Applications of Mathematics

David M. Gómez; Pablo Dartnell

Simple Monte Carlo integration with respect to Bernoulli convolutions

Applications of Mathematics, Vol. 57 (2012), No. 6, 617--626

Persistent URL: http://dml.cz/dmlcz/143006

57 (2012) APPLICATIONS OF MATHEMATICS No. 6, 617–626

SIMPLE MONTE CARLO INTEGRATION WITH RESPECT TO BERNOULLI CONVOLUTIONS*

David M. Gómez, Pablo Dartnell, Santiago de Chile

(Received December 21, 2010)

Abstract. We apply a Markov chain Monte Carlo method to approximate the integral of a continuous function with respect to the asymmetric Bernoulli convolution and, in particular, with respect to a binomial measure. This method—inspired by a cognitive model of memory decay—is extremely easy to implement, because it samples only Bernoulli random variables and combines them in a simple way so as to obtain a sequence of empirical measures converging almost surely to the Bernoulli convolution. We give explicit bounds for the bias and the standard deviation for this approximation, and present numerical simulations showing that it outperforms a general Monte Carlo method using the same number of Bernoulli random samples.

Keywords: MCMC, Bernoulli convolution, binomial measure, Monte Carlo integration, empirical measures

MSC 2010: 60G57, 65C05, 65D30

1. Introduction

Let $(X_n\colon\ n \geqslant 0)$ be an independent and identically distributed (i.i.d.) sequence of Bernoulli-distributed random variables in $\{-1, +1\}$ with $\mathbb{P}(X_0 = +1) = \alpha \in (0, 1)$. Given $\lambda \in (0, 1)$, the distribution $\nu_{\lambda,\alpha}$ of the random series $\sum_{n\geqslant 0} X_n \lambda^n$ is known as the asymmetric Bernoulli convolution. This distribution has been studied for more than 70 years, revealing important connections with several areas of mathematics, such as algebraic number theory, dynamical systems [9], and multifractal theory, since $\nu_{1/2,\alpha}$, also known as binomial measure, is considered a paradigmatic example of multifractal measure [8], [10]. It is a continuous distribution [6] whose support is contained in the interval $[-(1-\lambda)^{-1}, (1-\lambda)^{-1}]$.

* This research was partly supported by the Associative Research Program of CONICYT (CIE-05 and BASAL-CMM grants).

In [4], Calabrò and Corbo Esposito have proposed a number of quadrature formulae for binomial measures, although extending their work to general Bernoulli convolutions is not straightforward. Another possible approach to integration with respect to these distributions is given by Monte Carlo procedures. In general, these procedures provide random approximations of an integral $\int H \,\mathrm{d}F$, for a given function H and a distribution function F. The most general and conceptually simple method [7] consists in computing a sample $U_0, U_1, \ldots, U_n$ drawn from F and estimating

$$\int H \,\mathrm{d}F \approx I_n = \frac{1}{n+1} \sum_{k=0}^{n} H(U_k). \tag{1.1}$$

This approximation is guaranteed to converge almost surely to $I = \int H \,\mathrm{d}F$ as n grows because of the Law of Large Numbers, but its practical applicability is limited by the difficulty of sampling F. Some authors [11] proposed a way of sampling a piecewise linear approximation to $\nu_{\lambda,\alpha}$, allowing us to compute in a straightforward manner the Monte Carlo estimate. However, keeping this offline approximation and using it to compute random samples by—for instance—inverse transform sampling may involve important memory and computational demands.

Dovgoshey and colleagues [5] obtained a recursive expression for the moments $m_n = \int x^n \,\mathrm{d}C(x)$ of the Cantor function. As the Cantor function is a special case of Bernoulli convolution when $\lambda = \frac{1}{3}$ (up to a change of scale in the x-axis), a general integration method with respect to Bernoulli convolutions would provide a way to compute integrals of the kind $\int H(x) \,\mathrm{d}C(x)$ for some broader class of functions H.

