Bridge Monte-Carlo: a novel approach to rare events of Gaussian processes

S. Giordano[♮], M. Gubinelli[†], M. Pagano[♮]

[♮] Dept. of Information Engineering, University of Pisa, Via Caruso, I-56122 Pisa, Italy

[†] Dept. of Applied Mathematics, University of Pisa, Italy Via Bonanno 25/b, I-56125 Pisa, Italy

Abstract

In this work we describe a new technique, alternative to Importance Sampling (IS), for the Monte-Carlo estimation of rare events of Gaussian processes which we call Bridge Monte-Carlo (BMC). This topic is relevant in teletraffic engineering where queueing systems can be fed by long-range dependent stochastic processes usually modelled through fraction Brownian motion. We show that the proposed approach has clear advantages over the widespread *single-twist* IS and at the same time has the same computational complexity.

1 Introduction

We analyse the efficient simulation of a single server queueing system with infinite buffer loaded by a generic Gaussian traffic. In particular, we focus on the estimation of the overflow probability (defined as the probability that the steady-state queue-length Q exceeds a given threshold b) that represents an upper bound for the loss probability in the corresponding finite-buffer system. The overflow event $\{Q \geq b\}$ can be written as

$$\{Q \geq b\} = \{\sup_{t \geq 0}(X_t - \varphi_t) \geq 0\} \tag{1}$$

where $\varphi_t = b + t\mu$, μ is the difference between the mean values of the service and arrival rates and $X = \{X_t\}$ is a Gaussian noise modelling the fluctuation of the input traffic. For a general Gaussian process this probability has not a closed form and in many cases of interest (in telecommunication and in finance, for instance) it is very small, which renders usual MC estimation ineffective.

To overcome this problem IS techniques can be applied. It is well known that the efficiency of an IS-based algorithm depends on the choice of a proper change of measure to reduce the variance of the estimate. Recent works [2, 4] show that single-twist IS (which consists in a change of measure chosen within the class of pdfs that differ from the original one only by a shift in the mean value) cannot be efficient in the Large Deviations limit if the input is fractional Brownian motion (fBm) with Hurst parameter $H \neq 0.5$. Recently there have been proposed other two methods [4, 5] based on IS which are able to achieve asymptotic optimality, but at the cost of higher computational costs and complexity in the implementation.

Here we propose an alternative strategy to the efficient MC estimation of the overflow probability, referred in the following as Bridge Monte-Carlo (BMC), which exploits the Gaussian nature of the input process (the only assumption is the knowledge of the correlation structure of the incoming traffic flow). BMC estimator, which does not rely on a change of measure, can be derived by expressing the overflow probability as the expectation of a function of the *Bridge* Y of the Gaussian input process X, i.e. the process obtained by conditioning X to reach a certain level at some prefixed time.

2 Setting of the problem

Let $\{X_t\}_{t\in\mathcal{I}}$ be a continuous real Gaussian process, defined on the set $\mathcal{I}$, with zero mean and covariance $\Gamma_{t,s} = \mathrm{E}[X_t X_s]$, $t, s \in \mathcal{I}$. We will assume that $\Gamma_{t,t} > 0$ for any $t \in \mathcal{I}$ except a point $t_0 \in \mathcal{I}$ for which $\Gamma_{t_0,t_0} = 0$. We will consider a finite index set $\mathcal{I}$: in this case the process X is just a random vector in $\mathcal{X} = \mathbb{R}^{|\mathcal{I}|}$ ($|\mathcal{I}|$ is the cardinality of $\mathcal{I}$) which we will consider a Banach space using the Euclidean norm.

Let us denote by $\mathcal{H}$ the *reproducing kernel Hilbert space* of X and let us introduce the metric $|\cdot|_{\mathcal{H}}$ on $\mathcal{X}$ associated to the finite-dimensional Gaussian process X and defined as

$$|\rho|^2_{\mathcal{H}} = \langle \rho, \rho \rangle_{\mathcal{H}} = \langle \rho, \Gamma^{-1}\rho \rangle = \sum_{t,s\in\mathcal{I}} [\Gamma^{-1}]_{t,s}\rho_t\rho_s \tag{2}$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean scalar product of $\mathcal{X}$ and Γ^{-1} is the inverse of the covariance matrix Γ.

Taking into account equation (1), we are interested in the evaluation of the following probability

$$\mathrm{P}\left(\sup_{t\in\mathcal{I}}(X_t - \varphi_t) \geq 0\right) \tag{3}$$

for some fixed function $\varphi : \mathcal{I} \to \mathbb{R}$.

