The Use of Analytical-Statistical Simulation Approach in Operational Risk Analysis

Rustam Islamov[1]
Alexey Olkov[2]

Presented at
2010 Enterprise Risk Management Symposium
Society of Actuaries
April 12-15, 2010

[1] International Nuclear Safety Center, Moscow, Russia; islamov@insc.ru.
[2] The Agency for Housing Mortgage Lending, Moscow, Russia; asolkov@ahml.ru.

Abstract

Quantitative operational risk assessment is essentially based on stochastic scenario modeling of operational loss sequences. Given the lack of reliable historical data in most cases, mathematical methods should meet even stronger requirements in terms of results received and include specific features, namely: rare event analysis, uncertainty analysis and human factor analysis, etc. In this paper, the analytical-statistical simulation approach (ASSA) is considered as the most flexible approach to stochastic scenario loss sequence modeling and compared with Markov chain, fault/event tree and Monte Carlo, in terms of scenario loss sequence model adequacy and calculation cost. Probabilistic risk analysis software PRAISE and some applications are given.

1. Introduction

A series of corporate scandals and downfalls in the last two decades showed how operational risk could be a long-term killer even for the most respected companies all over the globe. As noted in Böcker and Klüppelberg (2005), probably the most dramatic and well-documented example of a bank collapse caused by operational risk losses was Britain's oldest merchant bank Barings after the rogue trader Nick Leeson had been hiding loss-making positions in financial derivatives. Other well-known lessons the financial industry was taught by are: Société Générale loss of €4.9 billion due to trader fraud: foreign currency trader staff fraud at National Australia Bank ($A360 million loss); currency trader fraud at Allfirst bank, then part of AIB Group (a loss near US $691 million); and Bear Stearns' near death since it was not able to price its mortgage portfolios. These examples obviously show the increased importance of effective and reliable operational risk management, based on risk identification, monitoring and reporting, risk mitigation, risk controlling and risk quantification. It is evident that most of these catastrophic losses would have never been prevented just by estimation of a loss distribution on historical data or naïve loss distribution approach (LDA) simulation techniques as they were due to flaws in internal control frameworks and ineffective internal audit functions. The first step in risk mitigation is effective management and internal control processes.

As for the relevance of operational risk modeling, the simplest argument is regulatory requirements. The most popular framework of Basel II, being de facto the standard of risk management in financial institutions, provides the argument for quantification of operational risk for every financial institution. In this respect, the main intention for the so-called advanced measurement approaches (AMAs) is to calculate a capital charge as a buffer against potential operational risk losses. Another reason for building models, besides to make predictions, is that models can help us to gain a deeper understanding of a subject matter as that is noted in Böcker and Klüppelberg (2009).

In Giacomelli and Pelizzon (2009), three classes of models for operational risk assessment and quantification are discussed. The most popular method, known as the LDA, is the parametric estimation of a frequency and a severity distribution for individual loss types and their subsequent aggregation that may incorporate dependencies. This general approach also includes the modeling of operational risk via extreme value theory (EVT), common Poisson shock models, and also some models based on ruin theory.

The second wide class of operational risk models also employs statistical techniques to quantify operational risk but uses mainly qualitative measures to calibrate the model. Key features of these models are qualitative-based scenario analyses and scorecards. Expert judgment, risk and control self assessments also fall within this class.

The third class of models focuses on the functional modeling of operational loss sequences. Functional processes and dependencies of operational risk events are defined and modeled via interdependence of the individual processes. The proposed approach of quantitative scenario modeling of operational losses, based on causal event tree analysis, is time-dependent, thorough and a realistic representation of real-world business processes going on in organization.

The proposed approach, which falls within the third class, is particularly useful when the underlying distributions exhibit rare events and complex dependence structures that influence the tails, which produce significant challenges for Monte Carlo modeling.

This paper is organized as follows. The second section describes some popular logical probabilistic techniques for scenario modeling used in industry and the role event tree analysis plays among them. The third and forth sections describe mathematics that is used to quantify the models described in the second section. The fifth section illustrates advantages of analytical statistical simulation approach (ASSA) on a simple example. The sixth section concludes. The appendix includes terms and definitions used in the paper.