In this note we present a Markov chain Monte Carlo procedure (abbreviated as MCMC, see [1] for an introduction) for integrating with respect to $\nu_{\lambda,\alpha}$, namely generating a sequence (U_k) of non-independent random variables whose empirical distributions converge almost surely to $\nu_{\lambda,\alpha}$. This method turns out to have rather low requirements in terms of memory and computational power.

The motivation for developing this method comes from a cognitive model of memory: think of an agent who observes a sequence of i.i.d. random outcomes $X_0, X_1, \ldots, X_n$, one at a time. It is well known that, in statistical terms, in order to estimate $\mathbb{E}(X_0)$ the best that this agent can do is to compute the sample mean $(n+1)^{-1}(X_0 + \ldots + X_n)$. Nonetheless, a more realistic model for a living agent is to take into account the effect of memory decay, so that the most recent outcomes are weighted more than the distant ones. This topic has raised a wide debate in psychological literature, and one of the main ways of incorporating memory strength decay is by what is called exponential decay [12], which implies that the

agent estimates $\mathbb{E}(X_0)$ by means of a weighted average

$$\frac{X_0\lambda^n + X_1\lambda^{n-1} + \ldots + X_{n-1}\lambda + X_n}{\lambda^n + \lambda^{n-1} + \ldots + \lambda + 1}, \tag{1.2}$$

where $\lambda \in (0,1)$ is a parameter idiosyncratic to the agent, called *rate of forgetting*. Since the denominator in expression (1.2) converges to a constant, the asymptotic behavior of this weighted average is determined by the numerator $U_n = \sum_{k=0}^{n} X_{n-k}\lambda^k$. The properties of this random sum are the ones leading to the method presented in the next section.

2. The method

Our procedure simplifies the computation of I_n by using a non-independent, easy-to-compute sample $U_0, U_1, \ldots, U_n$ converging in law to $\nu_{\lambda,\alpha}$. Specifically, based on a random sample $X_0, X_1, \ldots, X_n$ of a Bernoulli distribution in $\{-1,+1\}$ we build the sequence $(U_i\colon\ i = 0,\ldots,n)$ by means of the recursive procedure

$$\begin{cases} U_0 = X_0, \\ U_i = X_i + \lambda \cdot U_{i-1} \quad (i = 1,\ldots,n). \end{cases} \tag{2.1}$$

The sequence (U_n) can be thought of as a Markov chain in the countable state space

$$A = \left\{ \sum_{k=0}^{n} a_k\lambda^k\colon\ n \geqslant 0,\ a_0,\ldots,a_n \in \{-1,+1\} \right\},$$

with transitions $a \to f_+(a) = 1 + \lambda a$ and $a \to f_-(a) = -1 + \lambda a$ and associated probabilities α and $1-\alpha$. This approach can be seen as sampling the iterated function system [2] associated with $\nu_{\lambda,\alpha}$ given by f_+, f_-, and α.

Since (X_k) is i.i.d., U_n has the same law as $S_n = \sum_{k=0}^{n} X_k\lambda^k$. The space (S_n) converges almost surely to $S_\infty = \sum_{k\geqslant 0} X_k\lambda^k$, thus it converges in law to $\nu_{\lambda,\alpha}$ and the same can then be said of (U_n). However, we observe that (U_n) converges almost nowhere. This follows from the fact that for any $d \geqslant 1$, with probability one $X_m = X_{m+1} = \ldots = X_{m+d-1} = +1$ for infinitely many values of m. For every such m we have that

$$\begin{aligned} U_{m+d-1} &= \sum_{k=d}^{m+d-1} X_{m+d-1-k}\lambda^k + \sum_{k=0}^{d-1}\lambda^k = \lambda^d \sum_{k=0}^{m-1} X_{m-1-k}\lambda^k + \frac{1-\lambda^d}{1-\lambda} \\ &\geqslant \frac{-\lambda^d}{1-\lambda} + \frac{1-\lambda^d}{1-\lambda} = \frac{1-2\lambda^d}{1-\lambda}, \end{aligned}$$

showing that $(1-\lambda)^{-1}$ is an accumulation point of (U_n). Applying the same reasoning to the sequence $(\tilde{X}_k = -X_k)$ we conclude that (U_n) also accumulates on $-(1-\lambda)^{-1}$, being then non-convergent.