To study the behaviour of the estimators when the probability of interest is small, we will introduce a small parameter ε in the problem and consider the probabilities p_ε defined as

$$p_\varepsilon = \mathrm{P}\left(\sup_{t\in\mathcal{I}}(\varepsilon X_t - \varphi_t) \geq 0\right) \tag{4}$$

In a network framework, equation (4) naturally arises in the *many-sources regime*, i.e. when n i.i.d. Gaussian sources are aggregated and queueing resources (buffer size and service rate) are scaled with n. It is well known that buffer overflow (over level nb) becomes a rare event when $n \to \infty$ as a consequence of statistical multiplexing; the overflow probability, in this case, is given by

$$\begin{aligned}
\mathrm{P}\left(Q_t \geq nb\right) &= \mathrm{P}\left(\sup_{t\in\mathcal{I}}\left(\sum_{i=0}^{n} X_t^{(i)} - n\mu t\right) \geq nb\right) \\
&= \mathrm{P}\left(\sup_{t\in\mathcal{I}}\left(\sqrt{1/n}X_t - \varphi_t\right) \geq 0\right)
\end{aligned}$$

which, for $\varepsilon = 1/\sqrt{n}$ corresponds to the previous definition of p_ε.

A trivial approach to the estimation of the probability p_ε is to draw an i.i.d. sample $X^{(1)}, \ldots, X^{(n)}$ from X and consider the MC estimator

$$p_\varepsilon \simeq \frac{1}{N} \sum_{i=1}^{N} 1_{A_\varepsilon}(X^{(i)})$$

where A_ε is the event

$$A_\varepsilon = \{x \in \mathcal{X} : \sup_{t \in \mathcal{I}} [\varepsilon x_t - \varphi_t] \geq 0\}.$$

However when $p_\varepsilon \to 0$ the number N of samples to obtain a reliable estimate should grow roughly as p_ε^{-1}. Estimation of very small probabilities (e.g. $p_\varepsilon \simeq 10^{-10}$ or smaller) becomes impossible or computationally heavy.

Importance Sampling (IS) is a popular technique devised to build unbiased estimators which do not suffer from the smallness of p_ε. This is achieved by changing the law of the process so that to favour the occurrence of the event A_ε and taking this change into account by reweighting the estimator according to the density of the original law with respect to the new law.

The efficiency of an IS-based algorithm depends on the choice of a proper change of measure to reduce the variance of the estimate. In particular, the simplest IS estimator obtained by *twisting* the original measure cannot be asymptotically efficient [2, 4] if the input is fBm. This was our main motivation to introduce the Bridge technique which we are going to explain in the next section.

Before describing our method, we recall a basic result which is the identification of the exponential rate at which p_ε goes to zero using the Large Deviation Theory for Gaussian processes [3] (for sake of simplicity, in the following we will denote by A the event A_ε for $\varepsilon = 1$, i.e. $A \equiv A_1$):

Theorem 2.1 *We have*

$$-\lim_{\varepsilon \to 0} \varepsilon^2 \log p_\varepsilon = \frac{1}{2} \inf_{\rho \in A} |\rho|^2_{\mathcal{H}} = \inf_{t \in \mathcal{I}} \frac{\varphi_t^2}{\Gamma_{t,t}} = |\rho^*|^2_{\mathcal{H}} \tag{5}$$

where $\rho^*_s = \varphi_{t^*} \Gamma_{st^*} / \Gamma_{t^* t^*}$ *is called a* most-likely path *and where* t^* *is a* most-likely time*, i.e. a time which satisfy*

$$\inf_{t \in \mathcal{I}} \frac{\varphi_t^2}{\Gamma_{t,t}} = \frac{\varphi_{t^*}^2}{\Gamma_{t^*,t^*}}.$$

The proof of this result can be found in many works related to IS for Gaussian processes, e.g. [1, 4, 8]. The most-likely path (mlp) and the related most-likely time (which for a general process X and a general barrier φ need not to be unique) play an important röle in the behaviour of the IS estimators. As we will see they are important concepts also in the analysis of the BMC.

The name *most-likely path* comes from the fact that heuristically, as $\varepsilon \to 0$ (i.e. in the large deviation regime) the majority of the samples of the process which attains the level φ are concentrated around ρ^*. When the mlp is unique, this heuristics can be confirmed by a large deviation calculation using Varadhan's lemma which shows that the distribution of εX, conditional on the event A, weakly converges to a Dirac mass on ρ^*.

