Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

BETTER INPUT MODELING VIA MODEL AVERAGING

Wendy Xi Jiang
Barry L. Nelson

Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60208-3119, USA

ABSTRACT

Rather than the standard practice of selecting a single "best-fit" distribution from a candidate set, frequentist model averaging (FMA) forms a mixture distribution that is a weighted average of the candidate distributions with the weights tuned by cross-validation. In previous work we showed theoretically and empirically that FMA in the probability space leads to higher fidelity input distributions. In this paper we show that FMA can also be implemented in the quantile space, leading to fits that emphasize tail behavior. We also describe an R package for FMA that is easy to use and available for download.

1 INTRODUCTION

Stochastic simulation is a method for analyzing the performance of a system that includes interactions among stochastic processes. At a high level, a stochastic simulation consists of two parts: inputs and logic. The inputs are the uncertain components of a system, while the logic is a collection of rules and algorithms that govern the behavior of the system as a function of the inputs (Nelson 2013). Inputs are typically described by fully specified probability models, which includes the case of resampling a fixed set of data. *Input modeling*, as the name suggests, is the process of choosing simulation input models to approximate the uncertainty in the target system. In this paper we are interested in situations when the input models are "fit" to a relevant sample of real-world data. For instance, we later model data on lot sizes of surface mount capacitors in a manufacturing simulation from Wagner and Wilson (1996).

Despite many advances in input modeling for complex situations—including non-stationary arrival processes, time-series inputs and heterogeneous random-vector inputs—basic univariate input modeling in practice has not advanced much in decades: fit a collection of plausible distributions via methods such as maximum likelihood estimation (MLE) and select one of the candidates based on some ranking (e.g., goodness-of-fit test) or graphical assessment, perhaps giving extra consideration to a candidate suggested by the real process physics (e.g., sums of more basic random outcomes tend to be normally distributed). Yet under almost all circumstances the real-world data should not be expected to follow *any* of the candidate distributions exactly: mathematical distributions are idealizations that describe some precisely defined process physics (often in a limit), while the real-world processes generating the data have quirks and oddities that make them "real world."

The preceding paragraph might seem to suggest that one should avoid selecting a mathematical distribution altogether, and instead just resample the available real-world data to drive the simulation. This can be a good choice, but empirical distributions are inherently discrete (putting probability mass only on the sampled values), and therefore manifest gaps and lack a (possibly important) tail. Which begs the question, how much data are enough to forego the smoothing and inferred tails obtained by fitting a distribution?

Recently, Nelson et al. (2018) introduced frequentist model averaging (FMA) as an effective way to build input models that better represent the true, unknown input distributions, thereby reducing errors when

making inference back to the real world. The premise of FMA is that there *may*, and often will, be one or more parametric probability distributions that fit the real-world data reasonably well, but not perfectly if employed alone. Therefore, FMA averages or mixes the candidate set of distributions to extend their reach, with the ultimate goal of getting the most fidelity from a given candidate set. Cross-validation (CV) is employed to tune the mixture and avoid overfitting. Nelson et al. (2018) provide strong theoretical and experimental evidence that the FMA distribution gets as close as possible to the real-world distribution when using only the component distributions in the given candidate set. By also including the empirical distribution (ED) in the candidate set, FMA provides protection when *none* of the mathematical distributions is adequate, and explicitly indicates how adequate the ED is by the weight assigned to it.

The FMA distribution explored in Nelson et al. (2018) is fit in probability space: it minimizes the cross-validation error between the fitted and empirical cumulative probability distributions. Therefore, we refer to it as a probability model average estimator (PMAE). In this paper we introduce FMA in the quantile space, and refer to it as a quantile model average estimator (QMAE). As the name suggests, QMAE minimizes the cross-validation error between the fitted and empirical quantile functions. The choice between PMAE and QMAE depends on the application, and we show it is easy to fit both.

Although an FMA distribution is simple to use in a stochastic simulation once the fitting is complete, the fitting process itself requires solving a numerical optimization problem. The second contribution of this paper is to provide an R package for fitting and variate generation that can be downloaded from http://users.iems.northwestern.edu/~nelsonb/FMA/fma.R.

The paper is organized as follows. We describe our input model averaging method in Section 2, and the fitting of the QMAE in Section 3 (fitting the PMAE is similar and is described in Nelson et al. (2018)). Documentation of the R package we created follows in Section 4. An illustration using the package to model two datasets is found in Section 5, followed by conclusions.

2 INPUT MODELING AND INPUT MODEL AVERAGING

The most common method for selecting a marginal distribution F to represent an independent and identically distributed (i.i.d.) input process—as described in textbooks (e.g., Law and Kelton (1991)) and employed by distribution fitting software (e.g., BestFit®)—is some variation of the following:

1. Given: $x_1, x_2, ..., x_N$ an i.i.d. sample from some unknown input distribution F^c.
2. Fit $q \geq 1$ candidate parametric families of distributions $\mathscr{F} = \{F_1, F_2, ..., F_q\}$ using methods such as MLE. This yields a set of fitted distributions $\widehat{\mathscr{F}} = \{\widehat{F}_1, \widehat{F}_2, ..., \widehat{F}_q\}$.
3. Rank the fitted distributions using one or more goodness-of-fit measures and evaluate the top choices. Standard measures are hypothesis-test statistics such as chi-squared, Kolmogorov-Smirnov, Anderson-Darling and Cramér-von Mises, and likelihood-based statistics such as AIC and BIC.
4. Select $\widehat{F} = \text{Best Choice}\{\widehat{F}_1, \widehat{F}_2, \ldots, \widehat{F}_q\}$. Alternatively, use the (possibly smoothed) empirical distribution of $x_1, x_2, ..., x_N$ if nothing fits well.