We thus build our approximation I_n as given by Equation (1.1). The following lemma is a crucial step for stating the convergence of I_n.

Lemma 2.1. *Let F_n be the empirical distribution function of $U_0, \ldots, U_n$:*

$$F_n(y) = \frac{1}{n+1} \sum_{k=0}^{n} \mathbf{I}\{U_k \leqslant y\} \quad \textit{for any } y \in \mathbb{R}.$$

Then $I_n = \int H \, \mathrm{d}F_n$.

P r o o f. For $k = 0, \ldots, n$, let A_k be the (finite) atom set of U_k, and denote $B_k = A_0 \cup \ldots \cup A_k$. We can write $H(U_k)$ as $\sum\limits_{y \in B_n} H(y)\mathbf{I}\{U_k = y\}$, which implies

$$\begin{aligned} I_n &= \frac{1}{n+1} \sum_{k=0}^{n} \sum_{y \in B_n} H(y)\mathbf{I}\{U_k = y\} = \sum_{y \in B_n} \frac{H(y)}{n+1} \sum_{k=0}^{n} \mathrm{ind}\{U_k = y\} \\ &= \sum_{y \in B_n} H(y)(F_n(y) - F_n(y^-)) = \int H \, \mathrm{d}F_n. \end{aligned}$$

□

Theorem 2.1. *Let $H\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. With probability one,*

$$\lim_{n \to \infty} I_n = \int H \, \mathrm{d}\nu_{\lambda,\alpha}.$$

P r o o f. Denote by F the distribution function of $\nu_{\lambda,\alpha}$. Given our previous lemma and the continuity of F, we just need to show that for any $y \in \mathbb{R}$

$$\lim_{n \to \infty} F_n(y) = F(y) \quad \text{almost surely.} \tag{2.2}$$

In order to prove this, we employ a result proved in [3], providing almost sure versions of a wealth of weak limit theorems. Specifically, we are concerned with an almost sure version of the convergence in law of (U_n) to $\nu_{\lambda,\alpha}$. We observe that for $k < n$, U_n can be written as $\lambda^{n-k}U_k + U'_{n,k}$, where $U'_{n,k}$ depends only on $X_{k+1}, \ldots, X_n$. Hence, denoting $c_n = \lambda^{-n}$, the dependence of U_n on the initial variables $X_0, \ldots, X_k$ is given by the term $\lambda^{n-k}U_k$, whose absolute value is bounded by $(1-\lambda)^{-1}c_k/c_n$.

The sequence (c_n) fulfils all the hypotheses required in Theorem 1 in [3], implying that for any $y \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{d_0 + \ldots + d_n} \sum_{k=0}^{n} d_k \mathbf{I}\{U_k < y\} = F(y) \quad \text{almost surely,}$$

where the weights d_k are given by $d_k = \log(c_{k+1}/c_k)$. We note, however, that in our case these weights are constant: $d_k = |\log \lambda|$, meaning that

$$\lim_{n\to\infty} F_n(y^-) = F(y) \quad \text{almost surely.}$$

From $F_n(y) - F_n(y^-) = (n+1)^{-1} \sum_{k=0}^{n} \mathbf{I}\{U_k = y\}$ it follows that $F_n(y)$ and $F_n(y^-)$ are equal for all but an at most countable number of $y \in \mathbb{R}$, implying Equation (2.2) for this class of y and consequently for all $y \in \mathbb{R}$. □

Given the fact that we do not sample directly $\nu_{\lambda,\alpha}$, I_n turns out to be biased. The following result gives an explicit bound for the bias when H is minimally regular.

