A Fast Bootstrap Method Using EM Posteriors

Robert H. Leary, Michael R. Dunlavey, and Jason Chittenden

Certara/Pharsight® Corporation

Overview and Objectives

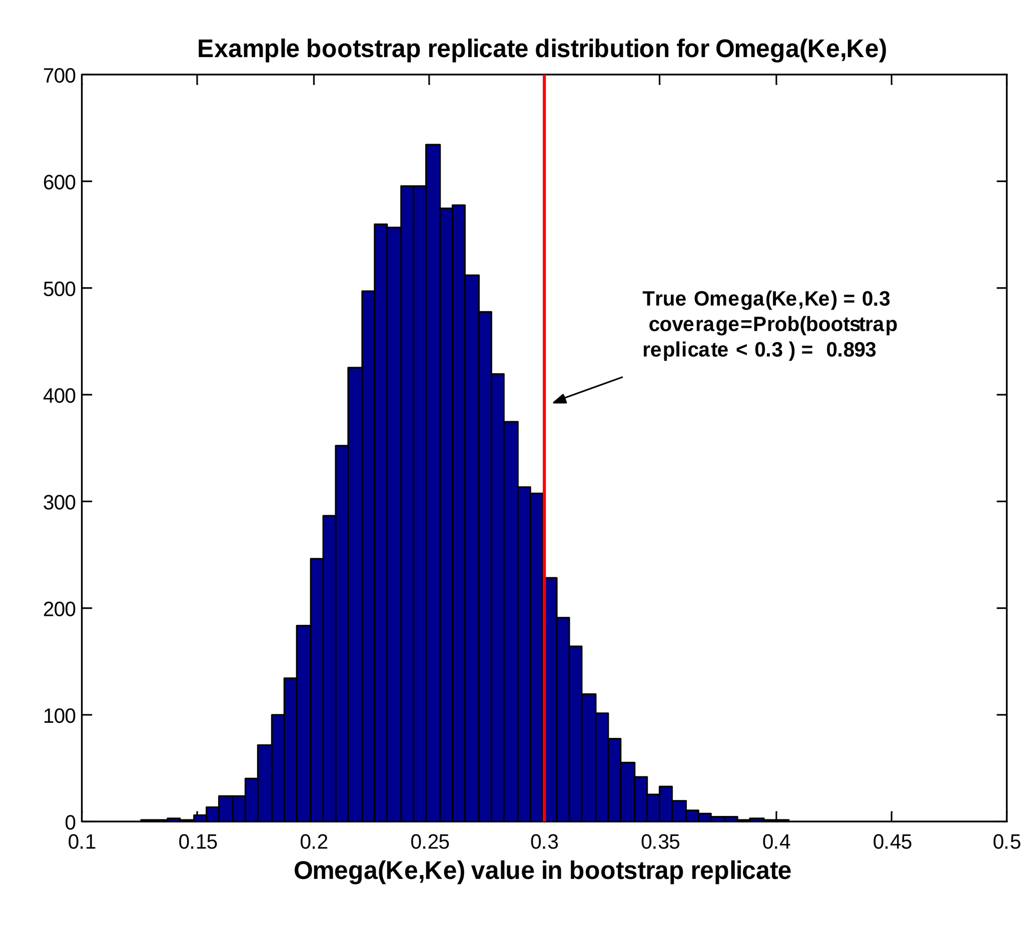
In the most usual form of NLME bootstrapping for parameter uncertainty estimation, each replicate data set is constructed by pooling *Nsub* (= total number of subjects) random selections of individual data sets with equal probability *1/Nsub* on each individual. In effect, the bootstrap replicates are the concatenation of *Nsub* individual data sets from a mixture distribution of individual data sets with equal probabilities on each set. An intriguing analogy occurs in the optimal NLME estimation solution via EM methods. At least in simple cases where all fixed effects are interpretable as structural parameter means, at the optimum (maximum likelihood) solution, the fixed effects parameters (thetas) are the means, and the random effect parameters (Omega) are the variances/covariances, of the mixture distribution of posteriors. Thus in these simple cases knowing the means and covariance matrices for each subject's posterior at the maximum likelihood point is sufficient to recover the optimal fixed and random effect values with a very fast computation involving just some basic linear algebra.

This suggests the possibility of a very fast bootstrapping procedure. Rather than assembling new replicate data sets and resolving the estimation problem for each one, simply resample the posteriors (means and variance/covariance matrices) and compute fixed and random effects from the resampled posteriors. The objective here is to evaluate the performance of this procedure on a simple test problem.

