

Rare event analysis in technological catastrophes

G. Rubino

Paris, ICT-DM'19

Outline

- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions

Outline

- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions

Motivation

- A **rare event** is an event occurring with a very small probability. How small? It depends on the area, the context, etc.
- In general, serious or catastrophic system's failures are, or must be, rare events.
- In this conference, the focus is on what to do *after* the occurrence of a catastrophic event on a communication service, how to prepare for those rare events in the networking area, etc.
- In our team at INRIA, we are interested in *predicting* things about rare events *before* their (next) occurrence.
- The most basic property of a rare event is its probability. How to evaluate it is the focus of this presentation.

Motivation (cont'd.)

- With the simplest assumptions, if the probability of occurrence of a rare event (implicitly, on some small period of time with length δ uot -units of time), is γ , then, on the average, we can expect δ/γ uot between those rare events, giving an idea of how often they happen.
- With better models (and more complex assumptions, and data) we can perhaps obtain more precise descriptions about the inter-event time (moments, distributions), or (in general, harder) about the time until the next occurrence.

Two main classes of rare events

- We often call **critical** those events with the property that when they occur, they have serious consequences (for instance, the crash of a communication network).
- Critical events are (should be) rare, and can be classified into two types: those with **artificial** causes (typically, failures), and those produced by **natural** phenomena.
- We will discuss only the analysis of the former; in general the tools available to analyze both types of events are different.
- Very roughly speaking, when the origin is artificial, the basic initial tool is the Central Limit Theorem. When causes are natural, it is Extreme Value Theory.

Multi-component systems

- Analyzing the properties of rare events of artificial origins is always done using **models** of the considered systems.
- In order to analyze the types of systems we are interested in, that is, communication ones, we always decompose the model into parts, and we call **components** the atomic ones.
- Both components and systems are supposed (in this talk) to be in one out of two possible states, working (\equiv up, \equiv operational) or failed (\equiv down, \equiv nonoperational). Most of the work is done inside this *binary* world. We typically code 'up = 1' and 'down = 0'.
- Once we have identified the criterion that defines a working system and a failed one, we must also understand the **structure** of the system, that is, how the different possible subset of components that failed, make or not that the whole system failed as well.

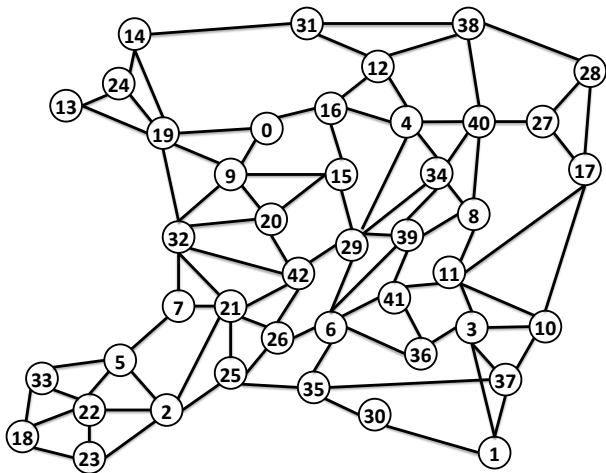
Static models versus dynamic models

- Models can be **static**, meaning that time doesn't play any role, or **dynamic**, where the model consists of some kind of stochastic process (for example, a Markov chain, in discrete or in continuous time).
- Since they are simpler, we will focus here on static systems' representations, and then analysis.
- Next section describes our reference model for the rest of the talk.

Outline

- 1 Introduction
- 2 Static models**
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions

A real-life example



European optical comm. infrastructure (43 nodes, 90 edges)

- The vertices of the graph are the nodes of the network, and the edges are the links or channels between nodes. In this example, we have (usual assumption) that the components (the atomic object that *can fail*) are the links. This is a modeler's choice.
- Say that $X_i = 1$ if link i is working, $X_i = 0$ if not, and suppose that these 90 Binary r.v.s are independent. We say that X_i is the **state of link i** .
- Assume we know (after measuring inside a lab) the number $r_i = \mathbb{P}(X_i = 1)$ for all link i , called the **reliability of component i** , or the **elementary reliability of i** .
- The **structure** of the system is now the mapping from 2^{90} possible vector of link states to $\{0, 1\}$, saying when the system is up or down.

- An important aspect of the robustness of the network topology is the fact that, when some links fail, it remains at least one path composed of working links only, between every pair of nodes, or between two particular nodes, or between the nodes in some subset of vertices.
- Denote by R the probability of one of these previously presented connectivity events, a central dependability parameter. Call it “system’s reliability”, or, in words, probability that the system works.
- The system’s structure is given by the graph plus the connectivity criterion defining when it is up, when it is down.
- If the failure of the system is a rare event, $R \approx 1$.

- Once transformed into a binary output problem, computing R is, in the general case, an NP-hard problem.
- Detail: for technical reasons, it is better to work with $\gamma = 1 - R$, the system's *unreliability*, rather than with R . So, we can write $\gamma \ll 1$, or $\gamma \approx 0$. We also denote $u_i = 1 - r_i$, the (elementary) unreliability of link i .
- For instance, previous example in slide 10 is completely out-of-reach for any of the many algorithms available for the exact evaluation of R (or equivalently, of γ).
- It remains the Monte Carlo approach: to **estimate** $\gamma = 1 - R$, perform $N \gg 1$ times the following:
 - **sample the state of each of the 90 links**, that is, sample the 90 Bernoulli variables X_1, \dots, X_{90} ;
 - **for each sampled instance, check if the system fails** (check the chosen connectivity criterion);

at the end, return the # of times the network failed divided by N .
- This is called **standard** or **naive** or **crude** Monte Carlo.

Confidence intervals

- Call $Y^{(n)}$ the r.v. 1 (the n th network's copy *fails*) in the execution of the standard Monte Carlo process.
- So, $Y^{(1)}, \dots, Y^{(N)}$ are N independent copies of $Y \sim \text{Bernoulli}$, with parameter γ : we have $Y = 0$ or 1 , and

$$\mathbb{P}(Y = 1) = \mathbb{P}(\text{network fails}) = \gamma.$$

Recall that $\mathbb{E}(Y) = \gamma$ and $\mathbb{V}(Y) = \gamma(1 - \gamma)$.