Lemma 2.2. *Let H be a function of class $\mathcal{C}^1$. Then the bias of I_n is $O(n^{-1})$.*

P r o o f. We first notice that for all $k \geqslant 0$, $\mathbb{E}(H(U_k)) = \mathbb{E}(H(S_k))$, and that $\int H \, \mathrm{d}\nu_{\lambda,\alpha} = \mathbb{E}(H(S_\infty))$. Then

$$\begin{aligned}
\left| \int H \, \mathrm{d}\nu_{\lambda,\alpha} - \mathbb{E}(I_n) \right| &\leqslant \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{E}(|H(S_\infty) - H(S_k)|) \\
&\leqslant \frac{\|H'\|_\infty}{n+1} \sum_{k=0}^{n} \mathbb{E}(|S_\infty - S_k|) \leqslant \frac{\|H'\|_\infty}{n+1} \sum_{k=0}^{n} \sum_{j \geqslant k+1} \lambda^j \\
&\leqslant \frac{\lambda \|H'\|_\infty}{(1-\lambda)^2 (n+1)},
\end{aligned}$$

which concludes the proof (we consider $\|H'\|_\infty$ to be the maximum computed over the bounded interval $[-(1-\lambda)^{-1}, (1-\lambda)^{-1}]$, which contains the support of $\nu_{\lambda,\alpha}$). □

We remark that a better speed of convergence for the bias can be obtained with slight changes of I_n, and essentially the same proof: for instance, given $\delta \in (0,1)$, the bias of $(n - \lceil \delta n \rceil + 1)^{-1} \sum_{k=\lceil \delta n \rceil}^{n} H(U_k)$ is $O(n^{-1}\lambda^{\delta n})$ for any function H of class $\mathcal{C}^1$.

3. Standard deviation

In the general Monte Carlo approach where the samples are independent, it is known that the standard deviation of I_n is proportional to $n^{-1/2}$. In our case, by observing that for any $j > i$, the quantity $\Delta_{ij} = U_j - \lambda^{j-i}U_i$ depends only on $X_{i+1}, \dots, X_j$, we conclude that $\operatorname{Cov}(U_i, U_j) = \lambda^{j-i}\operatorname{Var}(U_i) > 0$ and thus the decay of the variance of I_n may, in principle, be slower. However, our last result will show that if H has one more continuous derivative then the decay of the standard deviation of I_n is still $O(n^{-1/2})$.

Theorem 3.1. *Let H be a function of class $\mathcal{C}^2$. Then the standard deviation of I_n is $O(n^{-1/2})$.*

Proof. By using the Taylor expansion of H around $y = \Delta_{ij}$, we can write for $j > i$

$$H(U_j) = H(\Delta_{ij}) + H'(\Delta_{ij})\lambda^{j-i}U_i + \varepsilon_{ij}.$$

Recalling that U_i and Δ_{ij} are independent, we write $C_{ij} = \operatorname{Cov}(H(U_i), H(U_j))$ as

$$\begin{aligned} C_{ij} &= \lambda^{j-i}\operatorname{Cov}(H(U_i), H'(\Delta_{ij})U_i) + \operatorname{Cov}(H(U_i), \varepsilon_{ij}) \\ &= \lambda^{j-i}\mathbb{E}(H'(\Delta_{ij}))\operatorname{Cov}(H(U_i), U_i) + \operatorname{Cov}(H(U_i), \varepsilon_{ij}). \end{aligned}$$

Given that the magnitude of the residual ε_{ij} is bounded by $\frac{1}{2}U_i^2\|H''\|_\infty\lambda^{2(j-i)}$, we can bound also this covariance as follows:

$$\begin{aligned} |C_{ij}| &\leqslant \operatorname{Var}(H(U_i))^{1/2}(\lambda^{j-i}|\mathbb{E}(H'(\Delta_{ij}))|\operatorname{Var}(U_i)^{1/2} + \operatorname{Var}(\varepsilon_{ij})^{1/2}) \\ &\leqslant \operatorname{Var}(H(U_i))^{1/2}(\lambda^{j-i}\|H'\|_\infty\operatorname{Var}(U_i)^{1/2} + \mathbb{E}(\varepsilon_{ij}^2)^{1/2}) \\ &\leqslant \lambda^{j-i}\underbrace{\operatorname{Var}(H(U_i))^{1/2}\Big(\|H'\|_\infty\operatorname{Var}(U_i)^{1/2} + \frac{1}{2}\mathbb{E}(U_i^4)^{1/2}\|H''\|_\infty\Big)}_{K_i}. \end{aligned}$$