Test Problem and Evaluation Procedure

1000 simulated data sets were created with *Nsub*=100 subjects each for a simple IV bolus model C(t)=Dose*exp(-Ke*time)/V with additive residual error (standard deviation =0.05). Structural parameters Ke = exp(tvlogKe+etaKe) and V = exp(tvlogV+etaV) were log normally distributed with diagonal Omega where Omega(V,V)=0.4, Omega(Ke,Ke)=0.3,. Fixed effects were tvlogV=log(0.6), tvlogKe=log(0.2). All subjects had a common single bolus Dose=1 at t=0 and fixed sampling times t=[1,2,3,4].

For each data set, parameters were estimated with the Phoenix NLME QRPEM algorithm to obtain posterior means and variance/covariance matrices for each subject at optimality. For each data set, 10000 bootstrap replicates were taken from these posteriors and the fixed and random effects parameters computed. Figure 1 below shows a histogram of the 10000 bootstrap Omega(Ke,Ke) values for a single example data set



For each parameter and data set a coverage statistic CVG was generated as the cumulative distribution function of the true value of the parameter within the bootstrap distribution of that parameter. In the above example, 8912 out of 10000 of the Omega(Ke,Ke) values in the bootstrap distribution were below the true value Omega(Ke,Ke)=0.3, so the Omega(Ke,Ke) coverage statistic CVG for this data set is 0.891. Note that ideally the coverage statistics should have an empirical uniform [0,1] distribution over all 1000 data sets. In this ideal case, an upper X% confidence limit generated as the upper X percentile of the bootstrap distribution is 'correct' in the sense that the observed frequency of the true value falling below this limit in the 1000 data sets is approximately X%. So a convenient method of evaluation is to plot the empirical cdf of the 1000 coverage statistics and compare to the ideal linear behavior of the cdf of a uniform[0,1] distribution.

Results

Fixed Effects

Figures 2 and 3 show the empirical cdfs (blue) of the 1000 CVG statistics for the fixed effects tvlogV and tvlogKe, respectively, against the ideal linear behavior (red)

Figure 2 - tvlogV Coverage

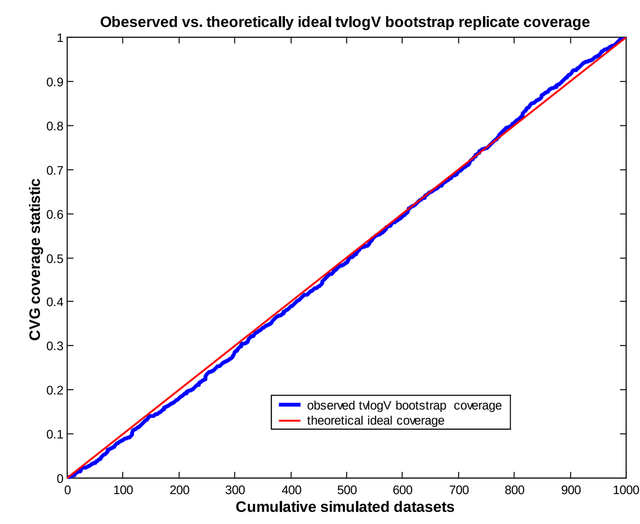
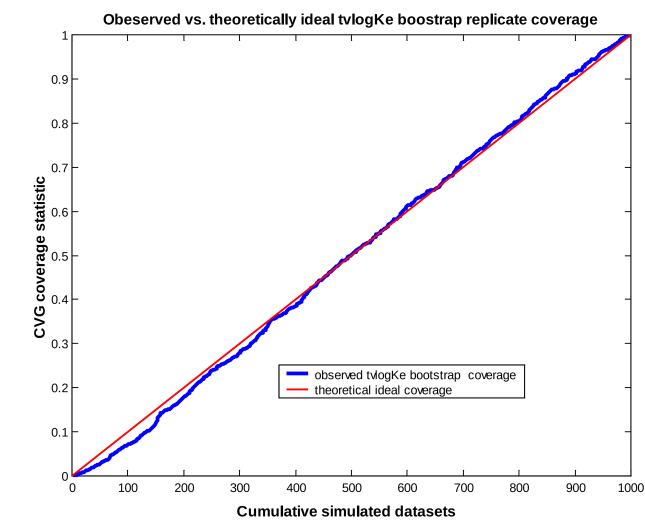


Figure 3 - tvlogKe Coverage



The random effect coverage is does not significantly deviate from ideal linear behavior. Thus the bootstrap distributions appear to be unbiased and effective for purposes of identifying reliable confidence limits with the proper coverage levels. For example, an upper 80% confidence limit for tvlogV should cover the true value tvlogV=log(0.6) in approximately 800 of the 1000 data sets – the actual observed number was 804 data sets.

