

Chapter 7

Uncertainty Analysis

7.1 Introduction

Economic consequences of natural, intentional, and accidental hazards include uncertainties. These uncertainties may arise due to variability in an event's magnitude, timing, duration, and location, as well as differing economic structures in various regions of interest. Quantification and propagation of these uncertainties result in probability distributions associated with various economic consequences. In this study, uncertainties associated with economic consequences are based on variability in stochastic regressors (predictor variables) within least squares and quantile regression models. Addressing uncertainties associated with regression model form (using linear predictor functions) was beyond the scope of this study.¹ Variability in stochastic regressors may arise due to inherent randomness (*aleatory* uncertainty) or incomplete knowledge (*epistemic* uncertainty) about underlying phenomena. *Epistemic* uncertainty may be reduced to *aleatory* uncertainty with more information, whereas *aleatory* uncertainty is not reducible. These consequence distributions, presented within a user-friendly and readily deployable tool, may be valuable for homeland security policy-makers conducting national risk assessments and for emergency management decision-making.

7.2 Overview

This chapter discusses the quantification, representation, propagation, and visualization of uncertainties in economic consequences within the E-CAT user interface. E-CAT displays inputs and outputs associated with hazardous events and their economic impacts with appropriate characterization of uncertainty. The economic

¹Regression parameter uncertainty will result in additional uncertainty associated with economic consequences.

consequences for each threat type are presented as probability distributions using input variables as: (1) point estimates, (2) mathematical intervals, and (3) triangular probability distributions. The uncertainty analysis is integrated with the CREATE Economic Consequence Analysis Framework (Rose 2009, 2015; Rose et al. 2014), which has expanded economic impact analysis to include resilience (actions to maintain system function and recover more rapidly), behavioral linkages (primarily fear), and remediation of consequences and spillover effects of countermeasures. Measures of uncertainty are aligned with various components of the framework and leverage prior work on quantifying uncertainties in direct hazard consequences (Chatterjee et al. 2015; Chatterjee et al. 2013a, b).

7.3 Uncertainty Quantification Tasks

The uncertainties in economic consequences may be characterized as statistical probability distributions using simulation methods. The research team implemented the following uncertainty quantification tasks:

- Monte Carlo sampling with variance reduction – This task involved Latin Hypercube sampling (Wyss and Jorgenson 1998), leading to more evenly distributed sample points across the sample space, to generate synthetic data associated with the E-CAT user interface input variables.
- Ordinary Least Squares regression (OLS) with stochastic regressors using synthetic data – This task produced estimates that approximate the conditional mean (given independent variables) of the dependent variable (i.e. economic consequences generated from CGE simulations).
- Quantile regression (QR) with stochastic regressors using synthetic data – This task produced estimates that approximate the conditional median (given independent variables) and other quantiles (i.e. 5, 25, 75, and 95 %) of the dependent variable. QR generates richer distributional data associated with the dependent variable and is more robust against outliers in the consequence estimates (Koenker and Bassett 1978; Koenker and Hallock 2001; Yu et al. 2003).

7.4 Uncertainty Representation

Uncertainties in quantitative models may emerge due to inherent randomness in samples or incomplete knowledge about fundamental phenomena (Paté-Cornell 1996). Representing these uncertainties appropriately is an important step for identifying knowns and unknowns among the modeling elements. Randomness may be addressed through the use of statistical probability distributions, whereas

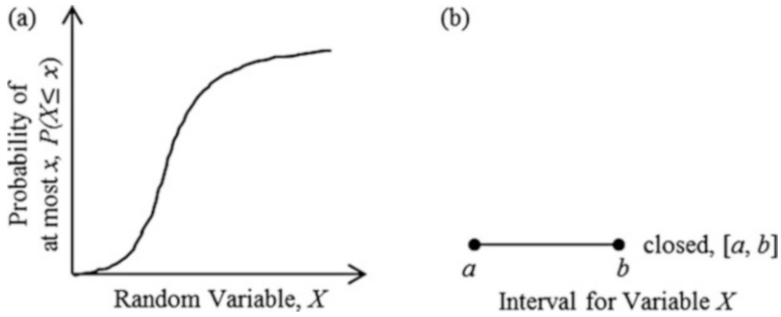


Fig. 7.1 Uncertainty representations for hypothetical variable, X . (a) Probability distribution. (b) Mathematical interval