2. Approaches to Scenario Modeling

Apart from quantitative scenario modeling, most financial institutions employ qualitative assessment of hazards by expert discussions and polling. Extensive literature is available on the use of expert aggregation techniques and questionnaires. Refer to Peters and Hübner (2009) for details. This paper is focused on formal logical probabilistic modeling of operational losses, which are based on probabilistic risk assessments (PRAs) originating from the complex system reliability engineering field, especially nuclear-power engineering. The proposed model is based on reliability stochastic models of complex system and operational loss sequences that are characterized by:

- logic and time cause-consequence relationships;
- random rare events are present in accident sequence.

In the nuclear industry, the most widely used method for system reliability and safety analysis is logic-probabilistic (based on Boolean algebra) using cut set definitions and the Markov process. The discussion below is mainly based on Papushkin, Islamov and Volkov (1999).

Let's formulate the problem for analysis.
System state is usually denoted by binary structural function:

$$f(x1, x2,..., xn)$$

or n-dimensional vector

$$\{xi\}, i = 1,..., n ,$$

where *xi* - binary vector component corresponds to system element state:
xi = 0 is efficiency state, *xi* = 1 is failure state,
n denotes full number of elements in considered structural scheme,
f = 1 denotes efficiency system state,
f = 0 denotes failure system state (operational risk realization).

When random vector $x(\omega)$ takes the value in the sampling space (**Ω**, A, **P**), where **Ω** is event space, A is event σ-algebra, **P** is probabilistic measure, then the range of $f(x)$ forms the state set **E**x(phase space), where system behavior random process $\xi(t)$ takes its value at the moment of time $t \in [0,T]$.

The total number of states can be equal to N(**E**x) = 2*n*. The process can be modeled as discrete Markov process crossing from one space **E**x state into other one. However, when the number of elements *n* > 20, phase space increases N(**E**x) > 106 and as a rule the Fault Tree method (with cut set tool) is used.

If the statistical simulation is used, there is a problem of sample volume. Random values, which define accident events, belong to so-called rare events. Satisfied number of sample is usually obtained as 102*np*-1. The number of simulation for one element is equal to 109 for probability estimation *p* = 10-7.

Analytical methods are simple to use and calculation-efficient, but statistical simulation ones have no restrictions on used models and have correct justification.

In order to use advantages of both approaches, the analytical-statistical simulation approach (ASSA) has been developed.

3. Scenario Model Essentials

Once initial events have been identified and grouped, it is necessary to determine the system response on each initiating event group. The modeling of the system response results in the generation of operational loss (accident) sequences due to operational risk.

Accident sequences are modeled in the form of event trees that are the logic diagrams reflecting success/failure of barrier functions required to prevent an accident in each initial event. There are two types of random events:

Random State: This state is described only by probability point estimation for the corresponding event in the considered accident sequence. In different accident sequences, the probability frequency estimation of the same barrier could be different.

Random Process: This process consists of parameter or parameters, which are the random values. A distribution function of this process and a time interval where the process is defined are considered to be known. In this case, one can obtain mean value, variance, other moments and confidence interval. Data for each barrier analysis depend on the type of the event.

One can obtain data for random state or random process by three ways: from modeling fault trees for functional events; from historical or general database; or from expert estimations. When the event frequency estimation or point probability estimation for the whole barrier in the considered accident sequence is known from historical evidence, the first way takes place. When estimations for the whole barrier in the considered accident sequence are unknown from statistics and fault frequencies estimation of elements which form the barrier are known, the second way takes place. In this case, the barrier is considered as a system or a part of the system. Given the data for elements one can construct fault trees. The required event frequency estimation for the whole barrier is defined with the aid of these fault trees. A system description, system boundaries, a system scheme, an element description, dependencies analysis and fault trees are initial data for the analysis. The third way takes place when data for the first and the second ways are unavailable. Initial data for this way are general databases, which are statistically analyzed or taken from experts' estimations.

