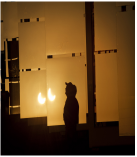


Exceptional service in the national interest

October 2, 2018

Conservative Estimation of Tail Probabilities from Limited Sample Data

Charles Jekel
Vicente Romero
V&V, UQ, and Simulation Credibility Processes Dept.
Sandia National Laboratories, Albuquerque, NM

Workshop: Risk Management Approaches in Engineering Applications University of Florida
October 1 - 2, 2018

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND NO. SAND2018-10998 C

Tail probability estimation with limited data

My work this summer:

- Estimating tail probabilities with limited data $2 \leq N \leq 20$
- Effective on exceedance probabilities on the order of $P < 10^{-2}$
- Challenging to provide conservative, but not overly conservative estimates
- Study performance of sparse-data UQ methods on 16 distributions
- Extending these methods to larger samples $N \approx 120$

Limited data methods

1. Tolerance Interval Equivalent Normal (TI-EN)
   - Construct a Normal distribution to represent sample
   - This is the same Normal distribution used for Tolerance Intervals
2. Superdistribution
   - Constructed from an ensemble of candidate Normal distributions that are consistent with the sparse samples of data
   - Data doesn't have to come from a Normal distribution to perform well in conservative tail probability estimation

Tolerance Interval Equivalent Normal (TI-EN)

Constructs an equivalent normal distribution from a sample

- This is the normal distribution used to construct Tolerance Intervals
- Control the conservatism with confidence level

$$\mathcal{N}(\tilde{\mu}, \sigma_{\mathsf{EN}}^2) \tag{1}$$

$$\sigma_{\mathsf{EN}} = k\tilde{\sigma} \tag{2}$$

- $\tilde{\mu}$ - sample mean
- σ_{EN} - equivalent normal standard deviation
- $\tilde{\sigma}$ - sample standard deviation
- k - correction factor based on confidence level

Superdistribution

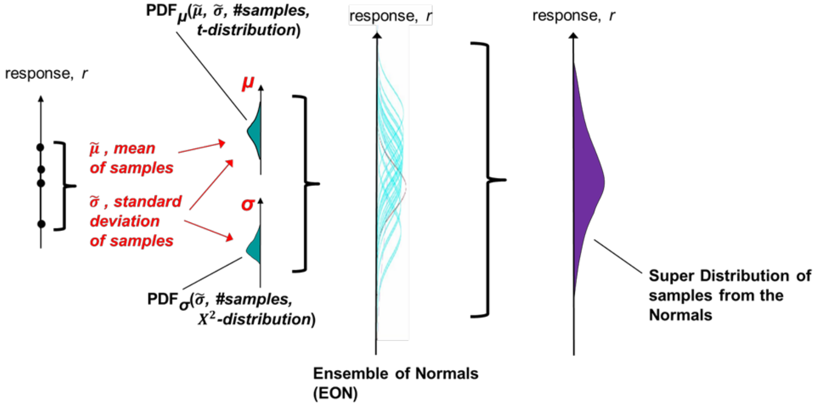


Figure: Construction of a Superdistribution from a sparse random sample of $N = 4$.

Superdistribution - exceedance probability

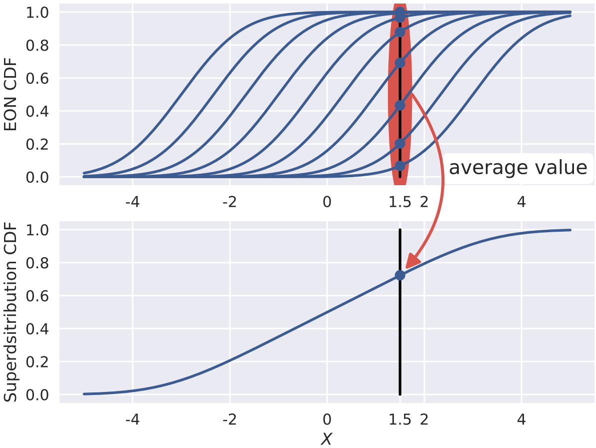


Figure: Predicted probability results from the average from each normal distribution in the Ensemble of Normals (EON).