In contrast to the method above that selects one element of $\widehat{\mathscr{F}}$, the FMA approach of Nelson et al. (2018) creates an "input model average" of the fitted distributions. This is different from finite-mixture models, such as McLachlan and Peel (2004), that assume the true distribution F^c is a mixture of instances of a common parametric family (e.g., normal). Instead, the premise is that there are one or more parametric families of distributions in $\mathscr{F}$ that are plausible choices, perhaps supported by real-world process physics. Therefore, the first two steps above are adopted, but rather than ranking the fitted candidate distributions and selecting the best, the mth fitted distribution is assigned a weight $w_m \geq 0$ with $\sum_{m=1}^{q} w_m = 1$. Thus, the model average distribution is

$$\widehat{F}(x, \mathbf{w}) = \sum_{m=1}^{q} w_m \widehat{F}_m(x), \tag{1}$$

where $\widehat{F}_m(x)$ is the mth fitted cumulative distribution function in the candidate set. Notice that some weights may be 0. Virtually any marginal distribution may be in the candidate set, including finite-mixture models and flexible families such as the generalized lambda distribution (Karian and Dudewicz 2000), as well as the standard choices of normal, lognormal, exponential, gamma, Weibull, etc. *The key to FMA is selecting the weights* $\mathbf{w}$ *to obtain a better fit.*

It is clear from (1) that $\widehat{F}(x,\mathbf{w})$ includes each of the individual fitted candidate distributions as a special case of $\mathbf{w}$, while increasing their flexibility by allowing mixtures. Thus, FMA maintains the benefits of the tried-and-true families which, for sound theoretical reasons, are often good approximations, but provides additional degrees of freedom for adjusting to the complexities of real data.

For PMAE, the optimal weight is obtained through minimizing the difference between $\widehat{F}(x,\mathbf{w})$ and $F^c(x)$ in a comprehensive way that guards against overfitting. Specifically, Nelson et al. (2018) solved for $\mathbf{w}$ to minimize the cross-validation squared error with the ED, a consistent and unbiased estimator of $F^c(x)$. They proved that when the true distribution is *not* in the candidate set—which we never expect it to be with real data—then the optimal cross-validation weights converge to the optimal weights as $N \to \infty$. Additionally, Nelson et al. (2018) showed that when the candidate set includes the ED then the PMAE is consistent for F^c in the sense that the weight on the ED $\widehat{w}_{\mathrm{ED}} \xrightarrow{P} 1$ as $N \to \infty$.

Once we have the fitted weights $\widehat{w}_m,\ m = 1,2,\ldots,q$, random-variate generation is easy:

1. Select $M = m$ with probability $\widehat{w}_m,\ m = 1,2,\ldots,q$.
2. Generate $X \sim \widehat{F}_M$.
3. Repeat.

Cross-validation in the probability space is just one possible choice for fitting an FMA. Because probabilities are between 0 and 1, differences between the fitted and ED are also bounded, so large differences (say) in the tails contribute small absolute differences. This suggests we might form FMA distributions with different tail behavior if we do cross-validation in the quantile space. We introduce this new idea here.

Recall that the distributions in the candidate set are $F_m(x)$, $m = 1,2,\ldots,q$, and $\widehat{F}_m(x)$ is the fitted estimator of $F^c(x)$ under the mth candidate distribution. Let $\widehat{G}_m(u)$ be the quantile function corresponding to $\widehat{F}_m(x)$, $G^c(u)$ the quantile function of the true distribution $F^c(x)$, and $\mathbf{v} = (v_1, v_2, \ldots, v_q)^T$ a weight vector belonging to the set $\mathscr{V} = \{\mathbf{v} \in [0,1]^q \colon \sum_{m=1}^q v_m = 1\}$. The quantile model average estimator (QMAE) of $G^c(x)$ is

$$\widehat{G}(u,\mathbf{v}) = \sum_{m=1}^{q} v_m \widehat{G}_m(u),$$

where $0 \le u \le 1$. Obviously, the QMAE is defined only for a candidate set of distributions with common support, which will be assumed for all of our tests and analysis.