- The unknown number γ is then estimated by the ratio (the average)

$$\tilde{\gamma} = \frac{1}{N} \sum_{n=1}^N Y^{(n)}.$$

- Observe that $\tilde{\gamma}$, the *standard estimator* of γ , is a r.v., with the property of being *unbiased*, which means that $\mathbb{E}(\tilde{\gamma}) = \gamma$. Also,

$$\mathbb{V}(\tilde{\gamma}) = \frac{1}{N^2} N \mathbb{V}(Y) = \frac{\gamma(1 - \gamma)}{N}.$$

- Instead of returning simply $\tilde{\gamma}$, the right procedure is to compute a measure of the accuracy of the estimation, typically, a Confidence Interval for γ .
- The idea is that instead of making the computer say
“My estimation of γ is $3.18 \cdot 10^{-9}$ ”,
without providing any idea about the quality of the estimation, a correct output (when using a “confidence level” of 95%, for instance), would take the form
“I got the **confidence interval** ($3.04 \cdot 10^{-9}$, $3.32 \cdot 10^{-9}$) for γ , with confidence level 0.95. The middle-point of the interval, $3.18 \cdot 10^{-9}$, is my **point-estimation** of γ .”

- A standard way to do so is to apply the **Central Limit Theorem** which leads to the Confidence Interval

$$\left(\tilde{\gamma} \mp c_{\alpha} \sqrt{\frac{\tilde{\gamma}(1 - \tilde{\gamma})}{N - 1}} \right).$$

- Parameter α is a “confidence level” given beforehand, a number close to 1 (typical values: 0.95, 0.99, 0.999);
- Then, $c_{\alpha} = \Phi^{-1}((1 + \alpha)/2)$, where Φ^{-1} denotes the inverse of the Standard Normal c.d.f. For instance, $c_{0.95} = 1.960$, $c_{0.99} = 2.576$, $c_{0.999} = 3.291$.

- The half-size of the Confidence Interval can be interpreted as a probabilistic upper bound of the absolute error of the estimation. If we divide it by the point-estimation, we can interpret the ratio as a probabilistic upper bound of the corresponding relative error.
- To simplify, we say that the relative error is

$$RE = c_\alpha \sqrt{\frac{\tilde{\gamma}(1-\tilde{\gamma})}{N-1}} \frac{1}{\tilde{\gamma}} = c_\alpha \sqrt{\frac{1-\tilde{\gamma}}{(N-1)\tilde{\gamma}}}$$

- Given the usual values of the coefficient c_α , sometimes we remove it from previous expression, and we call Relative Error the number

$$RE = \sqrt{\frac{1-\tilde{\gamma}}{(N-1)\tilde{\gamma}}} \approx \frac{1}{\sqrt{N\tilde{\gamma}}}.$$

The rare event problem

- Assume that $\gamma = 10^{-9}$ (a “reference value” in the area, frequently appearing in networking, in transportation of people, and in other areas).
- This means that when running a standard Monte Carlo procedure to estimate γ , on the average we must simulate 1 Billion networks to observe a failed one! And we need at least some hundreds of observation of the event of interest.
- This kills the use of the standard Monte Carlo approach as well.

The rare event problem (cont'd)

- If $\gamma \approx 0$ (the usual situation) and $N \gg 1$, then $\tilde{\gamma} \approx 0$ and

$$RE \approx \frac{c_\alpha}{\sqrt{N\tilde{\gamma}}}.$$

- Looking at RE better shows the problem of $\gamma \approx 0$ (that is, of rare events), when we use the standard Monte Carlo approach: when the target γ is very small, and thus, also is its estimation $\tilde{\gamma}$, the relative error is usually very large, unless N is really huge.

The rare event problem (cont'd)

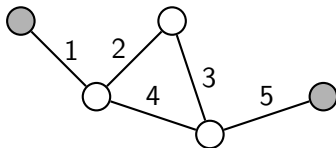
- Keep this $R = 1 - 10^{-9}$ and pick a classic 95% confidence for the estimations.
- Assume we want a modest 10% of relative error in the answer.
- Then, writing $1/\sqrt{N} \cdot 10^{-9} \leq 10^{-1}$ leads to $N \geq 10^{11}$. If each instance of the network needs 1 sec of CPU time for its processing (sampling and checking connectivity), we need about 12000 years to estimate γ . Too bad. And if each instance needs just 1 msec (optimistic), we still need 12 years to get the answer.

What to do?

- We must move to specialized methods that can deal with rare events, but with no “universality”, not in all cases, at least without supplementary efforts, etc. This is a hot research area.
- The techniques we will mention now often take seconds, or minutes, to evaluate models as the one shown before and with numbers as the preceding ones.

On exact computation of R or γ

- For instance, let \mathcal{G} be the graph



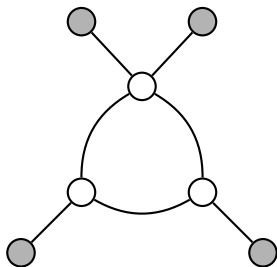
where we are interested in the possibility of communication between the two grey nodes.

- We have

$$R = r_1 [1 - (1 - r_4)(1 - r_2 r_3)] r_5 = r_1 (r_2 r_3 + r_4 - r_2 r_3 r_4) r_5$$

A “small Internet model”

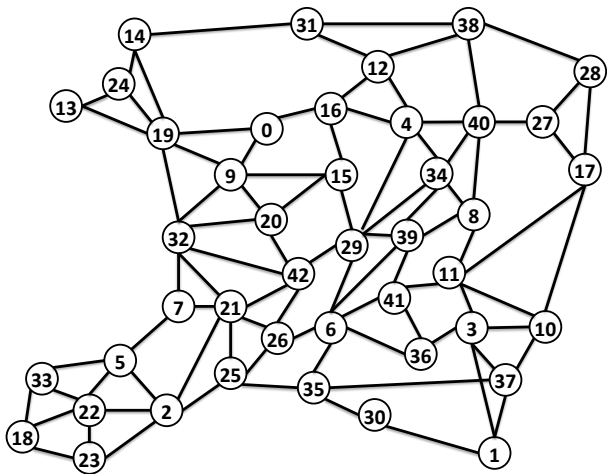
Consider this other example representing a typical communication network, with a set of terminal machines (the grey nodes) connected through a *backbone* (the ring of white nodes).