Thus,

$$\sum_{i=0}^{n}\sum_{j=i+1}^{n}|\operatorname{Cov}(H(U_i), H(U_j))| \leqslant \sum_{i=0}^{n}K_i\sum_{j=i+1}^{n}\lambda^{j-i} \leqslant \frac{\lambda}{1-\lambda}\sum_{i=0}^{n}K_i.$$

This gives, in turn, the following bound for $\operatorname{Var}(I_n)$:

$$\operatorname{Var}(I_n) \leqslant \frac{1}{(n+1)^2}\left[\sum_{k=0}^{n}\operatorname{Var}(H(U_k)) + \frac{2\lambda}{1-\lambda}\sum_{i=0}^{n}K_i\right].$$

Finally, the convergence in law of (U_i) to $\nu_{\lambda,\alpha}$ allows us to use the following asymptotically equivalent bound

$$\begin{aligned}(3.1)\quad &\frac{\operatorname{Var}(H(\nu_{\lambda,\alpha}))}{n+1}\\ &+\frac{2\lambda\cdot\operatorname{Var}(H(\nu_{\lambda,\alpha}))^{1/2}}{(1-\lambda)(n+1)}\Big(\|H'\|_\infty\operatorname{Var}(\nu_{\lambda,\alpha})^{1/2}+\frac{1}{2}\mathbb{E}(\nu_{\lambda,\alpha}^4)^{1/2}\|H''\|_\infty\Big).\end{aligned}$$

□

Equation (3.1) allows for a quick comparison with a general Monte Carlo method, whose variance is given by $(n+1)^{-1}\operatorname{Var}(H(\nu_{\lambda,\alpha}))$ when using $n+1$ independent samples of $\nu_{\lambda,\alpha}$. The overhead given by the second term in (3.1) can be seen as a penalty for building an estimate from sampling a much simpler distribution. However, in the next section we show that this overhead can in practice be smaller than the savings represented by avoiding the sampling of $\nu_{\lambda,\alpha}$.

4. Numerical assessment

In this last section, we show a numerical comparison between our proposed method and a general Monte Carlo approach in the case of the binomial measure $\mu_\alpha = \nu_{1/2,\alpha}$. The latter will be implemented by using independent samples of an approximation to μ_α.

We selected two functions from [4], namely[1] $H_1(y) = \sin(\pi y) + \frac{1}{4}\cos(8\pi y)$ and $H_2(y) = \sqrt{|y|}$. We deliberately chose H_2 not of class $\mathcal{C}^1$ in order to evaluate convergence when our bounds are uninformative.

For implementing the general Monte Carlo procedure, we remind the reader that $S_n = \sum_{k=0}^{n} X_k 2^{-k}$ converges in law to μ_α, and moreover, the distance between μ_α and the distribution of S_n in uniform norm is $O(M^n)$, with $M = \max\{\alpha, 1-\alpha\}$ (see Corollary 2.4 in [11]). Hence, a way of simulating approximately μ_α is sampling S_N for big N. We took this approach rather than computing an offline approximation to the distribution of μ_α, because it allows us to make a fair comparison between methods based on sampling only Bernoulli random variables, and with little memory consumption (as opposed to keeping a detailed offline approximation to μ_α, such as the one proposed in [11]). In fact, we chose the parameters for both methods so that the number of Bernoulli samples is equated. That is, whereas for our proposed

[1] Actually, H_2 is an adaptation of the function $\sqrt{y}$ these authors utilized. We did so because they assumed the binomial measure to be supported in the interval $[0,1]$, whereas we work in $[-2,2]$.