A good choice of weights $\mathbf{v}$, as in PMAE, will make $\widehat{G}(u,\mathbf{v})$ close to $G^c(u)$ in a comprehensive way. Of course, $G^c(u)$ is unknown. However, based to the fact the empirical cumulative distribution function,

$$\bar{F}(x) = N^{-1} \sum_{i=1}^{N} I\{x_i \le x\},$$

is consistent for $F_c(x)$, its inverse $\bar{G}(u)$ is also a consistent for $G^c(u)$. By definition, the quantile function of $\bar{F}(x)$ is

$$\bar{G}(u) = \bar{F}^{-1}(u) = \inf\{x \colon \bar{F}(x) \ge u\} = (1-\gamma)x_{(\lfloor uN \rfloor)} + \gamma x_{(\lfloor uN \rfloor + 1)},$$

where $x_{(i)}$ is the i^{th} smallest x of the i.i.d. samples from F^c, and $\gamma = 1$ if $uN - \lfloor uN \rfloor > 0$, and $\gamma = 0$ otherwise. Therefore, $\bar{G}(u)$ is simply an order statistic of the observed values. Similar to the PMAE, cross-validation with $\bar{G}(u)$ is used as we describe explicitly in Section 3 below.

Although QMAE is a weighted average of quantile functions, the distribution parameters of these quantile functions are identical to those of the cdfs for PMAE, which are simply MLEs for distributions in the given candidate set $\mathscr{F}$. The only difference between PMAE and QMAE is caused by the CV-estimated weights. Therefore, given a good mixture weight $\widehat{\mathbf{v}}$ that minimizes the CV criterion, variate generation is identical to that of PMAE: each time a value of X is needed, generate integer $M \sim \widehat{v}_m$ to select the distribution, then generate $X \sim \widehat{F}_M$.

Remark: Although less familiar than mixture distributions, there has been previous work on quantile mixture models. Carole and Vanduffel (2015) derive an explicit expression for the quantiles of a mixture of two random variables as a function of the quantiles of the component quantile functions. The validity of the expression is shown through examining all possible cases of discrete and continuous variables with possibly unbounded support. Karvanen (2006) suggests the method of L-moments or sample Trimmed L-moments (TL-moments) to estimate the parameters of quantile mixtures, where L-moments are linear combinations of order statistics and TL-moments are generalizations of L-moments with increased conceptual sample size. Although this paper proposes certain parametric families of distributions whose parameters can be estimated by the method of L-moments (or TL-moments) with higher reliability than those estimated using conventional moment matching, this method is difficult to apply in many cases because it is impossible to derive closed-form L-estimators for many commonly used distributions. All of this work differs from ours in that we do not assume that the true distribution F^c is a quantile mixture of known families of distributions.

3 FITTING

The FMA approach uses CV to provide a good fit without overfitting. The J-fold cross-validation we use is related to the Jackknife model averaging (JMA) of Hansen and Racine (2012), which was originally proposed for improving the quality of estimators in a heteroscedastic linear regression model. The JMA estimator was shown to outperform other estimators in terms of smaller asymptotic expected squared error.

To apply the JMA-like scheme for input modeling in stochastic simulations, we randomly divide the real-world data set $x_1, x_2, \ldots, x_N$ into J groups such that each group has $S = \lfloor N/J \rfloor$ observations. For the jth group, the observations are labeled $x_{(j-1)S+1}, \ldots, x_{jS}$, where $j = 1, 2, \ldots, J$. Let $\tilde{G}_m^{(-j)}(u)$ be the maximum likelihood estimator of $G^c(u)$ for the mth candidate distribution with observations from the jth group removed from the data set. Notice that this is just the inverse function of the MLE for candidate cdf F_m using the same data. Therefore, the QMAE with the jth group removed is

$$\tilde{G}^{(-j)}(u, \mathbf{v}) = \sum_{m=1}^{q} v_m \tilde{G}_m^{(-j)}(u).$$

Denote the ith smallest observation in the jth group as $x_{(i)}^{(j)}$. The ED estimator of $G^c(u)$ using observations from the jth group only is denoted by $\bar{G}^{(j)}(u)$. Our J-fold CV criterion is

$$\begin{aligned} CV_J(\mathbf{v}) &= \sum_{j=1}^{J} \sum_{i=1}^{S} \left\{ \tilde{G}^{(-j)}\left(\frac{i}{S+1}, \mathbf{v}\right) - \bar{G}^{(j)}\left(\frac{i}{S+1}\right) \right\}^2 \\ &= \sum_{j=1}^{J} \sum_{i=1}^{S} \left\{ \tilde{G}^{(-j)}\left(\frac{i}{S+1}, \mathbf{v}\right) - x_{(i)}^{(j)} \right\}^2. \end{aligned}$$

In other words, we consider the sum of squared differences between the QMAE constructed without the jth group of real-world data, and the empirical quantile function constructed from only the jth group, across all groups. The optimal weight vector resulting from this criterion is

$$\widehat{\mathbf{v}} = \text{argmin}_{\mathbf{v} \in \mathscr{V}} CV_J(\mathbf{v}),$$

resulting in the quantile model averaging estimator $\widehat{G}(u,\widehat{\mathbf{v}})$ of $G^c(u)$. This contrasts with PMAE where the weight $\widehat{\mathbf{w}}$ minimizes

$$CV_J(\mathbf{w}) = \sum_{j=1}^{J}\sum_{s=1}^{S}\left\{\widetilde{F}^{(-j)}(x_{(j-1)S+s},\mathbf{w}) - \bar{F}^{(j)}(x_{(j-1)S+s})\right\}^2.$$