If r is the elementary reliability of any node and R is the probability that the terminals can communicate, then

$$R = r^4(r^3 + 3r^2(1 - r)) = r^6(3 - 2r).$$

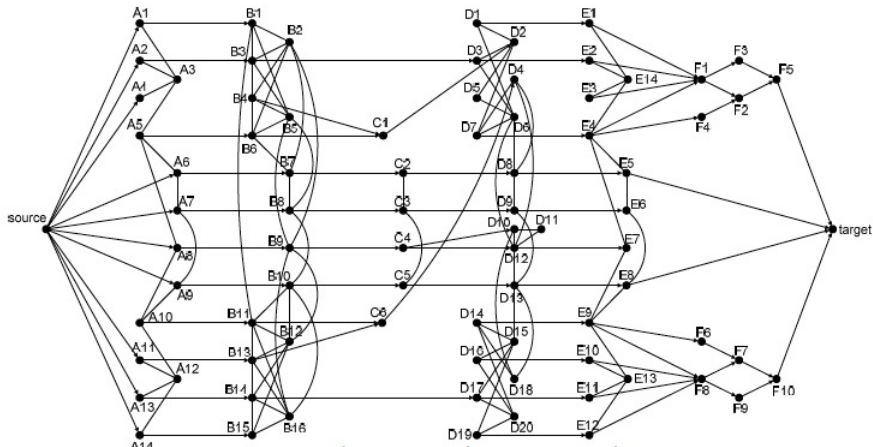
Previous out-of-reach example for exact analysis



Analyzed using Monte Carlo specialized techniques by our team.  

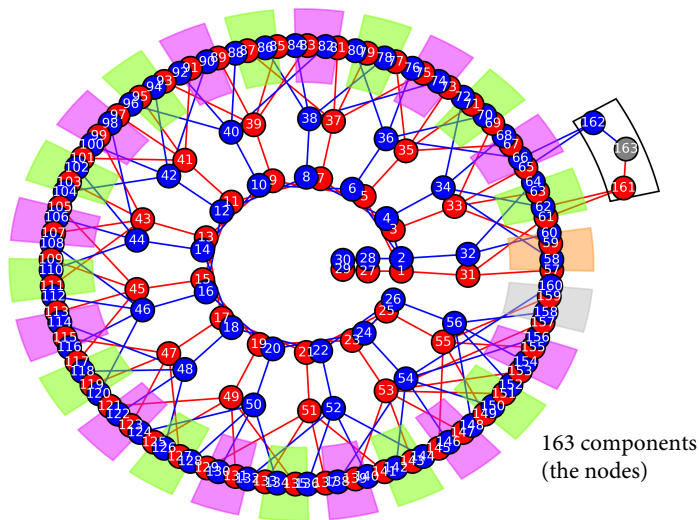
Another “out-of-reach” case

- One of many models used in the analysis of the Boeing Dreamliner 787 aircraft:



- 82 nodes, 171 links (arcs, here)

A subway communication network



- This last example concerns a specific communication system (and several other similar ones) designed by ALSTOM for their recent underground train systems.
- The model comes from a recent project developed by our team at INRIA, for Alstom.

Outline

- 1 Introduction
- 2 Static models
- 3 Dealing with rare events**
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions

Specialized techniques

The rarity aspect precludes the use of a naive (classical, standard, direct) approach: **specialized techniques are then needed.**

The specialized techniques designed for rare event analysis belong usually to some well studied families of methods. The main ones in the Monte Carlo area (simulation) are the following:

- **Importance Sampling simulation techniques**
(with, in particular, Zero-Variance versions)
- **Splitting simulation techniques** (mainly for dynamic models)
- Recursive Variance Reduction simulation techniques (mainly for static models)
- ...

Alternatives

Leaving the simulation world, we have (in very specific cases) other possibilities not covered today:

- **Bounding techniques.** In some rare event situations, we can build very tight bounds of the metrics of interest. More than that, in several cases of strong practical interest, the bounds improve when the event becomes rarer.
- **Mean field analysis.** Powerful and very useful approach in specific families of models.
- ...

In the sequel, we will provide some details about one specific technique belonging to the Monte Carlo family.

Outline

- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling**
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions

Importance Sampling

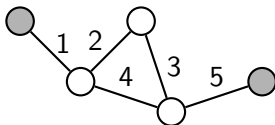
- To put things as simple as possible, assume that we have some random vector $\vec{X} \in \mathbb{N}^d$ and that we need to estimate $\gamma = \mathbb{P}(\vec{X} \in A) \ll 1$.
- Let $p()$ denote the probability mass function of \vec{X} . We have

$$\gamma = \sum_{\vec{x} \in \mathbb{N}^d} 1(\vec{x} \in A) p(\vec{x}).$$

- In many cases, $p()$ is unknown, but we can sample it. In other cases, $p()$ is known but the sum has a huge number of terms. In both situations, when we can not calculate γ numerically, we can often efficiently sample from $p()$, so that we can estimate γ by simulation.

Example

- Let's come back to



- Here, $\vec{X} = (X_1, \dots, X_5)$. Concerning its distribution, assuming $r_1 = \dots = r_5 = r$ to simplify things, some examples:
 $p(1, 1, 1, 1, 1) = r^5$, $p(1, 0, 1, 0, 1) = r^3(1 - r)^2$, etc.
- The set A is
 $\{(1, 1, 1, 1, 1), (1, 1, 1, 0, 1), (1, 0, 0, 1, 1), (1, 0, 1, 1, 1), (1, 1, 0, 1, 1)\}$.
- $\gamma = p(1, 1, 1, 1, 1) + p(1, 1, 1, 0, 1) + p(1, 0, 0, 1, 1) + p(1, 0, 1, 1, 1) + p(1, 1, 0, 1, 1)$.