method we consider $(n+1)$ random samples from a Bernoulli distribution used to build I_n, we define $n' = (n+1)/(N+1) - 1$ and take random samples $V_0, V_1, \ldots, V_{n'}$ of S_N for computing

$$I_{n'}^{MC} = \frac{1}{n'+1} \sum_{k=0}^{n'} H(V_k),$$

which is the approximation given by the general Monte Carlo approach. In what follows, we fix $N = 19$ and take several values for n. In order to estimate the variance of I_n and $I_{n'}^{MC}$, we computed 10000 simulations of the associated sampling processes. All computations were done using Octave version 3.2.3 (`http://www.octave.org/`). The actual values for $\int H_i \, \mathrm{d}\nu_{\lambda,\alpha}$, $i = 1, 2$, were approximated by substituting $\nu_{\lambda,\alpha}$ by an exact, offline computation of the distribution of S_{24}. This choice is due to the convergence in law of (S_n) to μ_α, and to the fact that the difference between $\int H \, \mathrm{d}S_n$ for $n = 23, 24$ is much smaller—by several orders of magnitude—than the approximation errors of the two random estimates I_n, $I_{n'}^{MC}$ that we compare in this section. However, other ways of computing the value of these integrals may be considered (such as the adaptive procedure presented in [4]), giving similar numbers.

In Fig. 1 A we show the average relative error of both the methods as a function of the number of Bernoulli random samples involved, whereas Fig. 1 B depicts their standard deviations. Notice that in average I_n is at least as good as the general Monte Carlo approximation $I_{n'}^{MC}$, and in some cases it gives much better results. Fig. 1 B shows that the standard deviation of the approximation I_n for the integral of H_1 decays as $n^{-1/2}$, just as it is the case for $I_{n'}^{MC}$. We omit the corresponding graphic for H_2, because the situation is qualitatively the same (despite the fact that H_2 has a discontinuous derivative at $y = 0$).

Acknowledgment. The authors are grateful to Alejandro Maass, Peter Grabner, and two anonymous referees for invaluable comments on previous versions of this manuscript.

References

[1] *C. Andrieu, N. De Freitas, A. Doucet, M. I. Jordan*: An introduction to MCMC for machine learning. Mach. Learn. *50* (2003), 5–43.

[2] *M. F. Barnsley, S. Demko*: Iterated function systems and the global construction of fractals. Proc. R. Soc. Lond., Ser. A *399* (1985), 243–275.

[3] *I. Berkes, E. Csáki*: A universal result in almost sure central limit theory. Stoch. Proc. Appl. *94* (2001), 105–134.

[4] *F. Calabrò, A. Corbo Esposito*: An efficient and reliable quadrature algorithm for integration with respect to binomial measures. BIT *48* (2008), 473–491.

[5] *O. Dovgoshey, O. Martio, V. Ryazanov, M. Vuorinen*: The Cantor function. Expo. Math. *24* (2006), 1–37.