The optimization problem we need to solve to find $\widehat{\mathbf{v}}$ can be formulated as a quadratic program (QP). Specifically,

$$\begin{aligned}
\text{Minimize: } CV_J(\mathbf{v}) &= \sum_{j=1}^{J}\sum_{i=1}^{S}\left\{\tilde{G}^{(-j)}\left(\frac{i}{S+1},\mathbf{v}\right) - x_{(i)}^{(j)}\right\}^2 \\
&= \sum_{j=1}^{J}\sum_{i=1}^{S}\left\{\sum_{m=1}^{q} v_m\left[\tilde{G}_m^{(-j)}\left(\frac{i}{S+1}\right) - x_{(i)}^{(j)}\right]\right\}^2 \\
&= \sum_{j=1}^{J}\sum_{i=1}^{S}\left\{\sum_{m=1}^{q} v_m b_{mjs}\right\}^2 \qquad (2)\\
&= \sum_{j=1}^{J}\sum_{i=1}^{S}\mathbf{v}^T\mathbf{B}_{js}\mathbf{v} \\
&= \mathbf{v}^T\mathbf{B}\mathbf{v} \\
\text{subject to: } \sum_{m=1}^{q} v_m &= 1 \\
v_m &\geq 0,\ m = 1,2,\ldots,q.
\end{aligned}$$

where the matrices $\mathbf{B}_{js}$ and $\mathbf{B}$ are defined in the obvious way. If the $q \times q$ matrix $\mathbf{B}$ is positive-definite, then the objective function is strictly convex and the QP has a unique optimal solution (refer to Chapter 16 in Nocedal and Wright (2006)). Since it is obtained from the quadratic term in (2), it is clear that the matrix $\mathbf{B}_{js}$ is positive semi-definite, and therefore so is its sum $\mathbf{B}$; that it is positive-definite with probability 1 can be shown in a similar way to Nelson et al. (2018) for PMAE. PMAE also leads to a QP.

Notice that the dimension of the QP is only the number of candidate distributions, q, and it only needs to be solved once. It is hard to imagine the number of candidates ever being larger than $q = 40$, which is a modest QP. In practice we expect that reasoned choices for the candidate set will lead to $2 \leq q \leq 5$, making it a very small QP. The computational burden is in computing the MLEs for each candidate distribution from all of the data, and from the data with each of the J folds removed (thus, $q(J+1)$ MLEs in total), and construction of the matrix $\mathbf{B}$. Again, these are one-time calculations. The number of observations N and folds J only affect the set up, not the size of the QP.

4 R PACKAGE

In this section we describe the R package we created for FMA fitting—either PMAE or QMAE—and variate generation from the fitted distribution. The software may be downloaded from http://users.iems.northwestern.edu/~nelsonb/FMA/fma.R. This section is written in the form of standard R documentation. We illustrate use of the software in the following section.

4.1 Description

Creation of an input model via the frequentist model averaging (FMA) approach and random-variate generation for the fitted input model.

4.2 Usage

```
fmafit(X,Fset,J,type)
rfma(n,myfit)
```

4.3 Arguments

`X`	a numeric vector of nonzero length containing data values for fitting
`Fset`	a list of character strings that specifies the set of candidate distributions; supported distributions are `'normal'`, `'lognormal'`, `'exponential'`, `'gamma'`, `'weibull'`, `'inverse gaussian'`,`'student t'`, `'uniform'`, `'cauchy'`, `'loglogistic'`, `'ED'`, `'beta'`, `'logistic'`, `'pareto'`, `'rayleigh'`
`J`	number of groups to divide the data into for cross-validation; if not specified, `J = 10`
`type`	a character string specifying the type of model averaging estimator, `'P'` for probability, `'Q'` for quantile; if not specified, `type = 'P'`
`n`	number of random variates to generate
`myfit`	a list object returned by `fmafit` containing the four components needed for random-variate generation: `w`, `MLE_list`, `Fset` and `data`

4.4 Details

`fmafit` first fits each candidate parametric distribution in `Fset` to the data `X` using maximum likelihood estimation, which yields a set of fitted distributions $\widehat{\mathscr{F}} = \{\widehat{F}_1, \widehat{F}_2, \ldots, \widehat{F}_q\}$. The MLEs for each distribution are returned as `MLE_list`. Next a weight vector $\mathtt{w} = \{w_1, w_2, ..., w_q\}$ is obtained through cross-validation and also returned. The resulting model-average estimator of the true cumulative distribution of the data is

$$\widehat{F}(x, \mathbf{w}) = \sum_{m=1}^{q} w_m \widehat{F}_m(x).$$

The model average fitting can be either in probability space or quantile space. The difference between the two types of model averaging is only in the weight vector associated with the candidate distributions, which is obtained through cross-validation in either probability or quantile space.

`rfma` generates random variates that have the distribution of the model-average estimator. Each time a random variate is needed, a distribution is selected with probability equal to the corresponding weight and then a random variate from the fitted distribution is generated.