Final states in event trees are grouped in categories. The first category is the safe state of the system representing zero operational loss. Other categories can be defined by two ways. In the first case, an operational loss interval from normal value of loss to maximum value of loss is divided into some parts. Each resulted interval corresponds to the defined category. In PRA software results, the total risk and the risk for each category are presented, so one can group final states in such a way that some chosen ones conclude in a category.

For each final state in each accident sequence, the total operational loss is defined. It is modeled by a random variable with one of many available probabilistic distributions. The data for a particular scenario loss distribution parameters assessment are taken either from historical-based statistics or experts' estimations.

4. Scenario Model Quantification

For the system first failure (sequence first operational accident) modeling, the situation when system or accident sequence (AS) reaches **at least one** failure (final) state on the denoted time interval is considered.

Let $\xi(t)$ represent stochastic process on the probabilistic space (**Ω**, A, **P**), taking the value from the finite set ***Z*** (possible states of the system). The process ξ can be also represented by a sequence of pairs of random values $\zeta_n = (\xi_n , \tau_n)$, where ξ_n denotes system state at the moment of time τ_n when process changes its state,

$$0 < \tau_0 < \tau_1 < ... < \tau_n < ...$$

The set $\boldsymbol{Z}_0 \subseteq \boldsymbol{Z}$ designates all the states of interest and denotes final events (human errors, equipment failures, etc.).

Suppose that the families of the conditional distributions

$$F_n (X, t ;Y_0 , Y_1 , Y_2 ,..,Y_{n-1}) = \mathbf{P}\{ \xi_n =X, \tau_n \le t \mid \zeta_0 =Y_0 , \zeta_1 = Y_1 ,.., \zeta_{n-1} = Y_{n-1} \},$$
$$F_0 (X) = P\{\xi_0 =X\}, n = 1,2,...,\tag{1}$$

where $Y_i =(X_i ,s_i)$, $X,X_i \in \boldsymbol{Z}$, $0 = s_0 <s_1 <...s_n <...$ are given.

Consider time interval $[0,T]$ and introduce the following event:

$$A = \{ \exists\, n : \xi_n \in Z_0 , \tau_n \le T \}\tag{2}$$

It is usually a rare event, i.e., the probability of this event is very small. The problem is to estimate $P\{A\}$. The method is based on modeling of random moments of time **only inside** the considered time interval $[0,T]$.

Put

$$A_n = \{ \xi_0 ,..., \xi_{n-1} \notin \boldsymbol{Z}_0 , \xi_n \in Z , \tau_n \le T \},\tag{3}$$

then

$$P\{A\} = P\{\xi_0 \in Z_0\} + \sum_{n=1}^{\infty} P\{A_n\}\tag{4}$$

Let us estimate $P\{A_n\}$. We can write

$$P\{A_n\} = \sum_{X_0 \notin Z_0} P\{A_n \big| \xi_0 = X_0\} P\{\xi_0 = X_0\}\tag{5}$$

Recurrently using the total probability formula we have for

k=1,..., n-2

$$P\{A_n \mid \zeta_0 = Y_0, \dots, \zeta_k = Y_k\} =$$
$$\sum_{X_{k+1} \notin Z_0} \int_{s_k}^{T} P\{A_n \mid \zeta_0 = Y_0, \dots, \zeta_{k+1} = Y_{k+1}\} F_{k+1}(X_{k+1}, ds_{k+1}; Y_0, \dots, Y_k) \qquad (6)$$

And finally for *k=n-1* we can write

$$P\{A_n \mid \zeta_0 = Y_0, \dots, \zeta_{n-1} = Y_{n-1}\} = \sum_{X_n \in Z_0} \int_{s_{n-1}}^{T} F_n(X_n, ds_n; Y_0, \dots, Y_{n-1}) \qquad (7)$$

Consequent application of formulas (5) - (7) yields function $P\{W\}$ expressed by conditional distributions (1).