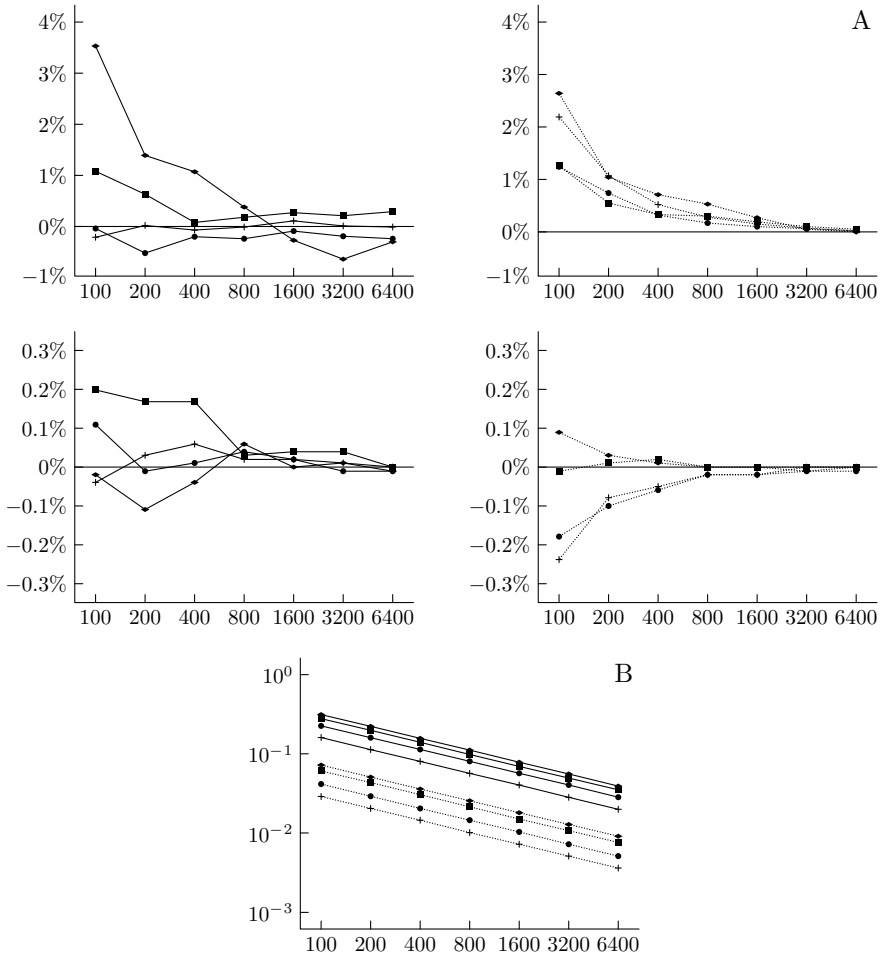

Figure 1. A. Average relative errors of the general Monte Carlo procedure ($I_{n'}^{MC}$, left column) and the one proposed in this work (I_n, right column), for the functions H_1 (top row) and H_2 (bottom row) defined in the main text.
B. Standard deviations of I_n (dotted lines) and $I_{n'}^{MC}$ (solid lines) for H_1. Symbols represent different values of α: crosses (0.1), circles (0.2), squares (0.3), diamonds (0.4). Horizontal axes (log-scale) depict simulation length, as measured by the number of Bernoulli samples used in each case.

[6] *B. Jessen, A. Wintner*: Distribution functions and the Riemann zeta function. Trans. Am. Math. Soc. *38* (1935), 48–88.

[7] *M. H. Kalos, P. A. Whitlock*: Monte Carlo Methods. Vol. I: Basics. Wiley, New York, 1986.

[8] *B. B. Mandelbrot, L. Calvet, A. Fisher*: A multifractal model of asset returns. Cowles Foundation Discussion Papers: 1164. 1997, Retrieved from `http://users.math.yale.edu/users/mandelbrot/web_pdfs/Cowles1164.pdf` (last access August 21, 2012).

[9] *Y. Peres, W. Schlag, B. Solomyak*: Sixty years of Bernoulli convolutions. Fractal Geometry and Stochastics II. Proceedings of the 2nd conference Greifswald/Koserow, Germany, August 28–September 2, 1998. Eds. C. Bandt at al. Prog. Probab. *46* (2000), 39–65.
[10] *R. H. Riedi*: Introduction to multifractals. Techn. Rep. Rice Univ., October 26, 1999, Retrieved from `http://rudolf.riedi.home.hefr.ch/Publ/PDF/intro.pdf` (last access August 21, 2012).
[11] *R. S. Strichartz, A. Taylor, T. Zhang*: Densities of self-similar measures on the line. Exp. Math. *4* (1995), 101–128.
[12] *K. G. White*: Forgetting functions. Animal Learning & Behavior *29* (2001), 193–207.

Author's address: *D. M. Gómez* (corresponding author), Department of Mathematical Engineering, University of Chile, Avenida Blanco Encalada 2120 piso 5, Santiago, Chile; currently at Cognitive Neuroscience Sector, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy, e-mail: `dgomez@sissa.it`; *P. Dartnell*, Department of Mathematical Engineering, University of Chile, Avenida Blanco Encalada 2120 piso 5, Santiago, Chile, and Centre for Mathematical Modeling, University of Chile, Avenida Blanco Encalada 2120 piso 7, Santiago, Chile, and Centre for Advanced Research in Education, University of Chile, Periodista José Carrasco Tapia 75, Santiago, Chile, e-mail: `dartnell@dim.uchile.cl`.