4.5 Values

`fmafit` returns an object called `myfit` which is a list with four components:

`w`	weight vector associated with distributions in `Fset`
`MLE_list`	list of MLEs for each candidate distribution with `'NA'` for `ED` (empirical distribution)
`Fset`	same as the input argument
`data`	same as input argument `X` (needed for ED)

`rfma` generates random variates from the distribution specified by `myfit`

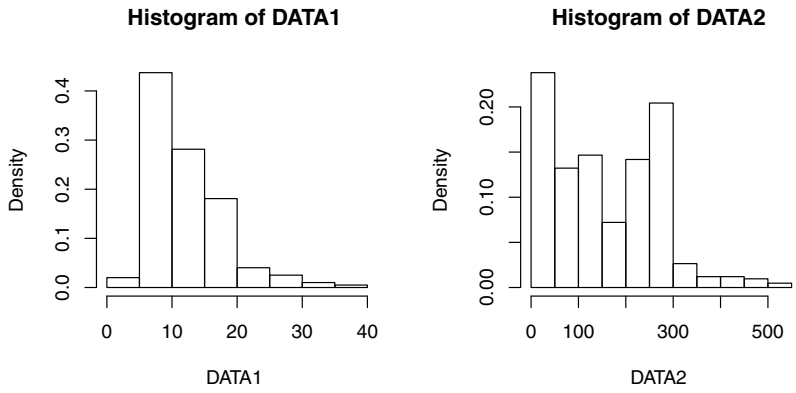

Figure 1: Histogram of financial returns (DATA1) and lot sizes (DATA2).

5 ILLUSTRATIONS

Nelson et al. (2018) provide a comprehensive empirical evaluation of PMAE by creating cases in which the true distribution F^c is known. Both the fidelity of the fit, and more importantly the fidelity of the simulation *output* with respect to the real-world system, tended to be greatly improved, especially when the tails of the input distributions matter. A queueing example, a highly reliable system and a stochastic activity network were considered, as well as a wide range of input distributions and candidate sets. Similar conclusions hold for QMAE.

In this section we illustrate PMAE, the new QMAE and our software on two real data sets, examining the fits that they provide and the protection afforded by including the ED in the candidate set. One data set is 200 financial returns, which we call DATA1. The second is 417 lot sizes of surface mount capacitors in a manufacturing simulation described in Wagner and Wilson (1996); we call this DATA2. The latter data set is bimodal and therefore is not well represented by the usual unimodal distribution choices. See Figure 1 for histograms of the two data sets. We used `J = 10` in both cases.

We first model DATA1 using `fmafit`. Of the distribution choices in `fmafit`, the best-fit distribution as measured by minimum AIC is `'gamma'`. Notice that `fmafit` can be employed to fit a single distribution, if desired, by only having a single choice in `Fset`. In Figures 2 and 3 we compare this single choice to a model average of `Fset=('gamma','weibull','lognormal')` for PMAE and QMAE, respectively. We also include the empirical cumulative distribution function (ECDF) in both plots for comparison. The R command for QMAE fitting is

```
myfit <- fmafit(DATA1, c('gamma','weibull','lognormal'), 10, 'Q')
```

The weight vectors obtained for PMAE and QMAE are $\mathbf{w}_P = (0, 0.0804, 0.9196)$ and $\mathbf{w}_Q = (0, 0.3627, 0.6373)$, respectively. From both figures it is clear that PMAE and QMAE are closer to the ECDF than the single best-fit gamma distribution, which indicates that the two FMA estimators better represent the distribution of DATA1. As illustrated in the figures and by the different weights, PMAE and QMAE lead to distinct fits. The Kolmogorov-Smirnov distance between each fit and the ED is 0.08 for the gamma, and 0.06 for both PMAE and QMAE, showing better conformance to the data for model averaging.

Interesting results are observed when fitting the bimodal lot size data, DATA2. The single best-fit distribution based on AIC is `'weibull'`. We compare it with the PMAE and QMAE when `Fset=('exponential', 'weibull','gamma','lognormal')`. The weight vectors associated with PMAE and QMAE are $\mathbf{w}_P = (0.4343, 0.5657, 0, 0)$ and $\mathbf{w}_Q = (0, 1, 0, 0)$, respectively. Figures 4 and 5 are plots of the CDFs of the resulting estimators. PMAE is better than the single best-fit distribution in the left tail of Figure 4 but worse in the right tail. QMAE, on the other hand, places all weight on `'weibull'`,

