Monte Carlo and Quasi-Monte Carlo Density Estimation via Conditioning

Pierre L'Ecuyer
Département d'Informatique et de Recherche Opérationnelle, Pavillon Aisenstadt, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7, lecuyer@iro.umontreal.ca

Florian Puchhammer
Basque Center for Applied Mathematics, Alameda de Mazarredo 14, 48009 Bilbao, Basque Country, Spain; and Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, fpuchhammer@bcamath.org

Amal Ben Abdellah
Département d'Informatique et de Recherche Opérationnelle, Pavillon Aisenstadt, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7, amal.ben.abdellah@umontreal.ca

Online Supplement

This Supplement contains additional examples and details for which there was not enough space in the main paper.

In Appendix A, we show with simple examples how one can prove that the HK variation of the CDE is bounded uniformly over the interval $[a, b]$ of interest. When this can be done, it proves that the MISE for the CDE with good RQMC points converges as $\mathcal{O}(n^{-2+\epsilon})$.

Appendix B provides additional examples showing how CDEs can be constructed, sometimes in non-trivial ways that are adapted to the problem at hand. In Section B.1, we consider the very simple example of a sum of independent normal random variables, for which the density is known, and the purpose is to see how each estimator behaves as a function of the dimension (the number of summands) and of the relative variance of the one we hide. In Section B.2, we consider a six-dimensional example taken from Schields and Zhang (2016). In Section B.3, we consider a multicomponent system in which each component fails at a certain random time, and we want to estimate the density of the failure time of the system. In Section B.4, we explain briefly how accurate density estimation is useful for computing a confidence interval on a quantile or on the expected shortfall.

Section C provides additional figures for examples in the paper.

Appendix A: Proving bounded HK variation for the CDE: some simple illustrations

Here we show how the HK variation of $g \equiv \tilde{g}'(x, \cdot)$ can be bounded uniformly in $x \in [a, b]$ in our CDE setting, for Examples 1 to 3 of the paper.

EXAMPLE 1. Consider a sum of random variables as in Example 1, with $\mathcal{G} = \mathcal{G}_{-k}$ summarized by the single real number S_{-k}. We have $F(x \mid \mathcal{G}) = F_k(x - S_{-k})$ and $f(x \mid \mathcal{G}) = f_k(x - S_{-k})$. Without loss of generality, let $k = d$. Suppose that each Y_j is generated by inversion from $U_j \sim \mathcal{U}(0,1)$, so $Y_j = F_j^{-1}(U_j)$ and $S_{-d} = F_1^{-1}(U_1) + \cdots + F_s^{-1}(U_s)$ with $s = d-1$. This gives $\tilde{g}(x, \mathbf{U}) = F_d(x - S_{-d}) = F_d(x - F_1^{-1}(U_1) - \cdots - F_s^{-1}(U_s))$ and $\tilde{g}'(x, \mathbf{U}) = f_d(x - S_{-d}) = f_d(x - F_1^{-1}(U_1) - \cdots - F_s^{-1}(U_s))$. The partial derivatives of this last function are

$$\tilde{g}'_{\mathfrak{v}}(x, \mathbf{U}_{\mathfrak{v}}, \mathbf{1}) = f_d^{(|\mathfrak{v}|)}(x - S_{-d}) \prod_{j \in \mathfrak{v}} \frac{\partial (F_j^{-1}(U_j))}{\partial U_j}.$$

So the functions F_j^{-1} must be differentiable over $(0,1)$ for $j = 1, \ldots, d-1$, the density f_d must be s times differentiable, and the integral of $|\tilde{g}'_{\mathfrak{v}}(x, \mathbf{u}_{\mathfrak{v}}, \mathbf{1})|$ with respect to $\mathbf{u}_{\mathfrak{v}}$ must be bounded uniformly in $x \in [a,b]$. Under these conditions, the HK variation is bounded uniformly in x over $[a,b]$.