- Now, assume that $\hat{p}(\cdot)$ is another mass probability function on \mathbb{N}^d , such that if $\vec{x} \in A$ and $p(\vec{x}) > 0$, then $\hat{p}(\vec{x}) > 0$ as well. We have

$$\gamma = \sum_{\vec{x} \in \mathbb{N}^d} 1(\vec{x} \in A) p(\vec{x}) = \sum_{\vec{x} \in \mathbb{N}^d: \vec{x} \in A, p(\vec{x}) > 0} p(\vec{x}) = \sum_{\vec{x} \in \mathbb{N}^d: \vec{x} \in A, p(\vec{x}) > 0} \frac{p(\vec{x})}{\hat{p}(\vec{x})} \hat{p}(\vec{x}).$$

- Then, denoting $L(\vec{x}) = p(\vec{x})/\hat{p}(\vec{x})$ when $\hat{p}(\vec{x}) > 0$, and 0 otherwise (called the *likelihood ratio*), $\gamma = \sum_{\vec{x} \in \mathbb{N}^d} 1(\vec{x} \in A) L(\vec{x}) \hat{p}(\vec{x})$ can be written

$$\gamma = \mathbb{E}_{\hat{p}}(1(\vec{Y} \in A) L(\vec{Y})).$$

- This means that a new estimator of γ is

$$\hat{\gamma} = \frac{1}{N} \sum_{n=1}^N 1(\vec{Y}^{(n)} \in A) L(\vec{Y}^{(n)}),$$

where $\vec{Y}^{(1)}, \dots, \vec{Y}^{(N)}$ are N independent copies of $\vec{Y} \sim \hat{p}$.

- The corresponding Confidence Interval is

$$\left(\hat{\gamma} \mp c_\alpha \frac{\hat{\sigma}}{N} \right),$$

where

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{n=1}^N \mathbf{1}(\vec{Y}^{(n)} \in A) L(\vec{Y}^{(n)}) - \frac{N}{N-1} \hat{\gamma}^2.$$

- So, we are estimating γ by estimating an artificial new thing (the expectation of the r.v. $1(\vec{Y} \in A)L(\vec{Y})$ where $\vec{Y} \in \mathbb{N}^d$ follows the new dynamics $\hat{p}()$).
- We say that we did a “change of measure” (c.o.m.), from $p()$ to $\hat{p}()$.
- The question is how good is this new estimator $\tilde{\gamma}$ (that is, **how small is its variance** and **how expensive is sampling from $\hat{p}()$**)?
- The answer is that by choosing $\hat{p}()$ carefully, we may obtain better estimators than the crude one, sometimes thousands of times more efficient.
- More precisely, with the same cost (execution time), which often reduces basically to take the same sample size N , the variance of the IS estimator can be thousands of times smaller than the variance of the standard one.
- Roughly speaking, as we will immediately see, we must use some $\hat{p}()$ such that **the rare event comes frequently**, that is, is not rare anymore, **but not too much**.

Comparing variances

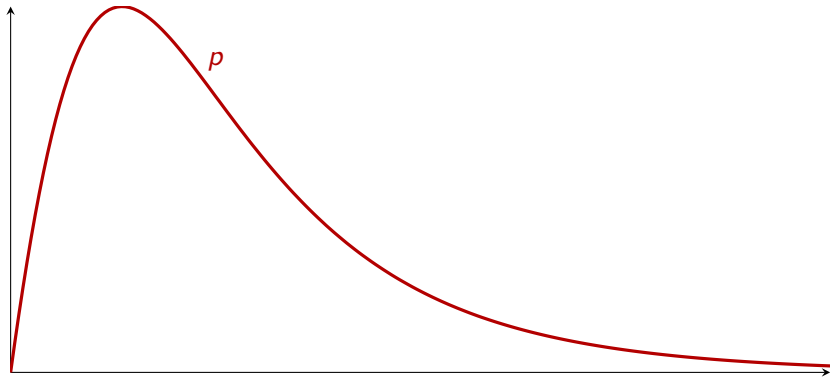
- Let us compare the variances of the standard estimator of γ , which we denote by $\tilde{\gamma}$, and of the IS estimator $\hat{\gamma}$, for the same sample size N .
- So, $\tilde{\gamma} = N^{-1} \sum_{n=1}^N \mathbf{1}(\vec{X}^{(n)} \in A)$ where $\vec{X}^{(1)}, \dots, \vec{X}^{(N)}$ are independent copies of $\vec{X} \sim p(\cdot)$, and $\hat{\gamma} = N^{-1} \sum_{n=1}^N \mathbf{1}(\vec{Y}^{(n)} \in A) L(\vec{Y}^{(n)})$ where $\vec{Y}^{(1)}, \dots, \vec{Y}^{(N)}$ are independent copies of $\vec{Y} \sim \hat{p}(\cdot)$.
- See that

$$\mathbb{V}(\tilde{\gamma}) = \frac{1}{N} \mathbb{V}(\mathbf{1}(\vec{X} \in A)) \quad \text{and} \quad \mathbb{V}(\hat{\gamma}) = \frac{1}{N} \mathbb{V}(\mathbf{1}(\vec{Y} \in A) L(\vec{Y})).$$

- Writing $\mathbb{V}(\tilde{\gamma}) = N^{-1} [\mathbb{E}(1(\vec{X} \in A)^2) - \gamma^2]$ and $\mathbb{V}(\hat{\gamma}) = N^{-1} [\mathbb{E}(1(\vec{Y} \in A)^2 L^2(Y)) - \gamma^2]$, comparing variances reduces to comparing second moments.
- So, IS is better than the standard estimator iff $\mathbb{E}(1(\vec{X} \in A)) > \mathbb{E}(1(\vec{Y} \in A) L^2(Y))$.
- We see then that the element that controls this issue is the value of the likelihood ratio, and what we need is $\hat{p}(\vec{x}) > p(\vec{x})$ when we are in A (and when $p(\vec{x}) > 0$).

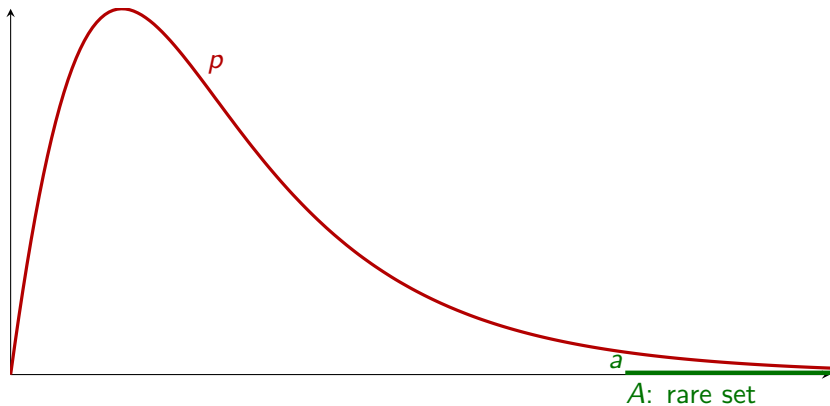
IS illustration

The original law of X (seen “from so far” that it looks continuous).



