Stochastic Event-Based Loss Assessment using OpenQuake

Within the OpenQuake event-based calculators, seismicity is simulated by generating a *stochastic event set* (also known as *synthetic catalog*) for a given time span T . For each rupture generated by a source, the number of occurrences in a time span T is simulated by sampling the corresponding probability distribution as given by $P_{rup}(k|T)$. A stochastic event set is therefore a *sample* of the full population of ruptures as defined by a seismic source model. Each rupture is present zero, one or more times, depending on its probability. Symbolically, we can define a stochastic event set spanning a time period T , $SES(T)$ as:

$$SES(T) = \{k \times rup,\ k \sim P_{rup}(k|T)\ \forall\ rup \in src\ \forall\ src \in SSM\}$$

where k , the number of occurrences, is a random sample of $P_{rup}(k|T)$, and $k \times rup$ means that rupture rup is repeated k times in the simulated stochastic event set (for more details, see Pagani et al., 2014).

For the purposes of estimating the Average Annual Loss (AAL) or the loss values corresponding to a set of return periods, the OpenQuake-engine allows the risk analyst two options for running the Monte Carlo simulation of events:

1) Generate one SES spanning n years
2) Generate n SES spanning 1 year each

Assuming a time-independent Poissonian recurrence model, the above two strategies are statistically equivalent for estimating the aggregate AAL. However, only strategy (2) is viable for the estimation of loss values corresponding to a set of chosen return periods[1]. For the rest of the discussion below, we therefore use option (2) to generate the event sets.

For any risk metric estimated via sampling, we are interested in answering some questions concerning the accuracy of the estimate, such as the following:

1. What is the 95% confidence interval for the estimate of the risk metric?
2. What is the minimum number of SES that should be simulated in order to keep the width (or half-width) of the 95% confidence interval below a certain chosen value?

[1] This assumes that both the AAL and the loss values corresponding to a set of chosen return periods are calculated considering the aggregate losses in each simulated year (i.e., summing up the losses from all events that occur in each year).

Average Annual Loss

Each year can have zero, one, or multiple events causing losses to a portfolio of assets. Let the annual loss, AL, be defined as the sum of all losses that occur within a year. For n stochastic event sets, each spanning one year, the set of annual loss values calculated from this event set, $(AL_1, AL_2, \ldots, AL_n)$, forms a sequence of i.i.d. random variables, which can be treated as samples from an unknown probability distribution. Our goal is to compute the Average Annual Loss, which is given by the expected value of this unknown distribution. We can estimate the AAL using the sample mean of the sequences of calculated annual losses as the estimator:

$$\overline{AL}_n = \frac{1}{n} \sum_{i=1}^{n} AL_i$$

The accuracy of $\overline{AL}_n$ in estimating the true AAL can be measured by using the standard error:

$$SE_{\overline{AL}_n} = \frac{s_n}{\sqrt{n}} = \sqrt{\frac{1}{n \times (n-1)} \sum_{i=1}^{n} \left(AL_i - \overline{AL}_n\right)^2}$$

Then, by the Central Limit Theorem, the $100(1-\delta)\%$ confidence interval for the AAL estimate is:

$$\left[\overline{AL}_n - z \times \frac{s_n}{\sqrt{n}}, \overline{AL}_n + z \times \frac{s_n}{\sqrt{n}}\right]$$

where $z = 1.64$ for the 90% confidence interval, and $z = 1.96$ for the 95% confidence interval.

Thus, the “error bars” associated with the 95% confidence interval for our estimate of the AAL would be:

$$\overline{AL}_n \pm 1.96 \times \frac{s_n}{\sqrt{n}}$$

In practice, the number of years of losses that should be simulated in order to keep the final confidence interval half-width equal to or below a value $\varepsilon \times \overline{\mu}_{AL,n}$ can be estimated by first simulating a small set of trial runs spanning n_0 years. The final number of years of simulated events required, n, is given by:

$$n \approx z^2 s^2{}_{n0} / \varepsilon^2 \overline{AL}^2{}_{n0}$$

Consider the following illustrative examples based on a portfolio of 28,596 residential assets in the San Francisco Bay Area. An initial calculation using $10,000$ SES spanning 1 year each produces the following results:

$$\overline{AL}_{10,000} = 0.17\%$$
$$s_{10,000} = 1.03\%$$
$$SE_{\overline{AL}_{10,000}} = \frac{s_{10,000}}{\sqrt{10,000}} = 0.01\%$$

