From data to probabilistic modeling in engineering risk analysis

Jianjun Qin, Assistant Professor, Shanghai Institute of Disaster Prevention and Relief, Tongji University, Shanghai, China
Kevin R. Mackie, Associate Professor, Civil, Environmental, and Construction Engineering Department, Univ. of Reliability, Orlando, USA
BozidarStojadinovic, Professor, Institute of Structural Engineering, ETH Zurich, Zurich,Switzerland

Abstract:

Probabilistic modeling is the basis of engineering risk analysis. For example, the Pacific Earthquake Engineering Research Center (PEER) framework to assess the risk of individual structure subject to seismic hazard disaggregates the problem into seismic intensity, engineering demand, damage, and decision intermediate probabilistic models. Quantification of seismic performance is based on the conditional probability distribution (fragility) functions used in the demand, damage and decision models. The probabilistic models in the risk analysis are always assumed to be parametric, and defined commonly as normal, lognormal, or some other distributions. The data used to estimate the distribution parameters are usually clustered in the small and moderate intensity ranges, forming the bulk of the data set, with few points associated with high intensities forming the high tail of the data set. Given that a single function is used to define the probabilistic model, the bulk of the data set, rather than its tail, govern the selection of the parameters of that function. However, the tails of the data set can be considerably different from the bulk, or the central part, of the data set (Caers and Maes, 1998). This difference (see Table 1) may significantly affect risk assessments based on the conditional probabilities formulated with the assumed parametric models.

Table 1. Characteristics of the bulk and the tails of the data set of a random variable.

	Data points	Target events	Consequences
Bulk	Large amounts	Small and moderate events	Low
Tails	Few	Extreme events	High

The present research work aims to improve the probability models used for engineering risk analysis by partitioning the bulk and the tails of the data set and by fitting different conditional probability distribution functions to represent these three data sub-sets (bulk and two tails). The data sets representing the parameter values are based on experimental data and/or numerical analysis. The probabilistic models in the risk analysis will be improved using tail power-law, extreme value, and other parametric probability distribution forms, while the bulk of the data set is modeled using a pre-defined distribution. The parameters of the power-law model are estimated using the approach by Clauset, et al. (2009). It will be explained the identification of the boundary between the bulk and the tail sub-sets, the

identification of the best distribution for the description of the tail sub-sets, the measurements, and the modification of the description of the distribution of the bulk. The PEER test-bed bridge seismic performance are re-evaluated using the proposed two-part demand and damage fragility models and compared to the results based only on lognormal fragilities presented in Mackie et al. (2008). This comparison will be used to examine the effectiveness of proposed strategy to improve the fragility models and the impact on risk when considering different hazard environments.

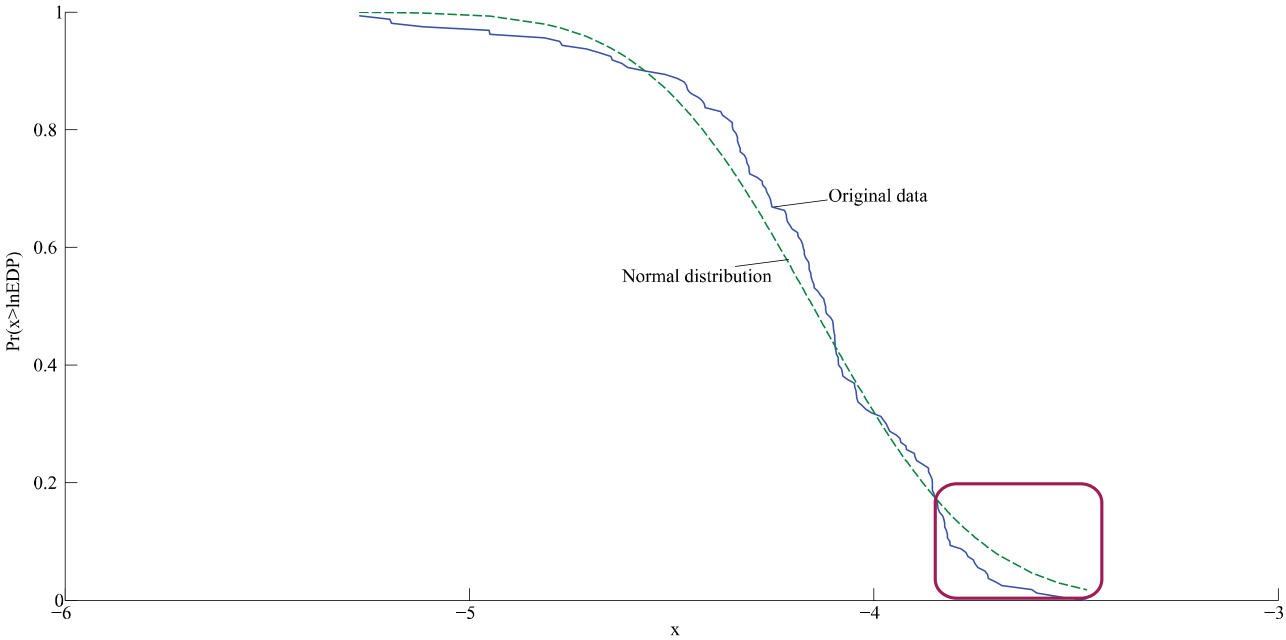

Figure 1. Comparison of the CCDF of numerically simulated variable data and a normal CCDF of the same variable

References:

Caers, J. & Maes, M. A. (1998), Identifying Tails, Bounds and End-Points of Random Variables, *Structural Safety*, **20**(1), 1-23.

Clauset, A., Shalizi, C. R. & Newman, M. E. J. (2009), Power-Law Distributions in Empirical Data, *SIAM Review*, **51**(4), 661-703.

Mackie, K. R., Wong, J.-M. & Stojadinovic, B. (2008), Integrated Probabilistic Performance-Based Evaluation of Benchmark Reinforced Concrete Bridges, PEER Center, University of California, Berkeley, pp. 199.