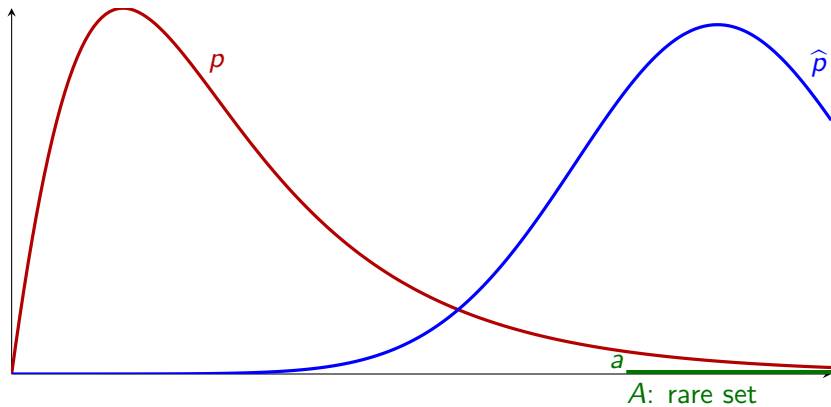
IS illustration

The rare set A .



IS illustration

The change of measure $p \rightarrow \hat{p}$.



On the perfect IS estimator

- Consider the c.o.m.

$$\hat{p}(\vec{x}) = \frac{\mathbf{1}(\vec{x} \in A)p(\vec{x})}{\gamma}.$$

- We have

$$\mathbb{V}(\hat{\gamma}) = \mathbb{V}(\mathbf{1}(\vec{Y} \in A)L(\vec{Y})) = 0.$$

- You “sample” only once ($N = 1$), and the result is exactly γ .
- Of course, this looks useless since we need the target (the value of γ) to implement it. More on this later.
- This suggests as a general rule, to try to make the new dynamics proportional to the numerator, something that “follows” the shape of $p(\cdot)$ on A .
- A more “high level” rule of thumb is “make rare things frequent” in the new dynamics ... “but not too much”.

Making rare things too frequent

A classic toy but illustrative example.

- Suppose that the life-time L of a system follows the Exponential distribution¹ with parameter λ . In such a context, λ is the system's *failure rate*. We want to estimate the probability γ that the failure arrives before a given time T .
- We assume that we know λ (in which case we of course know the answer:

$$\gamma = \mathbb{P}(L \leq T) = 1 - e^{-\lambda T}.)$$

- We will estimate γ using the standard estimator, and then, we will try IS using as the new measure another Exponential law with some rate $\hat{\lambda}$ that we will vary, to explore the efficiency of IS as a function of it.

¹This is reasonable for many types of systems.

- Write $\psi(t) = 1(t \leq T)$. The target is

$$\gamma = \int_0^{\infty} \psi(x) f(x) dx = \int_0^T \lambda e^{-\lambda x} dx = 1 - e^{-\lambda T}.$$

- If $t \geq 0$, the likelihood ratio at t is

$$L(t) = \frac{f(t)}{\widehat{f}(t)} = \frac{\lambda e^{-\lambda t}}{\widehat{\lambda} e^{-\widehat{\lambda} t}}.$$

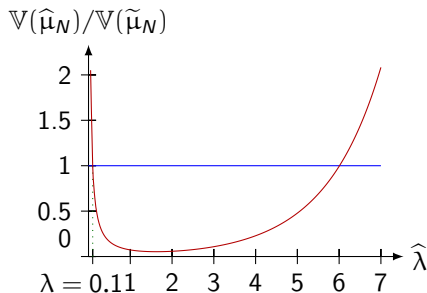
- If we compare the second moments, we have that the IS method is better (smaller variance) than the standard one if and only if

$$\frac{\int_0^{\infty} \psi^2(t) L(t) f(t) dt}{\int_0^{\infty} \psi^2(t) f(t) dt} = \frac{\int_0^T L(t) f(t) dt}{\int_0^T f(t) dt} \leq 1.$$

- For instance,

$$\int_0^T L(t)f(t) dt = \int_0^T \frac{\lambda e^{-\lambda t}}{\hat{\lambda} e^{-\hat{\lambda} t}} \lambda e^{-\lambda t} dt = \frac{\lambda^2}{\hat{\lambda}} \frac{1 - e^{-(2\lambda - \hat{\lambda})T}}{2\lambda - \hat{\lambda}}.$$

- Once we do the computations for the variance of the IS estimator (messy), we can analyze the evolution of the ratio $\mathbb{V}(\hat{\gamma})/\mathbb{V}(\tilde{\gamma})$ with $\hat{\lambda}$. See here a graphical view when we fix $\lambda = 1$ and $T = 0.1$:



- The point is that if we start with a small value of $\hat{\lambda}$, when we increase it the IS method improves in efficiency, and when $\hat{\lambda}$ becomes greater than some $\hat{\lambda}_0$, the IS estimator has a smaller variance than the standard one.
- But after a second threshold $\hat{\lambda}_1$, the IS estimator becomes worse than the standard one, and its variance grows without any bound as $\hat{\lambda} \uparrow \infty$.
- This simple example shows that we must make the rare event become frequent under the new measure, *but not “excessively” frequent*.
- We can anticipate then the fact that finding a good change of measure is in many cases hard, and there is no general way to do it. It strongly depends on the problem.