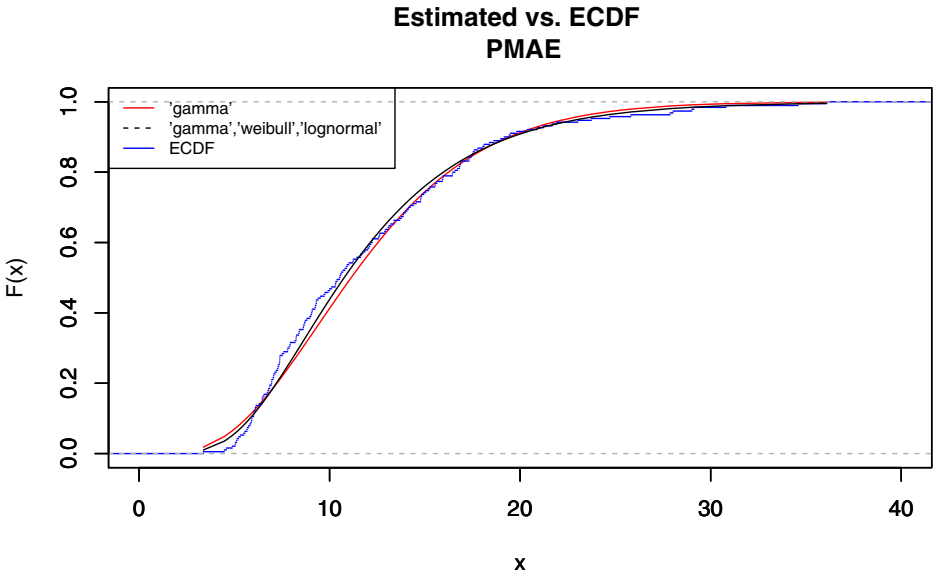

Figure 2: CDF of single best-fit distribution `'gamma'`, vs. PMAE with `Fset = ('gamma','weibull','lognormal')` for DATA1.

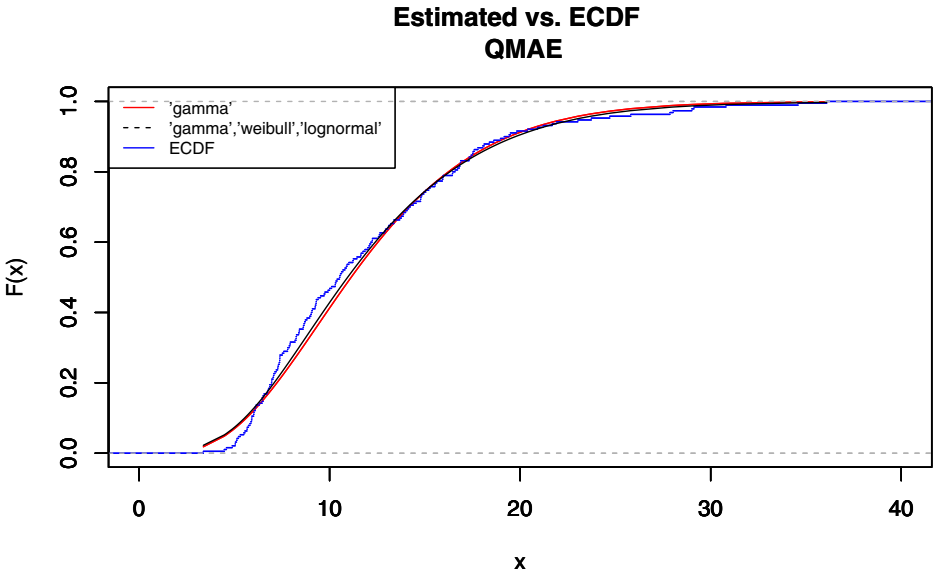

Figure 3: CDF of single best-fit distribution `'gamma'` vs. QMAE with `Fset = ('gamma','weibull','lognormal')` for DATA1.

which is exactly the same as the single-best distribution and better models the right tail. Thus, QMAE emphasizes the (long) right-tail behavior of the unknown input distribution more than PMAE does.

That said, neither fit is very good for this bimodal data when `Fset` includes only unimodal distributions. This is the common context when no standard distribution represents the data, even approximately, and including the ED as a candidate has significant value. To illustrate, we fit DATA2 with two candidate distribution sets, one including the ED (`Fset+ED`) and one without (`Fset`). The weight vector for the two PMAE fits are $\mathbf{w}_P = (0.4343, 0.5657, 0, 0)$ and $\mathbf{w}_P = (0.1596, 0, 0, 0, 0.8404)$, respectively. The corresponding weight vectors for the two QMAE fits are $\mathbf{w}_Q = (0, 1, 0, 0)$ and $\mathbf{w}_Q = (0.0180, 0.1583, 0, 0.0167, 0.8070)$. Both PMAE and QMAE place most, but not all, of the weight on the ED for `Fset+ED`. Thus, there is still some benefit of including the standard distributions. Figures 6 and 7 demonstrate the superior fit of `Fset+ED`, and thus the protection offered by including the ED in the candidate set of distributions. The Kolmogorov-Smirnov distance between each fit and the ED is 0.12 for the `'weibull'`, and both the PMAE and QMAE with `'exponential'`, `'weibull'`, `'gamma'` and `'lognormal'`; however, it is only 0.02 for both PMAE and QMAE when the `'ED'` is included, showing a substantially better fit.

6 CONCLUSIONS

In this paper we described two methods for "frequentist model averaging" that allow a simulation modeler to exploit the proven value of the standard families of distributions included in every simulation language (normal, lognormal, exponential, gamma, Weibull, etc.), while acknowledging that real-world input data will never perfectly conform to such distributions. Through model averaging we greatly extend the reach of these distributions, and by tuning the model average via cross-validation with the empirical distribution we insure that the fit is representative of the given real-world data. Including the empirical distribution as one of the candidates provides protection against data sets for which none of the standard distributions fit well.

Our R software `fmafit` makes fitting a model average distribution easy and fast. We recommend doing both probability and quantile fitting and comparing the results. The user may then take the weights and parameter estimates returned by `fmafit` and implement them as a simple mixture distribution in any simulation software, or use `rfma` to generate observations outside of the simulation model to read in as needed.