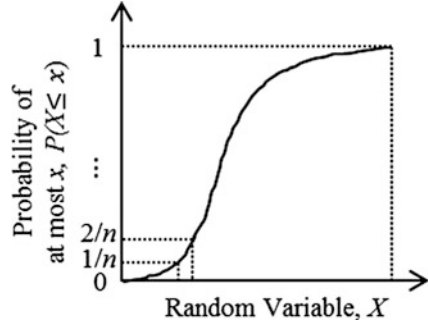
incomplete knowledge may be represented using mathematical intervals (Abrahamsson 2002).

Figure 7.1 presents two uncertainty representations (probability distribution and mathematical interval) for a hypothetical variable, X with uncertain values. Other uncertainty representations including probability bounds, probability boxes, and fuzzy sets are beyond the scope of this study. A probability distribution (see Fig. 7.1a) contains probabilities of occurrence of outcomes from a random experiment; and may be represented as a cumulative distribution function, $F(X) = P(X \leq x)$ that is a plot of probabilities of non-exceedance at various values (or estimates) associated with a random variable, X . Random variables with uncertain values may be *discrete* (with countable number of values; described using probability mass functions) or *continuous* (all values in a given interval; described using probability density functions). A mathematical interval (see Fig. 7.1b) is a set of real numbers between lower and upper bounds, $[a, b]$. The choice of uncertainty representation depends on data and knowledge associated with the variable of interest, i.e. economic consequences as GDP or employment losses in this study. Typically, with limited historical data for catastrophic events, probability distributions associated with reduced form model variables may be defined using a Bayesian approach (i.e. as degree of belief) with expert judgments.

7.5 Uncertainty Propagation

Approaches for propagating uncertainty to the output variables (i.e. GDP or employment losses) using reduced form regression models depend on the representations associated with the uncertain input variables. Let us assume \mathbf{x} representing a vector of m uncertain input variables; a single input variable is denoted as X ; and the regression model output y is a function of \mathbf{x} : $y = g(\mathbf{x})$. In this study, the function $g(\mathbf{x})$ represents the OLS and QR models that generate output y as

Fig. 7.2 Pictorial representation of Latin Hypercube sampling



conditional mean or quantiles (given independent variables \mathbf{x}) respectively. A Monte Carlo sampling approach is adopted in this study and is outlined below (for detailed discussion on additional approaches refer to: Abrahamsson 2002 and Cox 2012).

Let us assume an input random variable, X that has a cumulative distribution function $F(X) = P(X \leq x)$ and an inverse cumulative distribution function $F^{-1}(p) = x$. If $F(X)$ is strictly increasing and continuous, then $F^{-1}(p)$, where $p \in [0, 1]$, is a real number x such that $F(x) = p$. To generate a random sample value for an input random variable, X , a random number, r , is first generated between 0 and 1 (there are several random sampling schemes available in the literature (Abrahamsson 2002) including Latin hypercube sampling (a stratified sampling scheme without replacement—adopted in this study and presented in Fig. 7.2)). In the Latin Hypercube approach, $F(X)$ is segmented into n equally spaced intervals, where n represents the number of sampling iterations and a sample is drawn from each of these intervals. This sampled value, r , is then passed through the inverse cumulative distribution function $F^{-1}(r)$ to generate a random sample value, x . Similarly, random sample values for all m uncertain input variables may be generated resulting in a random sample vector, \mathbf{x} . The vector \mathbf{x} when passed through the function $g(\mathbf{x})$ produces a random output value of y . This Monte Carlo sampling process may be repeated several times to generate an empirical (simulation data-driven) probability distribution for the output random variable, Y . In this study, a Latin Hypercube sampling technique is adopted to sample from triangular probability distributions (with parameters as the minimum, most likely or mode, and maximum values) associated with the input random variables. Selecting values at equal intervals between the minimum and maximum values does not take into account the probabilistic structure associated with the input random variables. Also, this may not result in samples that are drawn from the overall distributional spread.