3 The BMC Approach

The overflow probability p_ε can be rewritten as the expectation of a function of the *Bridge* Y of the Gaussian process X. The *Bridge* Y is the process obtained by conditioning X to reach a certain level at some prefixed time $\bar{t}$. By the properties of Gaussian processes Y is still a Gaussian process with a covariance which is a simple function of the covariance of the original process X. It is interesting to note that the computational effort to simulate Y is equal to that of X.

Fix $\bar{t} \in \mathcal{I}$ and consider the following centered Gaussian process:

$$Y_t = X_t - \psi_t X_{\bar{t}}, \qquad \text{where} \qquad \psi_t := \frac{\Gamma_{t\bar{t}}}{\Gamma_{\bar{t}\bar{t}}},$$

which has the distribution of X conditioned to be in 0 at time $\bar{t}$. The joint process (X, Y) is still Gaussian and the process Y is independent of $X_{\bar{t}}$ since

$$\mathrm{E}[X_{\bar{t}} Y_t] = \Gamma_{\bar{t}t} - \frac{\Gamma_{t\bar{t}}}{\Gamma_{\bar{t}\bar{t}}} \Gamma_{\bar{t}\bar{t}} = 0$$

and has covariance function $\widetilde{\Gamma}$ given by

$$\widetilde{\Gamma}_{ts} = \Gamma_{ts} - \frac{\Gamma_{t\bar{t}} \Gamma_{s\bar{t}}}{\Gamma_{\bar{t}\bar{t}}}.$$

The process Y lives in a smaller space $\widetilde{\mathcal{X}}$ (since one of the coordinates is zero), in analogy with X we can define its reproducing kernel Hilbert space $\widetilde{\mathcal{H}}$ and the associated scalar product $\langle \cdot, \cdot \rangle_{\widetilde{\mathcal{H}}}$.

Note that $X_t = Y_t + \psi_t X_{\bar{t}}$ for any $t \in \mathcal{I}$. Then we can express the probability P of any event B as

$$\mathrm{P}(B) = \int_{\mathcal{X}} 1_B(x) d\gamma(x) = \int_{\widetilde{\mathcal{X}}} d\widetilde{\gamma}(y) \int_{\mathbb{R}} d\gamma^{\bar{t}}(\overline{x}) 1_B(y + \psi \overline{x}) \tag{6}$$

where $\widetilde{\gamma}$ is the law of Y and $\gamma^{\bar{t}}$ the law of $X_{\bar{t}}$.

For the event A_ε we have

$$\begin{aligned}
p_\varepsilon &= \mathrm{P}(\sup_{t \in \mathcal{I}} [\varepsilon X_t - \varphi_t] \geq 0) = \mathrm{P}(\sup_{t \in \mathcal{I}} [\varepsilon Y_t + \varepsilon \psi_t X_{\bar{t}} - \varphi_t] \geq 0) \\
&= \mathrm{P}(\inf_{t \in \mathcal{I}} \psi_t^{-1} [\varphi_t - \varepsilon Y_t] \leq \varepsilon X_{\bar{t}}) = \mathrm{E}\left[P(\inf_{t \in \mathcal{I}} \psi_t^{-1} [\varphi_t - \varepsilon Y_t] \leq \varepsilon X_{\bar{t}} | Y) \right] \\
&= \mathrm{E}\left[\Phi\left(\frac{\overline{Y}_\varepsilon}{\varepsilon \sqrt{\Gamma_{\bar{t}\bar{t}}}} \right) \right]
\end{aligned}$$

where

$$\overline{Y}_\varepsilon := \inf_{t\in\mathcal{I}} \frac{\varphi_t - \varepsilon Y_t}{\psi_t} \qquad \text{and} \qquad \Phi(x) := \int_x^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy.$$