We recommend keeping the candidate set, `Fset`, small, remembering that the weights are *estimates* that will be noisier the more candidate distributions q there are. A set of $q \leq 5$ distributions including any with the right physical basis for the situation (e.g., Weibull for failures), that have good fit measures (e.g., AIC), plus the ED is our suggested approach. Our method also provides a way to judge when it is acceptable to use the ED alone: when the weight on the ED is close to 1. On the other hand, when this weight is far from 1, it indicates that the ED alone is insufficient. In any event, mixing smooths the ED in a way that is less arbitrary than, say, linear interpolation, and extends the ED's tails, which is often desirable.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation Grant Number CMMI-1634982. We thank Art Owen for suggesting the idea of fitting in the quantile space.

REFERENCES

Carole, B., and S. Vanduffel. 2015. "Quantile of a Mixture with Application to Model Risk Assessment". *Dependence Modeling* 3(1):172–181.

Hansen, B. E., and J. S. Racine. 2012. "Jackknife Model Averaging". *Journal of Econometrics* 167(1):38–46.

Karian, Z. A., and E. J. Dudewicz. 2000. *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. New York: CRC Press.

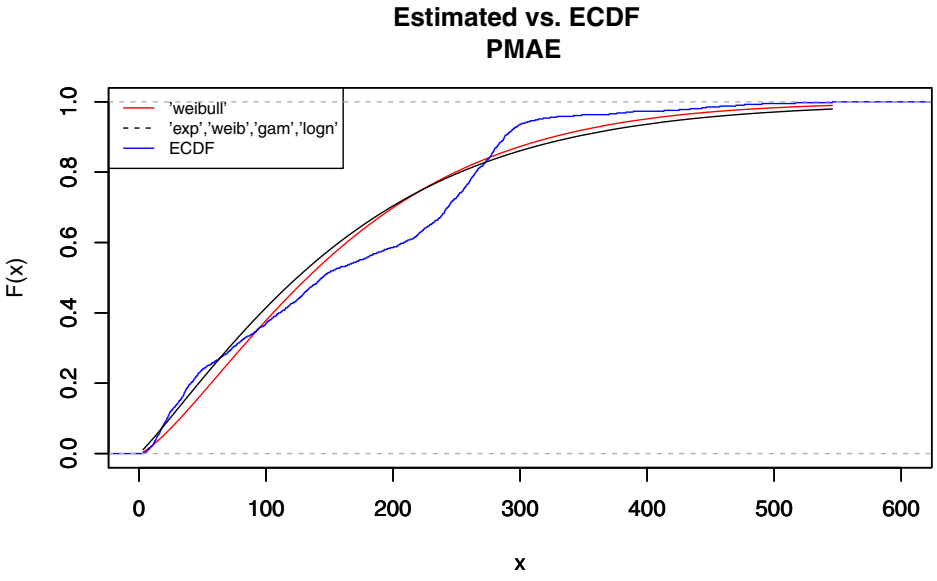

Figure 4: CDF of single best-fit distribution `'weibull'` vs. PMAE with `Fset = ('exponential','weibull','gamma','lognormal')` for DATA2.

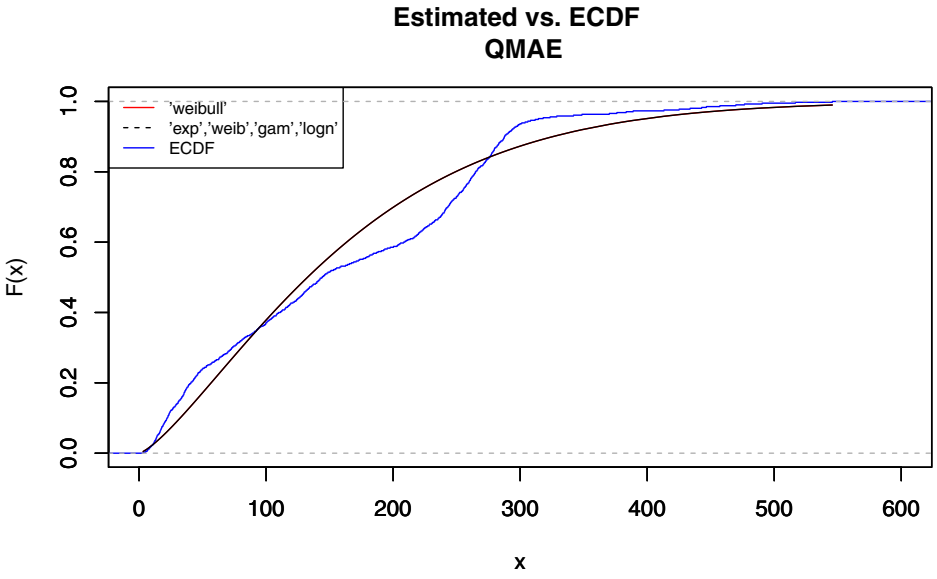

Figure 5: CDF of single best-fit distribution `'weibull'` vs. QMAE with `Fset = ('exponential','weibull','gamma','lognormal')` for DATA2.