Often times, an analyst may require summarizing the distribution of the output variable, Y using mathematical expectation, $E[Y]$. With the discrete

variable assumption: $E[Y] = \sum_{i=1}^{\infty} y_i \cdot p_i$; and with the continuous variable assumption, $E[Y] = \int_{-\infty}^{\infty} yf(y)dy$ where $f(y)$ is the probability density function. Also, various quantile values, $Q(p)$ may be computed as $\inf\{y \in \mathbb{R} : F(y) \geq p\}$ to identify the minimum value of y that results in $F(y) \geq p$. In this study, expected means and quantiles are computed using empirical consequence distributions under the discrete assumption.

For the case with interval representation of input variables, lower and upper bound values are passed through the reduced form regression models (both OLS and QR) to generate lower and upper bound estimates for the output variables.

7.6 Uncertainty Visualization

Uncertainty analysis outputs may be visualized in various forms, given user-specified inputs as point estimates, intervals, or triangular probability distributions (represented using minimum, most likely, and maximum estimate values of a , c , and b respectively—see Fig. 7.3). Triangular distributions were chosen due to the relative ease in eliciting expert judgments for distribution parameters a , c , and b . Figure 7.3a displays a notional probability density function and Fig. 7.3b presents a notional cumulative distribution function for a random variable, X with triangular probability distribution.

The following discussion includes numerical examples to demonstrate various uncertainty visualizations based on notional input estimates. Loss variable in the charts below refers to an economic loss output type, e.g., GDP or employment loss.

- *Input Variables as Point Estimates* – Figure 7.4 presents an empirical distribution function using the QR results. This chart provides probabilities of not

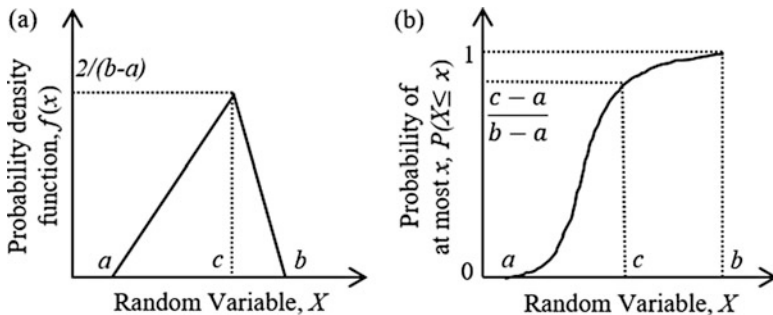


Fig. 7.3 Notional triangular probability density and cumulative distribution functions. (a) Triangular probability density function. (b) Triangular cumulative distribution function