The “error bars” associated with the 90% confidence interval for our estimate of the AAL would be:

$$90\%CI \approx \overline{AL}_{10,000} \pm 1.64 \times SE_{\overline{AL}_{10,000}} = [0.153\%,\ 0.187\%]$$

The relative half-width of the 90% confidence interval is $\varepsilon = (0.187 - 0.153) \div (2 \times 0.170) = 9.9\%$

The “error bars” associated with the 95% confidence interval for our estimate of the AAL would be:

$$95\%CI \approx \overline{AL}_{10,000} \pm 1.94 \times SE_{\overline{AL}_{10,000}} = [0.150\%,\ 0.190\%]$$

The relative half-width of the 95% confidence interval is $\varepsilon = (0.19 - 0.15) \div (2 \times 0.17) = 11.8\%$

If we desire a 95% confidence interval with a half-width of 10% of the estimate for the AAL $(0.10 \times 0.17\% = 0.017\%)$, we would have to choose the number of SES, n based on the approximate formula:

$$n \approx z^2 s^2_{n0} / \varepsilon^2 \overline{AL}^2_{n0} = 1.96^2 \times 1.03^2 / 0.017^2 = 14,022$$

Repeating the calculations using $100,000$ SES and $1,000,000$ SES spanning 1 year each , we get:

AAL	10,000 *SES*	100,000 *SES*	1,000,000 *SES*
Estimate	0.170	0.180	0.185
90% CI	[0.153, 0.187]	[0.175, 0.187]	[0.183, 0.186]
95% CI	[0.150, 0.190]	[0.174, 0.188]	[0.182, 0.187]

All values in %

Loss Corresponding To A Given Return Period

The loss corresponding to given return period RP , l_{RP} , is defined as follows:

$l_{RP} \equiv$ smallest value of l for which $Prob(AL \leq l) = 1 - RP^{-1}$

Once again, suppose we simulate n stochastic event sets spanning one year each, and calculate the set of aggregate annual loss values, $(AL_1, AL_2, \ldots, AL_n)$. We can estimate l_{RP} using:

$$\hat{l}_{RP,n} = sup \left\{ l : \tfrac{1}{n} \sum_{i=1}^{n} I(AL_i \leq l) < 1 - RP^{-1} \right\}$$

Unlike the previous case involving the estimator for the AAL, there is no simple formula for the standard error for the estimator in this case, since this estimator cannot be expressed as the mean of a random variable. However, we can obtain a bootstrap estimate of the standard error, and bootstrap estimates for the confidence interval in this case. Suppose that our calculated annual loss values, $AL_1, AL_2, \ldots, AL_n$, are samples from an unknown probability distribution F_{AL} . Let $\hat{F}_{AL}$ indicate the empirical probability distribution based on our calculated annual loss values:

$\hat{F}_{AL}$: probability mass $1/n$ on $AL_1, AL_2, \ldots, AL_n$

Now, in order to determine an approximate confidence interval for our estimate of l_{RP} , the Bootstrap algorithm proceeds as follows:

1. Draw B bootstrap samples (with replacement) from the empirical distribution $\hat{F}_{AL}$.
   Say these samples are:

   $y^*(1) = (AL^*_{11}, AL^*_{12}, \ldots, AL^*_{1n})$,
   $y^*(2) = (AL^*_{21}, AL^*_{22}, \ldots, AL^*_{2n})$,
   $\vdots$
   $y^*(B) = (AL^*_{B1}, AL^*_{B2}, \ldots, AL^*_{Bn})$.

2. For each bootstrap sample $y^*(b) = (AL^*_{b1}, AL^*_{b2}, \ldots, AL^*_{bn})$, calculate the loss value corresponding to the return period $\hat{l}^*_{RP,n}(b)$.

3. Obtain the approximate confidence intervals using percentiles from the bootstrap cumulative distribution function, $\hat{G}_{AL}(l) \equiv Prob_*(\hat{l}^*_{RP,n} < l)$:
   $\left[\hat{G}_{AL}^{-1}(\delta/2), \hat{G}_{AL}^{-1}(1 - \delta/2)\right]$ gives the $100(1 - \delta)\%$ confidence interval
   Eg. $\left[\hat{G}_{AL}^{-1}(0.025), \hat{G}_{AL}^{-1}(0.975)\right]$ gives the 95% confidence interval

Efron and Tibshirani (1986) indicate that $B = 250$ is an approximate minimum for the number of Bootstrap samples required to compute the percentile based confidence intervals.