For Example 2, with $\mathcal{G} = \mathcal{G}_{-2}$ and $Y_1 = U_1$, we have $\tilde{g}'(x, \mathbf{u}) = \tilde{g}'(x, U_1) = \mathbb{I}[U_1 \le x \le \epsilon + U_1]/\epsilon = \mathbb{I}[x - \epsilon \le U_1 \le x]/\epsilon$. This function is not continuous, but its HK variation (not given by (11) in this case) is $2/\epsilon < \infty$, because it is piecewise constant with only two jumps, each one of size $1/\epsilon$. Thus, the HK variation is unbounded when $\epsilon \to 0$, but it is finite for any fixed ϵ, independently of x. The behavior with $\mathcal{G} = \mathcal{G}_{-1}$ is similar and the HK variation is 2 in that case, which is much better.

For Example 3, if $\mathcal{G} = \mathcal{G}_{-2}$, we have $Y_1 = \sigma_1 \Phi^{-1}(U_1)$ where $U_1 \sim \mathcal{U}(0,1)$. Then, $F(x \mid \mathcal{G}_{-2}) = F_2(x - Y_1) = \Phi((x - Y_1)/\sigma_2)$ and $f(x \mid \mathcal{G}_{-2}) = \phi((x - \sigma_1 \Phi^{-1}(U_1))/\sigma_2)/\sigma_2 = \tilde{g}'(x, U_1)$. Taking the derivative with respect to u and noting that $\mathrm{d}\Phi^{-1}(u)/\mathrm{d}u = 1/(\phi(\Phi^{-1}(u)))$ yields

$$\tilde{g}'_{\mathfrak{v}}(x, u) = \frac{\phi'((x - \sigma_1 \Phi^{-1}(u))/\sigma_2)\, \sigma_1}{\sigma_2^2 \phi(\Phi^{-1}(u))}$$

for $\mathfrak{v} = \{1\} = \mathcal{S}$ (the only subset in this case). Integrating this with respect to u by making the change of variable $z = \Phi^{-1}(u)$ gives

$$\int_0^1 \tilde{g}'_{\mathfrak{v}}(x, u) \mathrm{d}u = \frac{\sigma_1}{\sigma_2^2} \int_{-\infty}^{\infty} |\phi'((x - \sigma_1 z)/\sigma_2)| \, \mathrm{d}z,$$

which is bounded uniformly in x, because $|\phi'(\cdot)|$ is bounded by $\phi(\cdot)$ multiplied by the absolute value of a polynomial of degree 1. So the HK variation is bounded uniformly in x.

Appendix B: Additional examples

B.1. A sum of normals

We start with a very simple example in which the density f is known beforehand, so there is no real need to estimate it, but this type of example is very convenient for testing the performance of various density estimators. Let $Z_1, \ldots, Z_d$ be independent standard normal random variables, i.e., with mean 0 and variance 1, and define

$$X = (a_1 Z_1 + \cdots + a_d Z_d)/\sigma, \qquad \text{where} \qquad \sigma^2 = a_1^2 + \cdots + a_d^2.$$

Then X is also standard normal, with density $f(x) = \phi(x) \stackrel{\text{def}}{=} \exp(-x^2/2)/\sqrt{2\pi}$ and cdf $\mathbb{P}[X \le x] = \Phi(x)$ for $x \in \mathbb{R}$. The term $a_j Z_j$ in the sum has variance a_j^2. We pretend we do not know this and we estimate $f(x)$ over the interval $[-2, 2]$, which contains slightly more than 95% of the density. We also tried larger intervals, such as $[-5, 5]$, and the IVs for the CDE were almost the same.

To construct the CDE, we define $\mathcal{G}_{-k}$ as in Example 1, for any $k = 1, \ldots, d$. That is, we hide Z_k and estimate the cdf by

$$F(x \mid \mathcal{G}_{-k}) = \mathbb{P}\left[a_k Z_k \le x\sigma - \sum_{j=1,\, j\neq k}^{d} a_j Z_j \,\middle|\, \mathcal{G}_{-k} \right] = \Phi\left(\frac{x\sigma}{a_k} - \frac{1}{a_k} \sum_{j=1,\, j\neq k}^{d} a_j Z_j \right).$$

The CDE becomes

$$f(x \mid \mathcal{G}_{-k}) = \phi\left(\frac{x\sigma}{a_k} - \frac{1}{a_k} \sum_{j=1,\, j\neq k}^{d} a_j Z_j \right) \frac{\sigma}{a_k} = \phi\left(\frac{x\sigma}{a_k} - \frac{1}{a_k} \sum_{j=1,\, j\neq k}^{d} a_j \Phi^{-1}(U_j) \right) \frac{\sigma}{a_k} \stackrel{\text{def}}{=} \tilde{g}'(x, \mathbf{U})$$

for $x \in \mathbb{R}$, where $\mathbf{U} = (U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_d)$, $Z_j = \Phi^{-1}(U_j)$, and the U_j are independent $\mathcal{U}(0,1)$ random variables. Assumption 1 is easily verified, so this CDE is unbiased.