Tail probability study

- 16 Distributions (8 analytical, 8 empirical)
- Generate 10,000 random sets of samples for each distribution
- Considering $N = 2, 3, 4, \cdots, 20$ number of samples
- Estimate the exceedance probability for each random sample
- Quantify the performance of the methods
  - Accuracy
  - Reliability

Analytical distributions

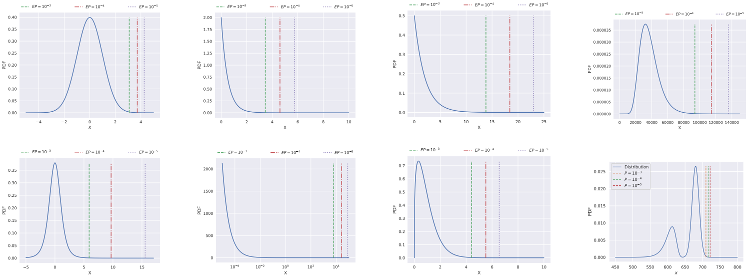


Figure: Analytical distributions and threshold locations for $EP = 10^{-3}, 10^{-4}, 10^{-5}$.

Empirical distributions

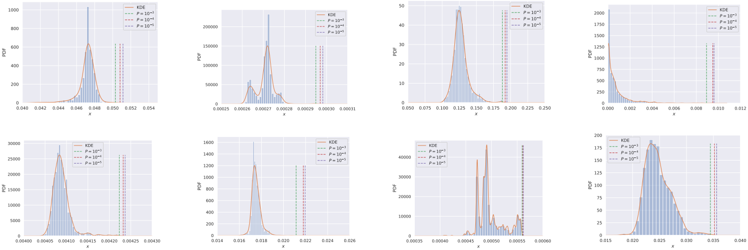
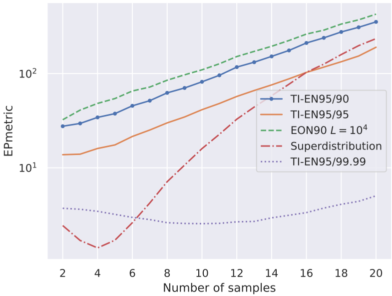
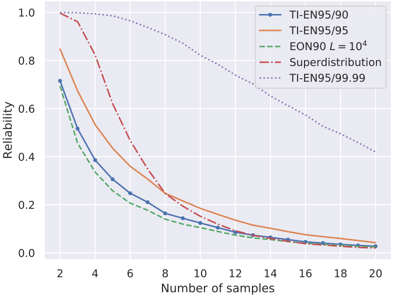


Figure: KDE fits to empirical data and threshold locations for $EP = 10^{-3}, 10^{-4}, 10^{-5}$.

Example result for $P = 10^{-4}$ on Exp Distribution.



Accuracy (lower is better)



Reliability (higher better)

Figure: Trade off between accuracy and reliability for low number of samples. The estimates become worse as the number of samples increase.

Results of the study

- Superdistribution (SD) was generally the most accurate method with conservative estimates
- Often the most accurate and conservative SD (optimal) was for small sample sizes between $N = 2$ - $N = 5$
- Results were distribution dependent and threshold dependent
- Smaller the exceedance probability the more conservative the estimates
- Reliability and accuracy often decrease as the number of samples increased
  - $N = 3, P \leq 10^{-3}$: 14 of 16 dists... reliability > 80%
  - $N = 4, P \leq 10^{-4}$: 14 of 16 dists... reliability > 80%
  - $N = 5, P \leq 10^{-5}$: 14 of 16 dists... reliability > 80%

% of conservative estimates for Superdistribution

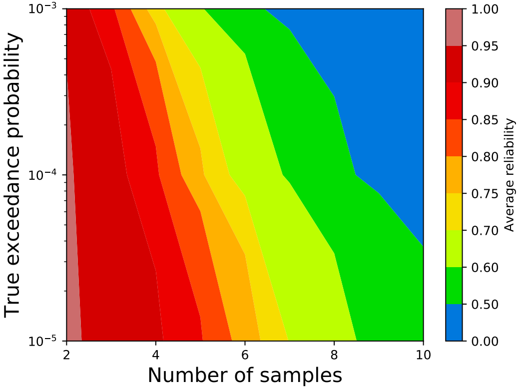


Figure: Reliability decays quickly as the number of samples increases, and less quickly with increasing order of magnitude of the exceedance probability.

Extending to larger sample sizes

The reliability and accuracy of the methods generally decreased as the number of samples increased.

What to do if you have $N \geq 4$ samples???