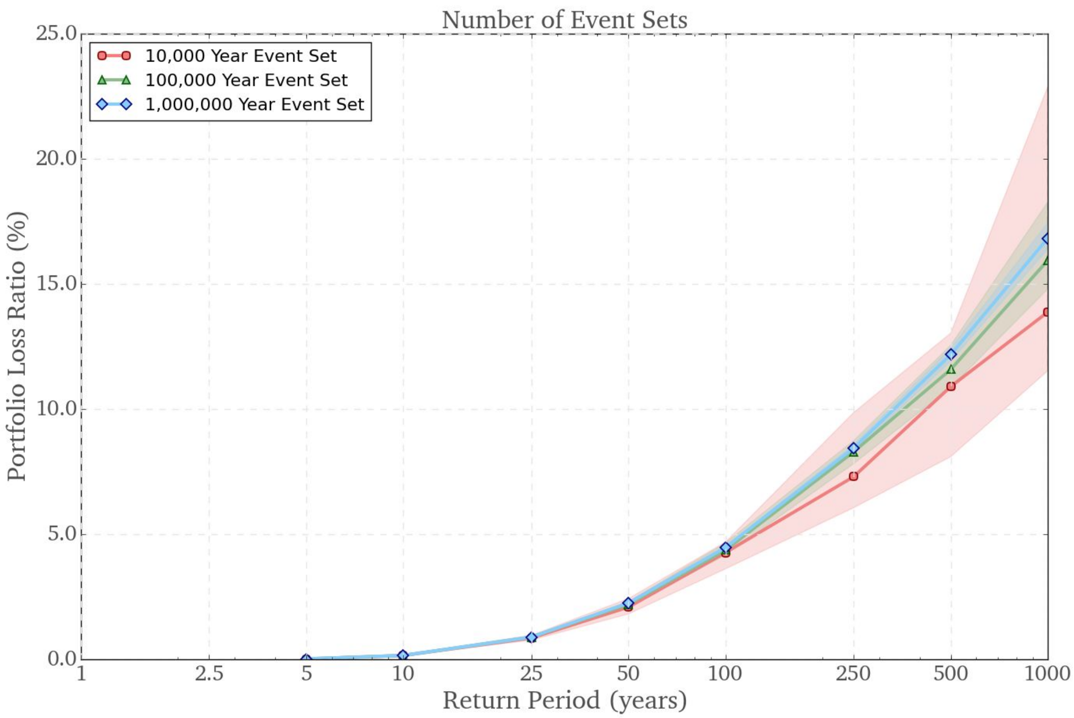
Since the width of the confidence interval in this case is not related to the standard error by a simple formula, we cannot easily estimate the number of years of losses that should be simulated in order to keep the final confidence interval width equal to or below a chosen value. Starting from a small number, one would have to progressively increase the simulated number of years until the calculated bootstrap confidence interval falls below the what the modeller is comfortable with. With the portfolio of residential assets in the San Francisco Bay Area, three calculations using $10,000\ SES$, $100,000\ SES$, and $1,000,000\ SES$ spanning 1 year each produce the following results ($B = 1,000$):