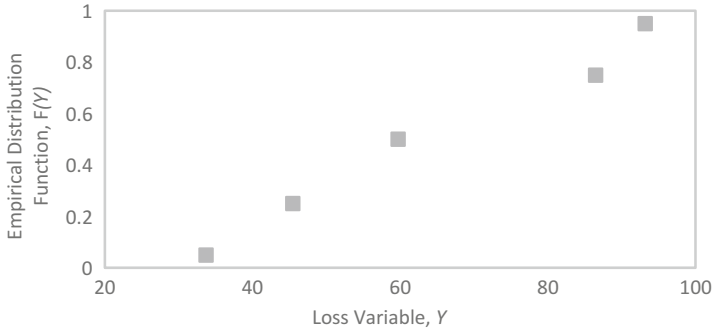


Fig. 7.4 Notional empirical distribution function

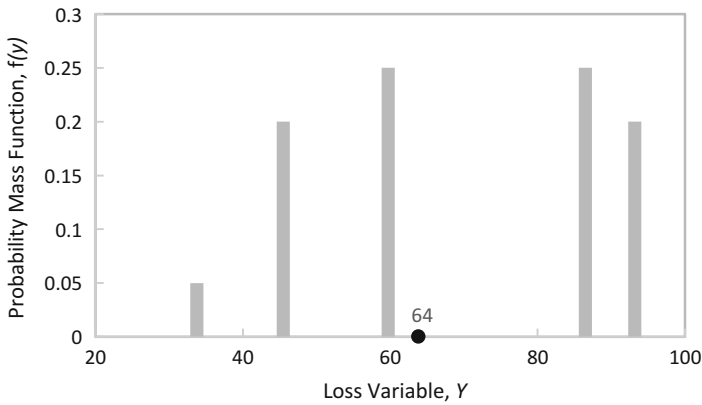


Fig. 7.5 Notional truncated probability mass function

exceeding certain levels of loss. For example, with probability of 0.5, losses will not exceed 59.74 units. Figure 7.5 presents a truncated probability mass function using the QR results and assuming economic loss as a discrete random variable. The bars in the plot represent probabilities of various levels of losses. For example, with probability of 0.05, losses will be 33.74 units. The mean loss is represented as a point value (at $y = 64$) from the OLS results. Figure 7.6 presents a box and whisker plot representing variability in the loss variable at different quantiles (5, 25, 50, 75, and 95 %) and the mean. We assume that the minimum and maximum losses correspond to the 5 and 95 % quantile losses. For example, with probability of 0.75, losses will not exceed 86.47 units.

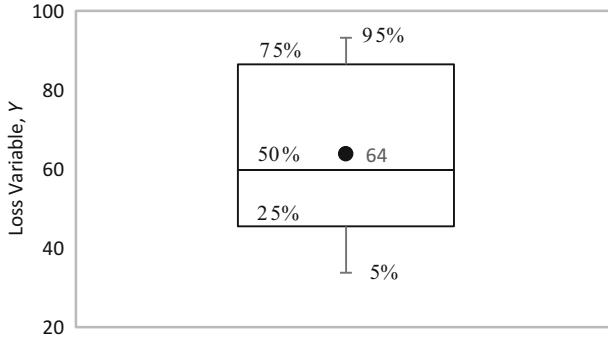


Fig. 7.6 Notional box and whisker plot

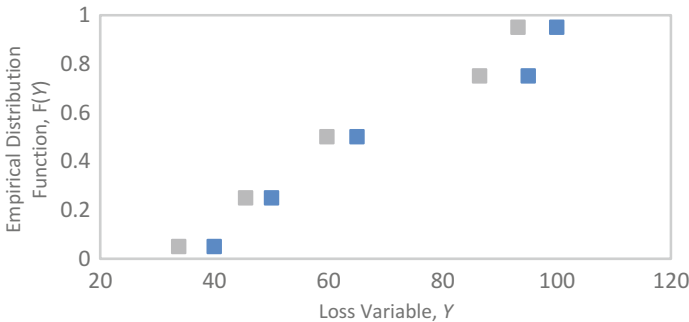


Fig. 7.7 Notional empirical distribution function with bounds (Note: lower bounds are in gray and upper bounds are in blue)

- Input Variables as Mathematical Intervals* – Figure 7.7 presents bounds for empirical distribution functions using the QR results. This chart provides probabilities of not exceeding certain bounded levels of loss. For example, with probability of 0.5, losses will not exceed a level between [59.74, 65] units. Figure 7.8 presents truncated probability mass functions for lower and upper bounds of economic losses using the QR results. The underlying assumption here is that the lower and upper bounds of economic losses are discrete random variables (In Fig. 7.5, lower bounds are in gray and upper bounds are in blue). The bars in the plot represent probabilities of various levels of losses. For example, with probability of 0.05, losses will be between [33.74, 40] units. The bounds on the mean loss (i.e. [64, 75]) are represented as point values from the OLS results. Figure 7.9 presents box and whisker plots, at the lower and upper bounds, representing variability in the loss variable at different quantiles (5, 25, 50, 75, and 95 %) and the mean. For example, with probability of 0.75, losses will not exceed a level between [86.47, 95] units.

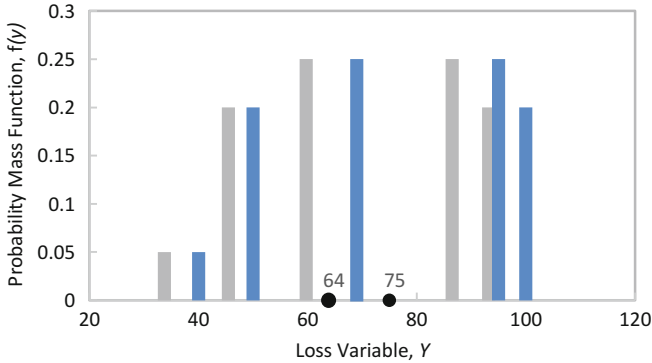


Fig. 7.8 Notional truncated probability mass function with bounds (Note: lower bounds are in gray and upper bounds are in blue)

