Optimizing the Quality of Risk Assessments of Biological Epidemics

Joachim Draeger (draeger@iabg.de)
IABG mbH, Ottobrunn, Germany

1 Motivation

The assessment of the risk R is a standard task in the safety and security domain. Formally, risk is defined as $R = \sum_i p_i L_i$ based on the frequency p_i and the losses L_i of unfortunate outcomes i. This means, that the risk is equal to the the expected loss in the sense of probability theory. As soon as the risk of safety- and/or security-critical systems is considered, the validity of the calculated risk value becomes an important aspect.

2 The SIR Epidemics Model

For the case study, we will consider risk assessments of biological epidemics. Such a situation is represented by a differential equation system

$$\begin{aligned} S' &= -\beta IS/N \\ I' &= \beta IS/N - \gamma I - \mu I \\ R' &= \gamma I \\ D' &= \mu I \end{aligned} \tag{1}$$

The designations S, I, R are a reminder on the the three infection states of the SIR model [1] representing *S*usceptible, *I*nfected, and *R*ecovered persons. The variable D designates the *D*ead. The overall population is designated as $N(t) := S(t) + I(t) + R(t)$. The parameters occurring in the equation system (1) are as follows:

Parameter with Description		Parameter Range
μ	Death rate	$\mu \in [0, 1]$
β	Infection rate	$\beta \geq 0$
γ	Recovery rate	$\gamma \in [0, 1]$

In (1), the population $N(t)$ is decreasing monotonically due to deaths of infected persons until the epidemics fades out. Accordingly, we assess the outcome based on the deaths cumulated over time. We define $L := \lim_{t\to\infty} D(t)$. An outcome is provided by simulating a specific scenario given by the corresponding model parameters (and initial values). Simulation means in this case the application of a numerical differential equation solver. If a parameter value is not known, we may assume a probability distribution of possible parameter values. In such a case, a Monte Carlo approach provides the corresponding outcome statistics. Accordingly, the risk R is determined by summing up the deaths L_i occurring in a scenario i weighted by the probabilities p_i of i.

3 Approximation Errors

Monte-Carlo sampling and simulation are algorithmic tools, which facilitate a computational determination of the risk R. They replace non-computable, mathematically exact terms by computable (finite) approximation processes. This leads immediately to corresponding approximation errors. We have to take the following types of errors into account.

Limitation	Corresponding Error
Finite simulation horizon t_H	Simulation stop error ε_H
Finite step size Δ of the simulation	Discretization error ε_Δ
Finite number n of Monte-Carlo samples	Stochastic sampling error ε_n

The overall result error ε_R — i.e. the difference between approximatively calculated and mathematically exact risk R — depends stochastically on ε_n and the outcome errors ε_y of the included scenario samples. ε_y, in turn, depends on ε_H and ε_Δ. These errors are not independent from each other; the discretization error ε_Δ, for example, influences the simulation stop error ε_H. Clearly, the more effort is invested — i.e. the later the approximation series are cut off — the smaller will be the errors ε_H, ε_Δ, ε_n and thus the result error ε_R as well. The number $\mathcal{N}$ of evaluations of (1) can be considered as the natural effort measure. It influences ε_R depending on the investment scheme $\mathcal{N} = \mathcal{N}_H \cdot \mathcal{N}_\Delta \cdot \mathcal{N}_n$, whereby $\mathcal{N}_H$, $\mathcal{N}_\Delta$, $\mathcal{N}_n$ determines the t_H, Δ, n, respectively. This means that ε_R will depend on the choice of the assigned fractions $\mathcal{N}_H$, $\mathcal{N}_\Delta$, $\mathcal{N}_n$. Our primary interest is an analysis, how ε_R changes corresponding to the chosen investment scheme $(\mathcal{N}_H, \mathcal{N}_\Delta, \mathcal{N}_n)$, and which investment scheme gives the smallest ε_R.

4 Recommended Analysis Strategy

The following questions are only suggestions. Feel free to skip questions, which are turning out to be intractable, or to add discussions, which seem to be of interest.

1. Dependence of ε_R on the chosen investment scheme ($\mathcal{N}_H$, $\mathcal{N}_\Delta$, $\mathcal{N}_n$) for a given $\mathcal{N} = \mathcal{N}_H \cdot \mathcal{N}_\Delta \cdot \mathcal{N}_n$?
2. Which investment scheme gives the smallest result error ε_R?
3. Dependence of the 'optimal' investment scheme on the model parameters?
4. What are the trade-offs e.g. between simulation step size Δ and simulation horizon t_H w.r.t. the result error?
5. How does the trade-off between statistical significance and numeric error look like?
6. Shape of the tails of the statistics of the individual outcome errors (The tails are of special interest for risk assessments)?

5 Closing Remarks

- A report documenting the results of the case study would be helpful.
- If interested in this and similar topics in general, please contact me via e-mail.
- In case of substantial results, it is planned to make the case-study accessible online (subject to the condition that this is appreciated by the team members).

References

[1] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721, 1927.