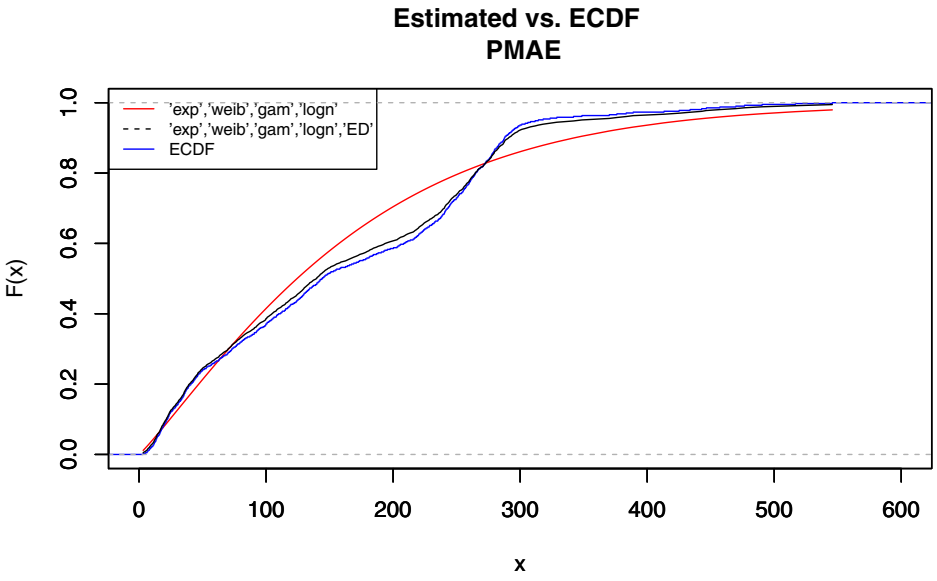

Figure 6: CDF of PMAE with `Fset = ('exponential','weibull','gamma','lognormal')` vs. PMAE with `Fset+ED` for DATA2.

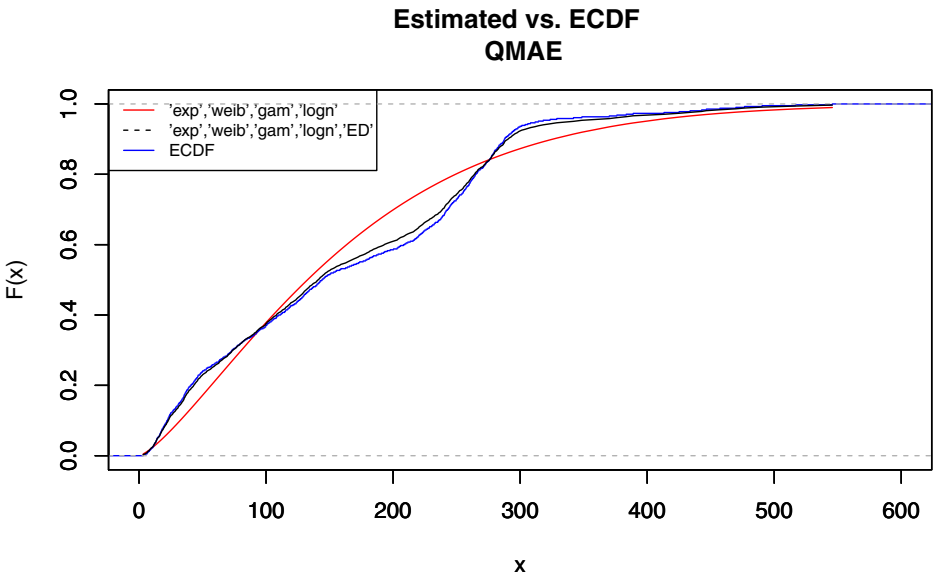

Figure 7: CDF of QMAE with `Fset = ('exponential','weibull','gamma','lognormal')` vs. QMAE with `Fset+ED` for DATA2.

Karvanen, J. 2006. "Estimation of Quantile Mixtures via L-moments and Trimmed L-moments". *Computational Statistics & Data Analysis* 51(2):947–959.

Law, A. M., and W. D. Kelton. 1991. *Simulation Modeling and Analysis*. 2 ed. New York: McGraw-Hill.

McLachlan, G., and D. Peel. 2004. *Finite Mixture Models*. New York: John Wiley & Sons.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer Science & Business Media.

Nelson, B. L., X. Jiang, A. Wan, and X. Zhang. 2018. "Reducing Simulation Input-Model Risk via Input Model Averaging". Technical report, Northwestern University, Evanston, IL USA. http://users.iems.northwestern.edu/~nelsonb/FMA.pdf.

Nocedal, J., and S. J. Wright. 2006. *Numerical Optimization*. 2 ed. New York: Springer.

Wagner, M. A. F., and J. R. Wilson. 1996. "Using Univariate Bézier Distributions to Model Simulation Input Processes". *IIE Transactions* 28(9):699–711.

AUTHOR BIOGRAPHIES

WENDY XI JIANG is a Ph.D. student in the Department of Industrial Engineering and Management Sciences at Northwestern University. She is majoring in applied statistics and statistical learning, with minors in analytics and optimization. Her research focus is on stochastic simulation methodology. Her e-mail address is xijiang2020@u.northwestern.edu.

BARRY L. NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IISE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address is nelsonb@northwestern.edu.