Statistical resampling

Reducing bias and variability in tail estimates

Bootstrapping:

- combinations with replacement
- each new sample has n number of points

Jackknifing:

- aggregates the $n-1$ sub-sample combinations
- predecessor to Bootstrapping

Generalized Jackknifing applied to limited data

- Consider all of the r sized sub-sample combinations
- Expressed as n choose r or nCr
- Compute the probability in each sub-sample
- Estimate results from the average of all estimated probabilities
- Total number of combinations:

$$\binom{n}{r} = \frac{n!}{k!(n-r)!} \tag{3}$$

Example on Exponential distribution

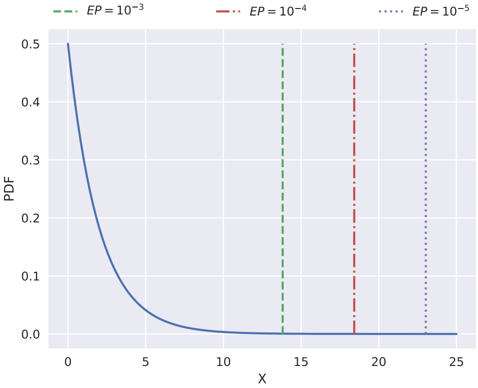


Figure: Exponential distribution and threshold locations for $EP = 10^{-3}, 10^{-4}, 10^{-5}$.

Limited data Jackknife results on a single distribution

Error on left (lower is better)
Reliability on right (higher is better)

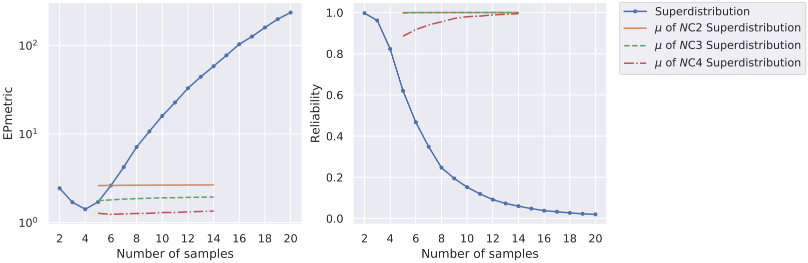


Figure: $nC4$ Jackknife Superdistribution method offers both improved accuracy and reliability for $N \geq 5$. This result is on an exponential distribution, and the results are distribution dependent.

Application with left tails on larger data $N \approx 120$

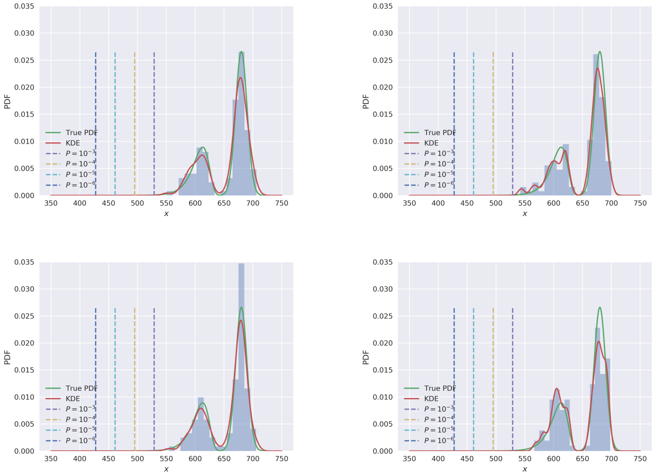


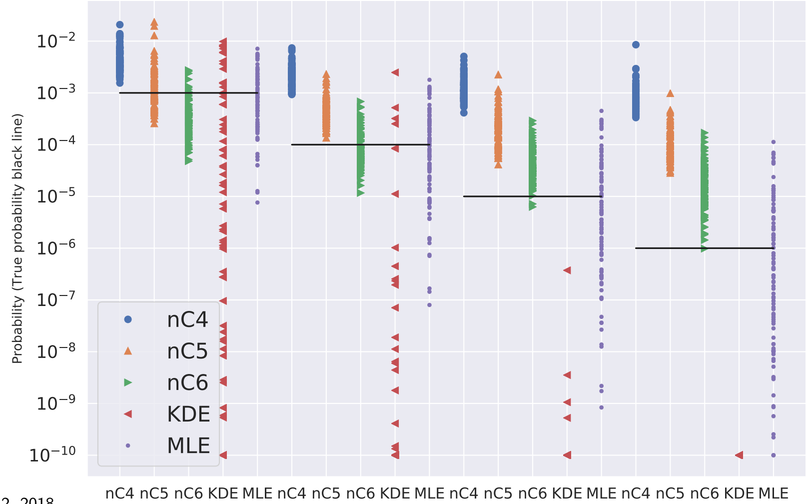
Figure: Examples of 4 random sets from the 100.