Random Effects

Figure 4 – Omega(V,V) Coverage

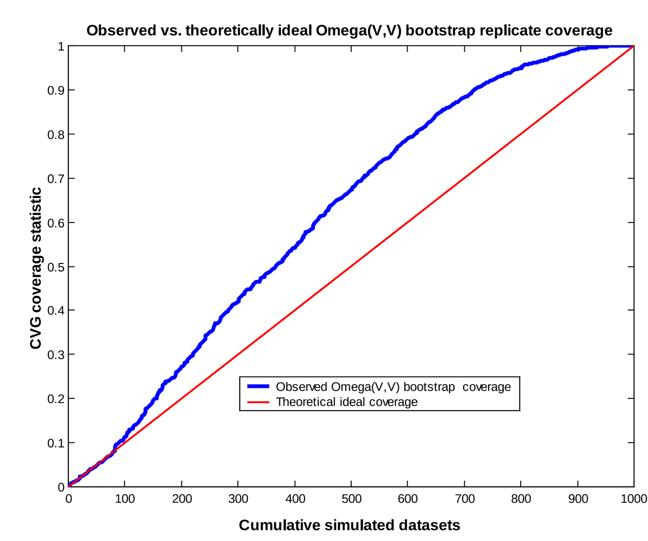
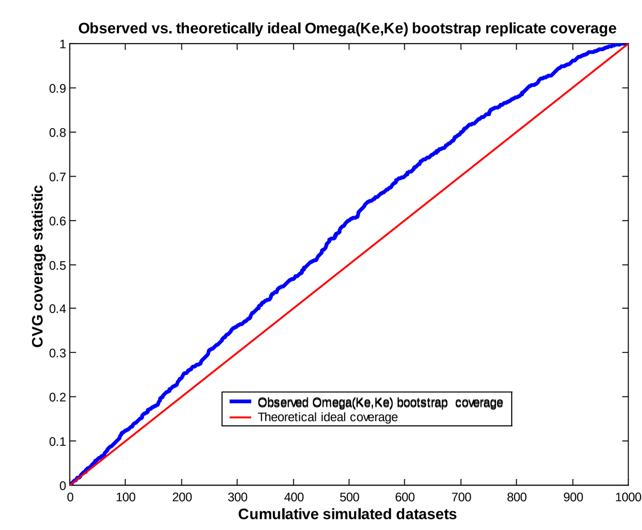


Figure 5 – Omega(Ke,Ke) Coverage



The observed CVG cdf's (blue) in both cases lies significantly above the ideal linear CVG cdf (red) over the entire range. This means that in general for this model, the bootstrap distribution is shifted left - i.e., the bootstrap estimates are biased low, and confidence levels are somewhat imprecise with lower actual coverage than the confidence level specifies. For example, an upper 80% confidence limit for Omega(Ke,Ke) in Figure 5 should in fact cover the true Omege(Ke,Ke) value 0.3 in 80% (800 of 1000) of the data sets – the actual coverage is 703 data sets (70.3%). The equivalent observed coverage for an upper 80% confidence limit on Omega(V,V) in Figure 4 is only 594 data sets out of 1000 (59.4%).

Discussion

A primary advantage of the posterior-bootstrapping methodology, at least in simple cases, is its extraordinary speed. The total computational time to evaluate 10000 bootstrap samples for a single data set was basically negligible (on the order of 1 sec). While we have not shown it here, the method is readily extensible to residual error parameters. The method can also be extended to handle fixed effects parameters (there are none in this example) that are not paired with random effects and hence are not directly estimable from just a knowledge of the basic statistics of the posterior distributions. However, this will raise the computational cost considerably, but it will still likely be well below that of conventional bootstrap methods.

The quality of the bootstrap coverage, at least for the fixed effects, is remarkably good over the entire cdf range. The quality of the coverage for random effects is considerably lower, but we believe may still be reasonable, for example, for the purposes of computing approximate standard errors. We are in the process of making comparisons with conventional bootstrap results in the random effects case, but as yet do not have definitive results.

It must be emphasized that this is a pilot study, with just a single very simple example model, so we cannot yet generalize to how well the method will perform on other models. However, the results, particularly in the case of fixed effects, suggest that at least for some purposes the method may be a reasonable and much faster alternative than the conventional bootstrap.

Contact

Robert H. Leary, PhD, Certara/Pharsight Corporation,
Ph: +1 919.852.4625, Email: Bob.Leary@Certara.com