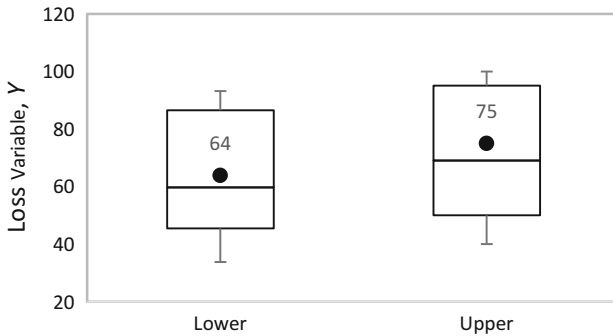


Fig. 7.9 Notional box and whisker plot with bounds

- Input Variables as Triangular Probability Distributions* – Figure 7.10 presents empirical cumulative distribution functions (ECDF) for the mean value, 5, and 95 % quantiles of an economic loss variable, based on empirical measures from the OLS and QR results. Lower to higher quantile distributions are presented as we navigate from left to right in the figure. These curves provide cumulative probabilities of non-exceedance at different levels of loss. The expected magnitudes of mean and quantile losses are estimated by evaluating the area above these curves. Figure 7.11 presents a relative frequency distribution for the mean value of an economic loss variable. A relative frequency distribution is a summary of the frequency proportions in a group of non-overlapping data

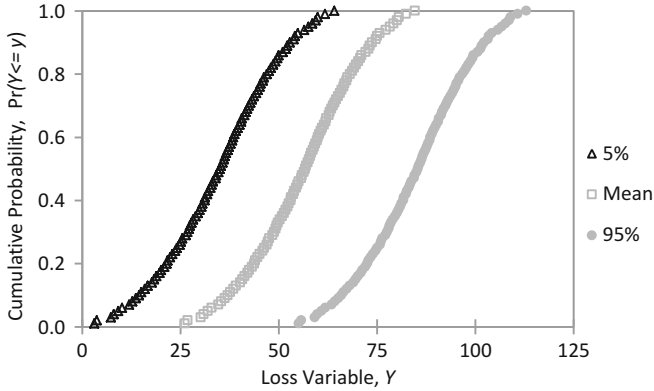


Fig. 7.10 Notional empirical cumulative distribution functions for mean and quantiles of the loss variable

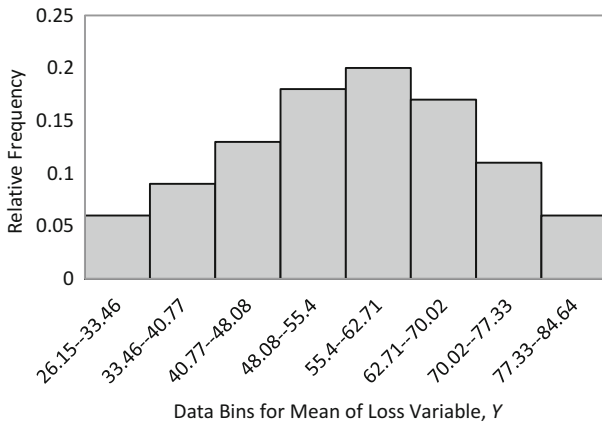


Fig. 7.11 Notional relative frequency distribution for mean of the loss variable

bins. Similar relative frequency plots were generated at other quantiles using the QR results.

As an example, based on the triangular probability distribution assumption, cumulative probability distributions at various quantiles and relative frequency plots for economic losses due to aviation system disruption are presented in Fig. 7.12.