Methods to consider left tail probability estimation

100 sets of random samples, $N = 120$ samples per set

- Kernel Density Estimation (KDE)
  - Find optimal bandwidth that maximizes the likelihood
  - Cross validation grid search
- Maximum Likelihood Estimation (MLE)
  - Finds 5 parameters of the true distribution
  - Most cases true distribution unknown
  - Global optimization
- Limited data Jackknife technique with Superdistribution
  - Need to choose appropriate r sub sample size
  - Compare $nC4$, $nC5$, $nC6$ results

Each estimated tail probability result



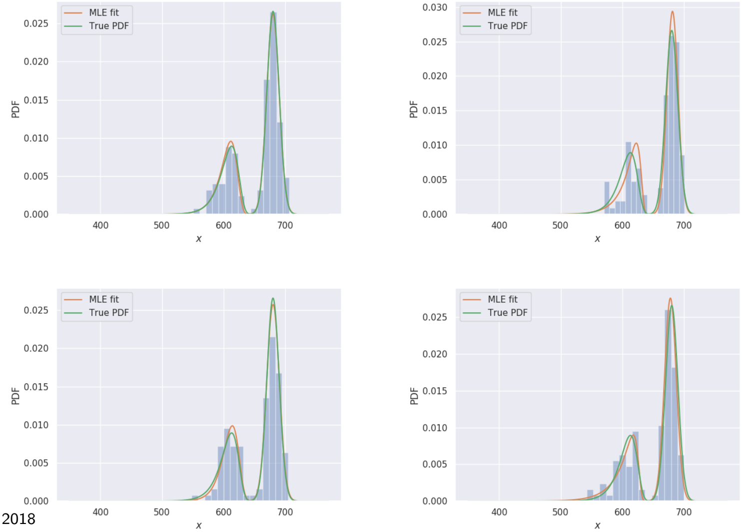
Summary

- Reasonably and reliably estimate extreme tail probabilities with limited samples
- The accuracy and reliability dependent upon: true distribution, level of exceedance probability, and the number of samples
- Control to how conservative the methods are
- Superdistribution (SD) was generally the most accurate method with conservative estimates
  - $N = 3, P \leq 10^{-3}$: 14 of 16 dists... reliability $>$ 80%
  - $N = 4, P \leq 10^{-4}$: 14 of 16 dists... reliability $>$ 80%
  - $N = 5, P \leq 10^{-5}$: 14 of 16 dists... reliability $>$ 80%
- Using SD with Jackknifing improved the conservative estimates for larger sample sizes beyond the optimal SD

Backup slides start here

:)

MLE Fit examples



Example of right tail thresholds

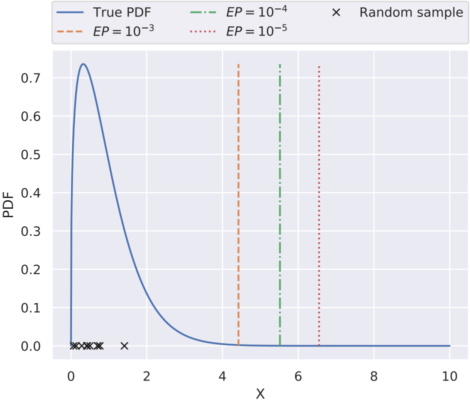


Figure: Thresholds of a Weibull distribution which have a right tail exceedance probability of $P = 10^{-3}, 10^{-4}, 10^{-5}$.

Reliability results

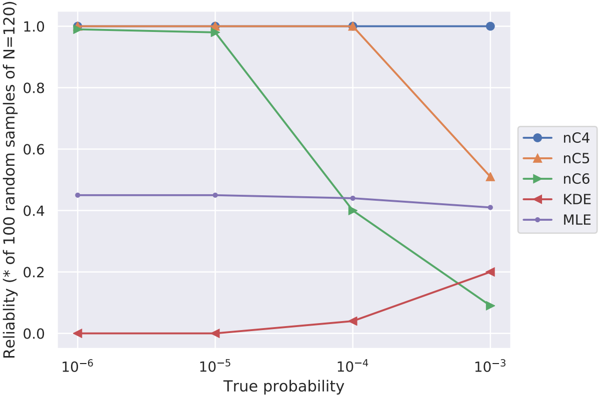


Figure: The *nCr* Jackknife Superdistribution method appears to be more reliable than KDE or MLE.