The $P\{A\}$ estimation using analytical-statistical simulation is as follows. Let's transform Equation (6) into a more appropriate form

$$\begin{aligned} &P\{A_n \mid \zeta_0 = Y_0, \dots, \zeta_k = Y_k\} = \sum_{X_{k+1} \notin Z_0} \int_{s_k}^{T} P\{A_n \mid \zeta_0 = Y_0, \dots, \zeta_{k+1} = Y_{k+1}\} \bullet \\ &[F_{k+1}(X_{k+1}, T; Y_0, \dots, Y_k) - F_{k+1}(X_{k+1}, s_k; Y_0, \dots, Y_k)] \bullet \\ &P\{ds_{k+1} \mid \zeta_0 = Y_0, \dots, \zeta_k = Y_k, \xi_{k+1} = X_{k+1}, s_k < s_{k+1} \le T\} \end{aligned} \qquad (8)$$

Each integral on the right-hand side of Equation (8) is estimated using the Monte Carlo method. At every step a random value *sk+1* is simulated with the condition that

$$\xi k+1 = Xk+1 \text{ , } sk < sk+1 \le T \quad \text{and} \quad \zeta 0 = Y0 \text{ ,..., } \zeta k = Yk \qquad (9)$$

The integrals on the last interval

$$\int_{s_{n-1}}^{T} F_n(X_n, ds_n; Y_0, \dots, Y_{n-1}) \qquad (10)$$

are estimated analytically.

The system final state has the following probability estimation property. The statistical estimation for *P{W}* under *m*-times simulation is derived by selecting (*P1,...,Pm*).

The value

$$h_m(W) = \frac{\sum_{i=1}^{m} P_i}{m} \qquad (11)$$

is point estimation for $P\{W\}$. It has asymptotically the Gaussian distribution with parameters

$$\begin{aligned} &\mathrm{E}hm\,(W) = P\{W\} \\ &\mathrm{D}hm\,(W) = P\{W\}(1\text{-}P\{W\})/\,m \end{aligned} \qquad (12)$$

The upper and lower confidence boundaries, $Pu\,(hm)$ and $Pl\,(hm)$ respectively, can be found by solving the following equation

$$P\left\{ h_m(W) - \frac{S}{m^{1/2}} t_{\alpha,m-1} < P(W) < h_m(W) + \frac{S}{m^{1/2}} t_{\alpha,m-1} \right\} = 1 - \alpha \qquad (13)$$

where,

$$S^2 = \frac{\sum_{i=1}^{m} \left[P_i - h_m(W) \right]^2}{m-1};$$

$t\ \alpha,m\text{-}1$ - is the correspondent α-quantile for the Student distribution with $m\text{-}1$ parameters,

α- is the confidence level.

Estimation for models (2)–(3) is similar to (13). One of the important estimation properties is that variance of these estimations is not greater than in standard simulation procedures.

The described mathematics of ASSA for event tree scenario model quantification is effectively implemented in Probabilistic Risk Assessment for Industrial Safety Evaluation software package (PRAISE code for short) developed by Xlerate Technologies LLC. This package handles event tree development and quantification easily with a user-friendly graphical user interface and combines graphical simplicity with powerful numerical procedures for importance, sensitivity and uncertainty analyses. The implementation details of these methods lie beyond the scope of this paper. Refer to both papers (Islamov, 1998) for details on methods used.

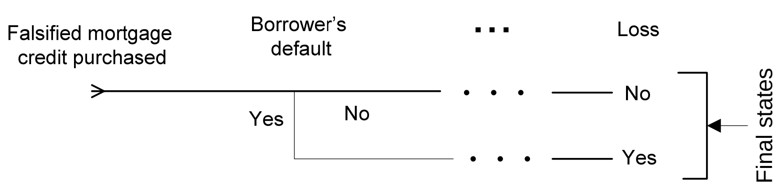
5. Advantages of ASSA

As mentioned in the previous paragraph, PRAISE/EventTree is focused on event tree representation of ASSA. PRAISE/EventTree, according to its documentation, has been developed to satisfy professional requirements in:

- analysis of time-dependant events;
- event random time simulation;
- event random probability function simulation;
- event deterministic time process modeling (mostly for application in engineering);
- point and interval probability and risk estimate;
- simulation speed optimization.

We'll illustrate the advantages of ASSA for quantitative operational risk analysis and modeling on a simplified scenario of two consecutive operational risk events.