Best c.o.m. inside a specific family

- Remember the design rule of thumb: “use a change of measure that makes frequent what was rare, but moderately”.
- In the static world, a **first and simple idea** is straightforward: change the individual unreliabilities (which are close to 0) into numbers that are, say, between 0.5 and 0.8. This is called “Failure Biasing” (FB) and it already works fine (even if it is possible to do much better).
- As a particular but frequent case, assume an homogeneous situation where the components' reliabilities are all the same, $= r$, and we want to use FB by changing r into some $r' < r$.
- We can prove that the optimal value r' (**optimal inside this specific family of FB c.o.m.**) is $r' = c/M$, where M is the number of links and c is the size of a cut of minimal size.

- As a more generic example, there is a methodology called **cross entropy** that can be used to find the best c.o.m. inside a given parametric family of distributions (see the bibliographic references at the end).
- It is possible to design a *sequential* way of simulating the system using IS, and modify the c.o.m. as a function of the evolution of the simulation.
- Previous way to operate is called state-dependent or adaptive IS. We typically find it when the model is a stochastic process, but it can appear when dealing with static models. The idea is to build an artificial stochastic process related to the static model (the “sequential way” mentioned above) and then, to adapt the c.o.m. according to observed events appearing during the simulation process.

Back at the optimal c.o.m.

- Recently, the optimal and in principle useless change of measure led to a very powerful idea, that became a hot topic these days in the area: **the zero-variance approach**.
- The idea is to transform this global view “ $\hat{p}(\vec{x}) = 1(\vec{x} \in A)p(\vec{x})/\gamma$ ” into “local relationships inside the model”, local connection between different quantities related to the many components of the model.
- As we could expect, the global form of the optimal \hat{p} that requires to know the exact answer to the original question (so, no need to simulate) translates into “atomic” or “detailed” relations where things like γ appear.
- The gain happens when instead of \hat{p} we use some \hat{p}^* , built using these detailed expressions, where we replace the forever unknown exact quantities by approximations, even rough ones. The resulting IS procedure is sometimes extremely efficient.

A last word on our reference model

- Consider the case of the source-to-terminal reliability estimation:
 - a connected non-oriented graph,
 - two nodes marked,
 - components are the edges, sharing the same elementary reliability r ,
 - the network works iff the two marked nodes can communicate.
- Let M be the number of edges, and c be the graph's breadth, that is, the minimal size of a mincut separating the two marked nodes.
- Then, if we consider the changes of measure in the family of the models where edges have the same elementary reliability, then the optimal one (inside the family) is when the new common elementary reliability is $1 - c/M$.

A powerful IS technique

- A few years ago, we published “*Balanced and Approximate Zero-Variance Recursive Estimators for the Static Communication Network Reliability Problem*”, TOMACS, 25(1): 5:1-5:19, 2015, co-authored with H. Cancela, M. El Khadiri and B. Tuffin.
- The paper describes a new approach combining a powerful recursive estimation technique with an Importance Sampling scheme approximating the zero-variance one.
- We claim that it is (at least) at the top positions in the state-of-the-art technology for solving these rare events problems associated with a network reliability problem.
- The paper allows also to illustrate important properties of the estimators, as well as the problems of proving them on specific cases.

Outline

- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators**
- 6 Numerical illustrations
- 7 Conclusions

Bounded Relative Error

- An unbiased estimator Y' of $\mathbb{E}[Y]$ has *Bounded Relative Error* (BRE) if RE remains bounded as the event becomes rarer.
- Formally, we have BRE if $\sqrt{\mathbb{V}(Y')}/\mathbb{E}[Y]$ seen as a family of functions indexed by the sample size N , is uniformly bounded when $\mathbb{E}[Y] \rightarrow 0$, or equivalently, if $\mathbb{E}(Y'^2)/\mathbb{E}^2(Y)$ is uniformly bounded as $\mathbb{E}[Y] \rightarrow 0$.
- BRE implies that the sample size required to get a given relative error is not sensitive to the rarity of the event.
- The standard estimator does not possess this property.

Other relevant properties of estimators

- **Weaker than BRE.** An unbiased estimator Y' of $\mathbb{E}[Y]$ is *Asymptotically Optimal* (AO) or *Logarithmically Efficient* (LE) if

$$\lim_{\mathbb{E}(Y) \rightarrow 0} \frac{\ln \mathbb{E}(Y'^2)}{\ln \mathbb{E}(Y)} = 2.$$

- **Stronger than BRE.** An unbiased estimator Y' of $\mathbb{E}[Y]$ verifies the *Vanishing Relative Error* (VRE) property if $\sqrt{\mathbb{V}(Y')/\mathbb{E}[Y]} \rightarrow 0$ as $\mathbb{E}[Y] \rightarrow 0$.

On the AO or LE property

- To understand the connection of the definition of AO with the subject, consider the IS estimator $\hat{\gamma}$ of previous section and recall the expression

$$\mathbb{V}(\hat{\gamma}) = \frac{1}{N} \left[\mathbb{E}_{\hat{\rho}}(\mathbf{1}(\vec{Y} \in A)^2 L^2(Y)) - \gamma^2 \right].$$

- Denote $E = \mathbb{E}_{\hat{\rho}}[\mathbf{1}(\vec{Y} \in A)L^2(Y)]$. We have $\mathbb{V}(\hat{\gamma}) = (E - \gamma^2)/N$, so, $E = N\mathbb{V}(\hat{\gamma}) + \gamma^2 \geq \gamma^2$.
- Then, $\ln E \geq 2 \ln \gamma$, and

$$\frac{\ln E}{\ln \gamma} \leq 2.$$
- See that $\ln E = 2 \ln \gamma$ leads to $\mathbb{V}(\hat{\gamma}) = 0$.

Rarity parameter

- More generally, we often use a *rarity parameter* to “move the model toward a rare situation” and explore the behavior of estimators (a kind of asymptotic analysis).
- For instance, in the static case, we can set
 - $r_i = r = 1 - \varepsilon$ (homogeneous model), and let $\varepsilon \rightarrow 0$,
 - or $r_i = 1 - a_i \varepsilon^{b_i}$ (heterogeneous model), with $a_i > 0$ and $b_i > 0$, and let $\varepsilon \rightarrow 0$,

and then define BRE, AO, VRE, etc., with respect to $\varepsilon \rightarrow 0$.

- If the model is a queuing one and the rare event is the saturation of the queue whose storage capacity is H , then the rarity parameter can be H when $H \uparrow \infty$.