Given an i.i.d. sequence $\{Y^{(i)},\ i = 1,\dots,N\}$ distributed as Y, we introduce the *Bridge Monte-Carlo* (BMC) estimator for p_ε as follows:

$$\widehat{p}_\varepsilon^N := \frac{1}{N}\sum_{i=1}^N \Phi\left(\frac{\overline{Y}_\varepsilon^{(i)}}{\varepsilon\sqrt{\Gamma_{\bar t \bar t}}}\right) \qquad \text{where} \qquad \overline{Y}_\varepsilon^{(i)} := \inf_{t\in\mathcal{I}} \frac{\varphi_t - \varepsilon Y_t^{(i)}}{\psi_t}$$

which is unbiased ($E\widehat{p}_\varepsilon^N = p_\varepsilon$) and whose variance is $\mathrm{Var}(\widehat{p}_\varepsilon^N) = \sigma_\varepsilon^2/N$ where

$$\sigma_\varepsilon^2 := \mathrm{Var}\left[\Phi\left(\frac{\overline{Y}_\varepsilon}{\varepsilon\sqrt{\Gamma_{\bar t \bar t}}}\right)\right] = E\left[\Phi\left(\frac{\overline{Y}_\varepsilon}{\varepsilon\sqrt{\Gamma_{\bar t \bar t}}}\right)^2\right] - p_\varepsilon^2$$

The asymptotic behaviour of $\widehat{p}_\varepsilon$ as $\varepsilon \to 0$ is once again determined by the Large Deviation Principle for continuous Gaussian processes [3]. We proved [6] that

Lemma 3.1

$$-\lim_{\varepsilon\to 0} \varepsilon^2 \log E\left[\Phi\left(\frac{\overline{Y}_\varepsilon}{\varepsilon\sqrt{\Gamma_{\bar t \bar t}}}\right)^2\right] = 2\inf_{t\in\mathcal{I}} \frac{\varphi_t^2}{2(\Gamma_{tt} + \widetilde{\Gamma}_{tt})}. \tag{7}$$

which implies the following result on the asymptotic performance of the BMC:

Theorem 3.1 *The Bridge Estimator is* asymptotically efficient *[7] if and only if*

$$\inf_{t\in\mathcal{I}} \frac{\varphi_t^2}{2(\Gamma_{tt} + \widetilde{\Gamma}_{tt})} = \inf_{t\in\mathcal{I}} \frac{\varphi_t^2}{2\Gamma_{tt}}.$$

This is equivalent to require that $\bar t$ must be a most-likely time.

4 Conclusions

We proposed a novel approach (BMC - Bridge Monte-Carlo) for fast simulations of rare events in broadband communication systems which *does not* rely on Importance Sampling. The computational cost of BMC is comparable to that of IS.

To explain the novelty of BMC form a computational perspective we could note that if the basic MC method can be seen as a numerical scheme to perform integration in a large number of variables, then BMC is a *hybrid* method in that it performs *one* of these integrations exactly exploiting properties of Gaussian processes while the *remaining integrations* are still performed using a MC scheme. When it comes to rare events estimation it happens that in the full space of the process the characteristic function of the rare event has support on a region with small probability and this renders direct MC estimation ineffective. However BMC smooth out the function to be integrated allowing a more efficient estimation by the MC part.

It is important to point out that the principle underlying the BMC method can be applied to any Gaussian process and in a wide variety of network systems (e.g. tandem queues, schedulers, etc.) and could be generalised with more than one conditioning or with dynamic choice of the parameters.

The authors would like to acknowledge support from the Network of Excellence EuroNGI (Design and Engineering of the Next Generation Internet).

References

[1] R. Addie, P. Mannersalo, and I. Norros. Most Probable Paths and performace formulae for buffers with Gaussian input traffic. *Eur. Trans. on Telecommunications*, 13, 2002.

[2] Paolo Baldi and Barbara Pacchiarotti. Importance Sampling for the Ruin Problem for General Gaussian Processes. Technical report, 2004.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[4] A. B. Dieker and M. R. H. Mandjes. Fast simulation of overflow probabilities in a queue with Gaussian input. Technical Report PNA-E0404, CWI, the Netherlands, 2004.

[5] P. Dupuis and H. Wang. Importance Sampling, Large Deviations and differential games. Technical report, Lefschetz Center for Dynamical Systems, Brown University, 2002.

[6] S. Giordano, M. Gubinelli, and M. Pagano. Estimation of Overflow probabilities of a Gaussian process via its Bridge. Technical report, Mittag-Leffer Institute, 2005.

[7] P. Heidelberger. Fast Simulation of Rare Events in Queueing and Reliability Models. *Performance Eval. of Computer and Commun. Syst.*, 729:165–202, 1993.

[8] O'Connell, N. and Procissi, G. On the Build-Up of Large Queues in a Queueing Model with Fractional Brownian Motion Input. Technical Report HPL-BRIMS-98-18, BRIMS, HP Labs. Bristol (U.K.), August 1998.