Consider the simplest event tree representing a logic model of operational loss event sequence (accident sequence). The first considerable event, which is called Initial Event, may cause several possible paths. Each of other events may form the event tree branches by logic Yes/No:



Pic. 1

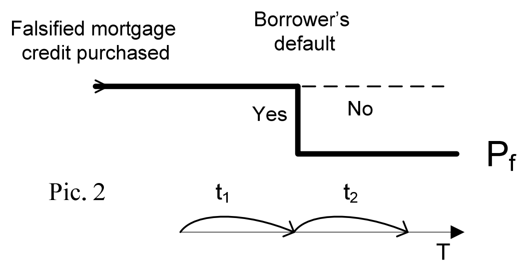
branch is determined by respective probability **P** of the event E:

$$P\{\mathrm{E = Yes}\} + P\{\mathrm{E = No}\} = 1$$

In the general case a final state probability is obtained as follows $P_f = P\left\{\bigcap_i E_i\right\}$. For independent events (in probabilistic sense) a final state probability can be given $P_f = \bigcap_i P\{E_i\}$, where event ***E*** relates to an appropriate logic function (Yes/No). The path final states with respective probability P_f and accident consequence C_f characterize the risk

$$R = \sum_f P_f \cdot C_f .$$

Consider a fragment from a simple model in Pic. 1. Shortcomings in the underwriting process cause the initiating event, an erroneous purchase of a fraudulent mortgage credit on a random time t_1, then a consequent event, failure to satisfy the terms of a loan obligation or a 90+ days payment delinquency (mortgage credit default) at random time t_2 follows. So we have an accident. Other functional events include the possible reselling to the originator of the loan in case of default, securitization of the loan, possible foreclosure and so on. For illustrative purposes and simplicity, we take only two events. So times t_1, t_2 are independent in a probabilistic sense and have distribution functions F_1, F_2 respectively. The mission time is T, which is the time interval risk analysis is conducted on.



Pic. 2

The accident final state probability is defined as $P_f = P\{t_1 + t_2 < T\}$

Let's compare standard Monte Carlo simulation approach with ASSA for a simple scenario of two consecutive events in Pic. 2.

Standard Monte Carlo Simulation

In standard simulation procedure we would have to:

1. simulate random values $0 < r < 1$ for each distribution function F_1, F_2;
2. estimate random times $t = F^{-1}(r)$ using inverse functions of F_1, F_2;
3. check the event $\{t_1+t_2<T\}$, whether it is successful or not;
4. repeat items1-3 N times to estimate the probability $P_f = m/N$, where m is a number of successful events.

It is obvious that, the lower probability $\mathbf{P_f}$ we estimate, the larger number of simulation N we have to carry out (so-called rare event problem):

$\mathbf{P_f}$	**m**	**M**
$\sim 2 \cdot 10^{-4}$	20	10^5
$\sim 2 \cdot 10^{-6}$	20	10^7

Analytical-Statistical Simulation Approach (ASSA)

Let's transform the expression:

$P_f = P\{t_1 + t_2 < T\}$:

$$P_f = P\{t_1 + t_2 < T\} =$$
$$P\{t_1 + t_2 < T \mid t_1 < T\} \cdot P\{t_1 < T\} =$$
$$\int_0^T P\{t_2 < T - u \mid t_1 < T\} dP\{t_1 < u \mid t_1 < T\} P\{t_1 < T\}.$$

The integral

$$\int_0^T P\{t_2 < T - u\} dP\{t_1 < u \mid t_1 < T\} P\{t_1 < T\} =$$
$$\int_0^T F_2\{T - u\} dP\{t_1 < u \mid t_1 < T\} P\{t_1 < T\}$$

may be considered as a mathematical expectation of the function $F_2(T - u)$, because

$$\int_0^T dP\{t_1 < u \mid t_1 < T\} = 1.$$

In ASSA simulation procedure we should:

1. estimate $F_1(T)$;
2. simulate random value $0 < r < 1$;
3. estimate random time $u = F_1^{-1}(r \cdot F_1(T))$ using inverse function of F_1;
4. estimate $F_2(T-u)$;
5. repeat items 2-4 m times to estimate the probability

$$P_f = \frac{\sum_{i=1}^{m} F_2(T - u_i)}{m}.$$

That is why ASSA has no "rare event" problem:

P_f	m
$\sim 2 \cdot 10^{-4}$	20
$\sim 2 \cdot 10^{-6}$	20

6. Conclusion

An important quantitative operational risk analysis benefits from the ability to estimate the precision of results received, which combines uncertainty analysis of the input data, the model uncertainty, sensitivity and importance analyses. Nowadays many of them are elaborated and well-documented steps described in probabilistic risk assessment procedure guides available.