Other measures

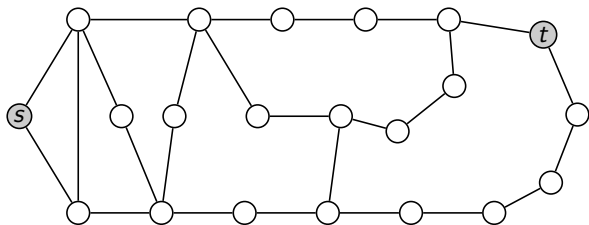
- There is another set of properties where we look not only at the variance but also to the cost in time of the execution of the algorithm implementing the estimator.
- Let us denote by $\tau(\hat{\gamma}_N)$ the mean execution time of the estimation code where we added explicitly the sample size N .
- Then what we want is to have $\mathbb{V}(\hat{\gamma}_N)$ small and also $\tau(\hat{\gamma}_N)$ small. This is the basis of another set of definitions. For instance, the relative efficiency of the estimator is defined as the ratio $\gamma / (\mathbb{V}(\hat{\gamma}_N)\tau(\hat{\gamma}_N))$.
- Then, we define Bounded Relative Efficiency, etc.

Outline

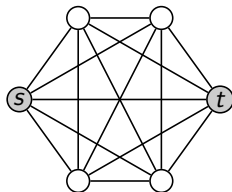
- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations**
- 7 Conclusions

Topologies: arpanet, complete graphs,

Arpanet:

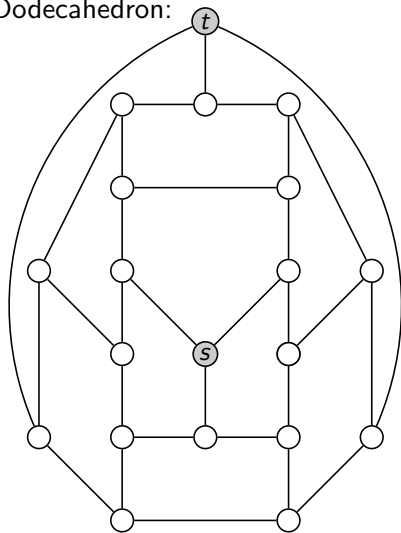


C_6 :

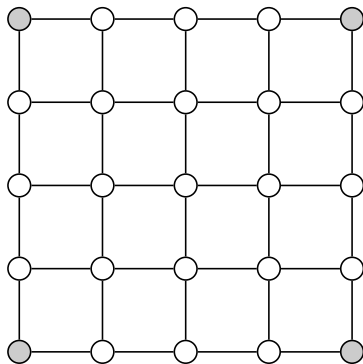


... dodecahedron and grids

Dodecahedron:



Grid, order 5:



Comparisons

The (normalized) relative error² for various methods and unreliabilities ε of links (homogeneous case), on the dodecahedron topology

Method	$\varepsilon = 0.1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$
[F]	2.6 e+00	1.3 e+00	4.3 e-01	1.4 e-02
[S1]	4.0 e+00	6.2 e+00	7.7 e+00	8.9 e+00
[S2]	4.6 e+00	7.1 e+00	8.6 e+00	8.8 e+00
[B]	3.0 e+00	4.2 e+00	4.3 e+00	4.4 e+00
[Z]	1.2 e+00	1.7 e-01	5.7 e-02	1.7 e-02
[R]	8.4 e-01	7.1 e-01	7.1 e-01	7.1 e-01
BRD [A]	9.5 e-01	7.0 e-01	7.1 e-01	7.1 e-01
AZVRD [A]	2.8 e-01	5.1 e-02	1.6 e-02	5.0 e-03

²Denoting RE_N the relative error, we use here $\sqrt{N} \cdot RE_N/z$, with, say, $z = 1.96$.

- F G. S. Fishman. 1986. A Monte Carlo sampling plan for estimating network reliability. *Operations Research* 34, 4, 581–594 (method based on bounds).
- S1 Z. I. Botev, P. L'Ecuyer, G. Rubino, R. Simard and B. Tuffin. 2013. Static network reliability estimation via generalized splitting. *INFORMS Journal on Computing* 25, 1, 56–71 (a generalization of splitting).
- S2 L. Murray, H. Cancela, and G. Rubino. 2013. A splitting algorithm for network reliability. *IIE Transactions* 45, 2, 177–189 (another adaptation of splitting to static problems).
- B I. B. Gertsbakh and Y. Shpungin. 2010. *Models of Network Reliability*. CRC Press, Boca Raton, FL, US (the so-called turnip method).
- Z P. L'Ecuyer, G. Rubino, S. Saggadi and B. Tuffin. 2011. Approximate zero-variance importance sampling for static network reliability estimation. *IEEE Transactions on Reliability* 8, 4, 590–604 (another zero-variance approximation).
- R H. Cancela and M. El Khadiri. 1995. A recursive variance-reduction algorithm for estimating communication-network reliability. *IEEE Transactions on Reliability* 44, 4, 595–602 (the original RVR technique).
- A H. Cancela, M. El Khadiri, G. Rubino and B. Tuffin. 2015. Balanced and Approximate Zero-Variance Recursive Estimators for the Static Communication Network Reliability Problem. *TOMACS*, 25(1): 5:1–5:19 (our paper's).