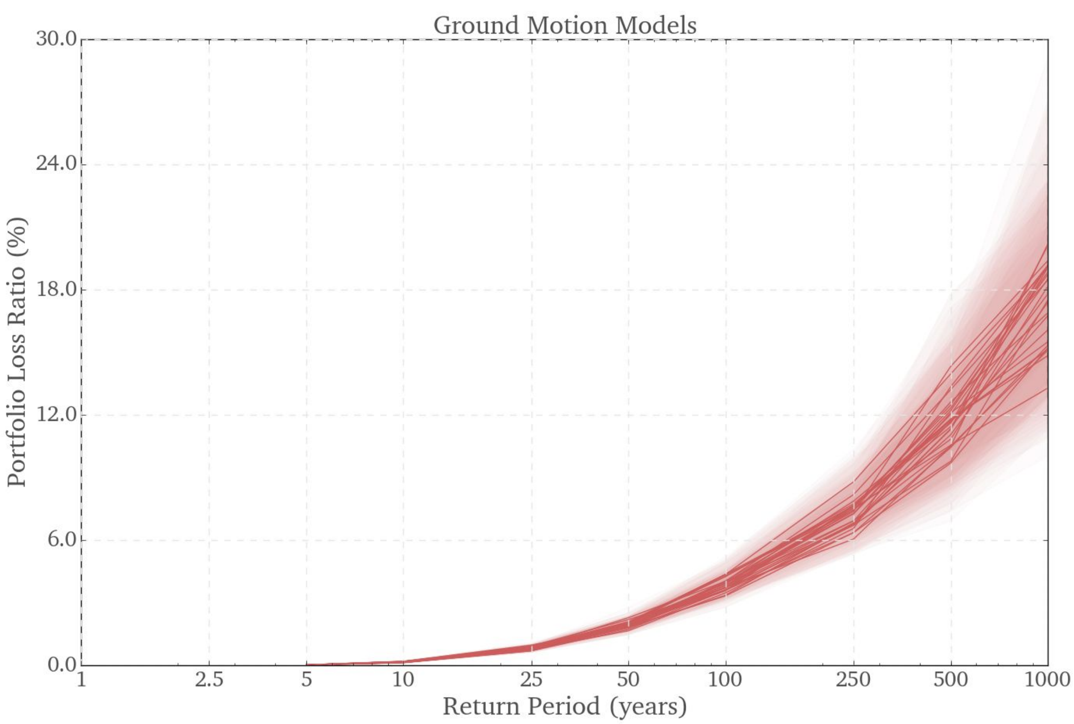
RP = 100 years	Statistic	10,000 *SES*	100,000 *SES*	1,000,000 *SES*
Point estimate	Estimate	4.264	4.363	4.466
Bootstrap estimates	Mean	4.221	4.384	4.467
	Median	4.264	4.361	4.467
	Stddev	0.285	0.130	0.037
	CoV	0.068	0.030	0.008
	90% CI	[3.652, 4.654]	[4.208, 4.630]	[4.406, 4.527]
	95% CI	[3.486, 4.684]	[4.166, 4.653]	[4.395, 4.544]

All values except CoV in %

RP = 250 years	Statistic	10,000 *SES*	100,000 *SES*	1,000,000 *SES*
Point estimate	Estimate	7.298	8.295	8.444
Bootstrap estimates	Mean	7.220	8.281	8.437
	Median	7.298	8.295	8.443
	Stddev	0.633	0.229	0.072
	CoV	0.088	0.028	0.008
	90% CI	[6.221, 8.097]	[7.913, 8.625]	[8.306, 8.555]
	95% CI	[6.092, 8.559]	[7.892, 8.741]	[8.272, 8.579]

RP = 1,000 years	Statistic	10,000 *SES*	100,000 *SES*	1,000,000 *SES*
Point estimate	Estimate	13.874	15.941	16.827
Bootstrap estimates	Mean	15.648	16.121	16.868
	Median	13.874	15.874	16.822
	Stddev	3.665	0.982	0.281
	CoV	0.234	0.061	0.017
	90% CI	[12.614, 22.749]	[15.052, 18.156]	[16.452, 17.331]
	95% CI	[11.709, 22.872]	[14.872, 18.271]	[16.396, 17.436]


Number of Event Sets
10,000 Year Event Set
100,000 Year Event Set
1,000,000 Year Event Set
Portfolio Loss Ratio (%)
Return Period (years)
25.0
20.0
15.0
10.0
5.0
0.0
1
2.5
5
10
25
50
100
250
500
1000


Ground Motion Models
Portfolio Loss Ratio (%)
Return Period (years)
30.0
24.0
18.0
12.0
6.0
0.0
1
2.5
5
10
25
50
100
250
500
1000

References

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–75. Retrieved from http://www.jstor.org/stable/2245500

Pagani, M., Monelli, D., Weatherill, G. A. and Garcia, J. (2014). The OpenQuake-engine Book: Hazard. *Global Earthquake Model (GEM) Technical Report 2014-08, doi: 10.13117/-GEM.OPENQUAKE.TR2014.08, 67 pages.*