Given powerful estimation methods like ASSA, one can fully use limited data available. These techniques are usually not used in simpler approaches, but nonrealization of a rare event is useful information, which helps to build one-sided interval estimations. Stochastic modeling itself provides instruments taking into account the uncertainty in observed or just roughly estimated frequencies and probabilities. That’s why quantitative scenario modeling, linked with uncertainty analysis, particularly helps fully use data available on frequency and probability estimations of rare operational risk events.

The ability to deal with analysis of time-dependant events (stochastic processes) makes the proposed model suitable for complex cases, including multi-event scenarios with complex dependencies.

The ability to deal with rare event problems, computational efficiency and flexibility of statistical simulation apparatus makes ASSA well-suited for nontrivial problems of reliability, engineering and quantitative operational risk analysis.

References

Böcker, K., and Klüppelberg, C. 2005. "Operational VAR: A Closed-Form Approximation." RISK Magazine December: 90–93.

Böcker, K., and Klüppelberg, C. 2009. "First Order Approximations to Operational Risk—Dependence and Consequences." In: Gregoriou, G.N., ed. *Operational Risk Toward Basel III*, *Best Practices and Issues in Modeling, Management and Regulation*. New York: Wiley, pp. 219–245.

Giacomelli, A., and Pelizzon, L. 2009. "Operational Risk Based on Complementary Loss Evaluations In: Gregoriou, G.N., ed. *Operational Risk Toward Basel III*, *Best Practices and Issues in Modeling, Management and Regulation*. New York: Wiley, pp. 69–84.

Peters, J-P., and Hübner, G. 2009. "Modeling Operational Risk Based on Multiple Expert's Opinions." In: Gregoriou, G.N., ed. *Operational Risk Toward Basel III*, *Best Practices and Issues in Modeling, Management and Regulation*. New York: Wiley, pp. 3–21.

Papushkin, V., Islamov, R., and Volkov, A. 1999. *Development of Standard Probabilistic Risk Assessment (PRA) Procedure Guide: System modeling*, *NSI PREDRAFT REPORT-1999*. Russian Academy of Science, Nuclear Safety Institute.

Islamov, R. 1998a. *Uncertainty Analysis*, IBRAE RAN, Report for NRC.

Islamov, R. 1998b. *Vulnerability Study*, SCIENTECH, Report for Sandia National Laboratory.

Appendix: Terms and Definitions

Throughout the paper some specialized terms and definitions are used originating from nuclear power safety engineering, operational risk and reliability analyses. Some of these terms and definitions are given below.

Initiating event—an event that creates a disturbance in the business process functionality that directly leads to operational loss depending on the successful operation of the various mitigating controls and systems in the business process.

Components—equipment, devices, operations and other elements designated to perform specific functions solely or as a part of the system and considered as a design structural element when performing reliability and safety analysis.

Failure—an event of disrupting operating conditions of a component, or whole system.

Event tree—a logical model that expresses in graphical form the different ways of operational loss sequence development for an initiating event group being considered that depends on accomplishing or failure of safety functions and successful or not successful personnel activities necessary for prevention of operational loss.

Fault tree—a logical model presenting the various failure combinations resulting to safety function non-performance.

Point Estimate (Parameter Point Estimate)—a single number that represents an estimate of a reliability parameter, such as the mean, median or mode developed through maximum likelihood (in some sense it gives the best estimate), moments method, Bayesian estimation methods or other methods. The Point Estimate does not contain any information on possible discrepancy of the parameter estimation.

System—a set of components designed for fulfilling specified functions.

Uncertainty—random variability in given parameters or measurable quantity or the imprecision in the knowledge of the parameter or model.