Illustration of the BRE and VRE properties

Network	ϵ	$q(G)$	$\frac{\sqrt{n} \times RE_{SMC}}{c_\alpha}$	$\frac{\sqrt{n} \times RE_{RVR}}{c_\alpha}$	$\frac{\sqrt{n} \times RE_{BRD}}{c_\alpha}$	$\frac{\sqrt{n} \times RE_{AZVRD}}{c_\alpha}$
Arpanet	5 e-01	9.6398994 e-01	1.93 e-01	6.33 e-02	4.16 e-01	4.27 e-01
Arpanet	3 e-01	6.8150724 e-01	6.84 e-01	3.20 e-01	1.10 e+00	1.35 e+00
Arpanet	1 e-01	9.5422918 e-02	3.08 e+00	1.27 e+00	2.01 e+00	3.24 e+00
Arpanet	1 e-03	6.0558106 e-06	4.06 e+02	2.09 e+01	1.24 e+00	9.67 e-01
Arpanet	1 e-05	6.0005600 e-10	4.08 e+04	2.11 e+02	1.26 e+00	9.82 e-02
Dod	5 e-01	7.0974499 e-01	6.39 e-01	1.77 e-01	9.17 e-01	5.17 e-01
Dod	3 e-01	1.6851806 e-01	2.22 e+00	5.70 e-01	1.93 e+00	7.70 e-01
Dod	1 e-01	2.8796013 e-03	1.86 e+01	8.37 e-01	9.53 e-01	2.76 e-01
Dod	1 e-03	2.0060181 e-09	2.23 e+04	7.08 e-01	7.06 e-01	1.59 e-02
Dod	1 e-05	2.0000600 e-15	2.24 e+07	7.07 e-01	7.07 e-01	1.58 e-03
Grid5	5 e-01	9.6062484 e-01	2.02 e-01	2.66 e-02	4.55 e-01	1.01 e-01
Grid5	3 e-01	5.2094890 e-01	9.59 e-01	1.53 e-01	1.17 e+00	2.29 e-01
Grid5	1 e-01	4.8160510 e-02	4.45 e+00	1.40 e-01	1.09 e+00	1.35 e-01
Grid5	1 e-03	4.0080020 e-06	4.99 e+02	1.58 e-02	1.14 e+00	1.37 e-02
Grid5	1 e-05	4.0000800 e-10	5.00 e+04	1.58 e-03	1.15 e+00	1.37 e-03
C6	5 e-01	7.6416016 e-02	3.48 e+00	1.15 e-01	3.43 e-01	1.12 e-01
C6	3 e-01	5.2672775 e-03	1.37 e+01	9.61 e-02	5.32 e-01	9.06 e-02
C6	1 e-01	2.0076587 e-05	2.23 e+02	1.78 e-02	7.53 e-01	1.71 e-02
C6	1 e-03	2.0000000 e-15	2.24 e+07	1.58 e-05	8.65 e-01	1.58 e-05
C6	1 e-05	2.0000001 e-25	2.24 e+12	1.89 e-08	8.66 e-01	1.89 e-08
C10	5 e-01	1.9550825 e-02	7.08 e+00	2.10 e-01	3.65 e+01	3.13 e-01
C10	3 e-01	1.9690832 e-04	7.13 e+01	2.21 e-01	7.33 e+01	4.35 e-01
C10	1 e-01	1.0000004 e-08	1.00 e+04	3.33 e-01	1.04 e+02	5.95 e-01
C10	1 e-03	5.9991786 e-27	1.29 e+13	5.27 e+00	1.17 e+01	4.99 e-01
C10	1 e-05	4.1102231 e-45	1.56 e+22	7.69 e+01	2.63 e+00	2.70 e-01

Outline

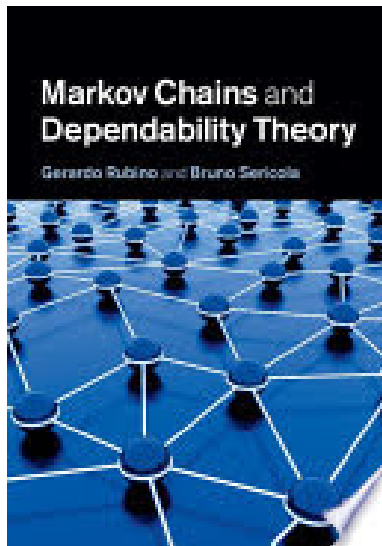
- 1 Introduction
- 2 Static models
- 3 Dealing with rare events
- 4 Importance Sampling
- 5 Main properties of estimators
- 6 Numerical illustrations
- 7 Conclusions**

Conclusions

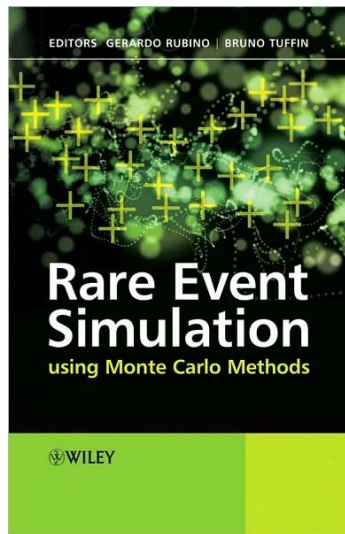
- When we must estimate the probability of a rare event defined on a complex model precluding any exact analysis, straightforward (naive, crude, standard) simulation doesn't work.
- This is also the case if we want to analyze some variable that (partially) depends on a rare event.
- Good news: the three families of methods previously mentioned may provide procedures that can evaluate in a very short time variables that would take excessive computing time with standard approaches.
- Bad news: there is no universal approach, and, often, ad hoc methods must be developed, with some effort. In difficult situation, finding an efficient technique becomes a research problem.
- But, all in all, very fast procedure can be exhibited for a large number of applications.

Cambridge University Press
2014

- Material on lumping, sojourn times, bounding techniques, etc.



- Book edited by G. Rubino and B. Tuffin. Covering many fundamental issues and techniques in the area.
- Content:
 - *Introduction to Rare Event Simulation* (G. Rubino and B. Tuffin)
 - Part I: Theory
 - *Importance Sampling* (P. L'Ecuyer, M. Mandjes and B. Tuffin)
 - *Splitting Techniques* (P. L'Ecuyer, F. Le Gland, P. Lezaud and B. Tuffin)
 - *Robustness Properties and Confidence Interval Reliability Issues* (P. Glynn, G. Rubino and B. Tuffin)



2009, Wiley

Book (cont'd)

The rest of the chapters (Part II: Applications)

- *Rare Event Simulation for Queues* (J. Blanchet and M. Mandjes)
- *Markovian Models for Dependability Analysis* (G. Rubino and B. Tuffin)
- *Rare Event Analysis by Monte Carlo Techniques in Static Models*, (H. Cancela, M. El Khadiri and G. Rubino)
- *Rare Event Simulation and Counting Problems* (J. Blanchet and D. Rudoy)
- *Rare Event Estimation for a Large-Scale Stochastic Hybrid System with Air Traffic Application* (H. A. P. Blom, G. J. Bakker and J. Krystul)
- *Particle Transport Applications* (T. Booth)
- *Rare Event Simulation Methodologies in Systems Biology* (W. Sandmann)