. Here the model was run over a wide range of scale factors λ with a standard 1-component ISD = $N(\mu, \lambda C)$ vs. a Hesterberg-type defensive mixture ISD = $0.5 \cdot N(\mu, C) + 0.5 \cdot N(\mu, \lambda C)$ using a quasi-random sample size of 1000.



Results and Discussion

For all values of the covariance expansion factor $\lambda > 5$, the defensive mixture reproduced the true ML estimates (obtained from using a very large sample size QRPEM run and confirmed independently with the high accuracy Adaptive Gaussian Quadrature engine in Phoenix NLME) very closely. However, the single component normal ISD started to show significant deviations from the ML estimates for $\lambda > 10$. Thus the window for acceptable performance for the single component normal ISD is $5 < \lambda < 10$ while that for the defensive mixture is far more robust and simply $\lambda > 5$. (In terms of the Acceptance Ratio, which allows the user to control the expansion factor and has a default value of 0.1 in QRPEM and 0.4 in IMP, the single component window is [0.03, 0.9], which is extremely narrow and lies outside both defaults)

Finally we note that the 1000 samples in the 'defensive mixture' were generated by first 500 samples from $N(\mu, C)$, and then 500 from $N(\mu, \lambda C)$. However, the mixture density was used in the importance sampling weighting factor for all 1000 samples. — i.e., the knowledge of which component a sample came from is not used. This forcing of the sample proportions to the true mixture weights is recommended by Hesterberg, but in fact is actually necessary to preserve the low discrepancy properties of the QR sampling and its advantage over Monte Carlo sampling.

A 2- and 3-component implementation of defensive mixtures (without control variates) will be available for Phoenix QRPEM in the next Phoenix 1.4 release.

References

1. T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions, Tech. report no. 148, Division of Biostatistics, Stanford University
2. A. Owen and Y. Zhou, "Safe and effective importance sampling", JASA 95, 2000, pp.135-143.
3. E. Plan et al., PAGE 19 (2010) Abstract 1880.
4. R. Leary and M. Dunlavey, PAGE 22 (2012), Abstract 2797.