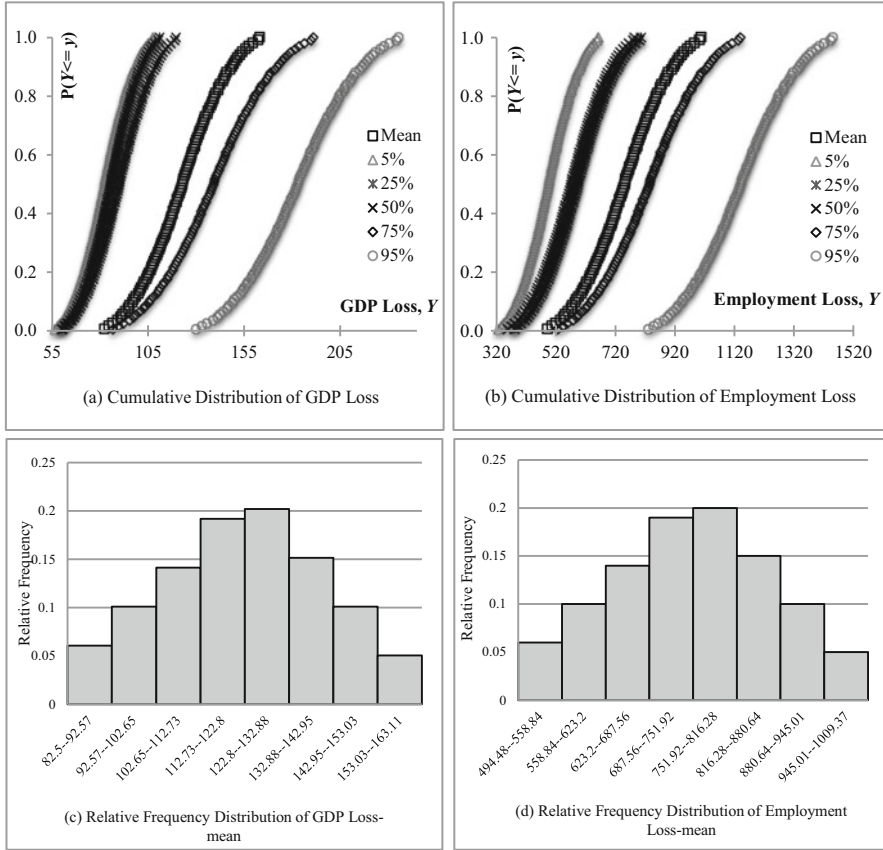


Fig. 7.12 Probability distributions of economic losses due to aviation system disruption

References

Abrahamsson M (2002) Uncertainty in quantitative risk analysis – characterization and methods of treatment, Report 1024. Department of Fire Safety Engineering, Lund University, Lund, p. 88

Chatterjee S, Salazar D E, Hora S C (2013a) Frequency-severity relationships for human – caused extreme events. Presentation at Society for Risk Analysis (SRA) Annual Meeting, Baltimore, Maryland

Chatterjee S, Salazar D E, Hora S C (2013b) Analyzing security portfolios amidst uncertain effectiveness. Presentation at Institute for Operations Research and the Management Sciences (INFORMS) Annual Meeting, Minneapolis, Minnesota

Chatterjee S, Hora SC, Rosoff H (2015) Portfolio analysis of layered security measures. *Risk Anal* 35(3):459–475

Cox LA (2012) Confronting deep uncertainties in risk analysis. *Risk Anal* 32(10):1607–1629

Koenker R, Bassett G Jr (1978) Regression quantiles. *Econometrica* 46(1):33–50

Koenker R, Hallock KF (2001) Quantile regression. *J Econ Perspect* 15(4):143–156

- Paté-Cornell ME (1996) Uncertainties in risk analysis: six levels of treatment. *Reliab Eng Sys Saf* 54:95–111
- Rose A (2009) A framework for analyzing and estimating the total economic impacts of a terrorist attack and natural disaster. *J Homeland Secur Emerg Manag* 6(1): Article 6
- Rose A (2015) Macroeconomic consequences of terrorist attacks: estimation for the analysis of policies and rules. In: Mansfield C, Smith VK (eds) *Benefit transfer for the analysis of DHS policies and rules*. Edward Elgar, Cheltenham
- Rose A, Avetisyan M, Chatterjee S (2014) A framework for analyzing the economic tradeoffs between urban commerce and security. *Risk Anal* 14(8):1554–1579
- Wyss GD, Jorgenson KH (1998) A user's guide to LHS: Sandia's Latin hypercube sampling software. Report Number SAND98–0210. Available at: <http://prod.sandia.gov/techlib/access-control.cgi/1998/980210.pdf>
- Yu K, Lu Z, Stander J (2003) Quantile regression: applications and current research areas. *Statistician* 52(3):331–350