For CMC+MC (independent sampling), we get an exact formula for the variance of the CDE from Example 3, by taking in that example $Y_2 = a_k Z_k/\sigma$ and $Y_1 = X - Y_2$, whose variances are $\sigma_2^2 = (a_k/\sigma)^2$ and $\sigma_1^2 = 1 - \sigma_2^2$, and plugging these values into (6). With the same argument as in the second part of Example 1 in the supplement, we can show that $V_{\text{HK}}(\tilde{g}'(x,\cdot)) < \infty$, uniformly in x over any bounded interval $[a, b]$, so Proposition 3 applies. We expect to observe this empirically.

For the GLRDE, with $Y_j = Z_j a_j/\sigma \sim \mathcal{N}(0, a_j^2/\sigma^2)$, we obtain $\partial(\log f_j(y_j))/\partial y_j = -y_j\sigma^2/a_j^2$, $h_j(y_j) = 1$, $h_{jj}(y_j) = 0$, and then $\Psi_j = -Y_j\sigma^2/a_j^2 = -Z_j\sigma/a_j$. Note that we could also replace Y_j by Z_j and f_j by ϕ_j (the standard normal density), which would give $\partial(\log\phi_j(z_j))/\partial z_j = -z_j$, $h_j(z_j) = a_j/\sigma$, $h_{jj}(y_j) = 0$, and again $\Psi_j = -Z_j\sigma/a_j$.

In our first experiment, we take $a_j = 1$ for all j, and $k = d$. By symmetry, the true IV is the same for any other k. Table 1 reports the estimated rate $\hat{\nu}$ and the estimated value of e19 $= -\log_2(\text{IV})$ for $n = 2^{19}$, for various values of d and sampling methods. The rows marked CDE-1 give the results for $k = d$, while those labeled CDE-Avg are for a convex combination (7) with equal weights $\beta_\ell = 1/d$ for all $\ell = k - 1$, after computing the CDE for each k from the same simulations.

For MC, the rates $\hat{\nu}$ agree with the (known) exact asymptotic rates of $\nu = 1$ for the CDE and GLRDE, and $\nu = 0.8$ for the KDE. By looking at e19, we see that the MISE with MC is much smaller for the CDE than for the GLRDE and KDE, for example for $d = 2$ by a factor of about 32 for CDE-1 and about 70 for CDE-avg. For $d = 20$, the gains are more modest. RQMC methods provide huge improvements for small d with the CDE. We observe rates $\hat{\nu}$ larger than 2 for $d = 2$ and 3. These rates also hold for larger d asymptotically, but they take longer to kick in, so we would need to have much larger values of n to observe them. By looking at the exponents e19, we see that for $d = 3$, for example, the MISE goes from 2^{-17} for the GLRDE and KDE to about 2^{-42} for CDE-avg with Sobol' points with LMS. This is a MISE reduction by a factor of about $2^{25} \approx 33$ millions! The large values of $\hat{\nu}$ imply of course that this factor is smaller for smaller n. When d is large, such as $d = 20$, RQMC brings only a small gain. The values of $\hat{\nu}$ are sometimes noisy. For the GLRDE with Lat+s and $d = 5$, for example, the large $\hat{\nu} = 1.45$ comes from the fact that the IV for $n = 2^{14}$ (not shown) is unusually large (an outlier). Looking at e19 gives a more robust assessment of the performance. The GLRDE performs better than the KDE under RQMC for small d, but is not competitive with the CDE. Under MC, the GLRDE is slightly worse than the KDE.

Table 1 Values of $\hat{\nu}$ and e19 for a CDE, a convex combination of CDEs, a GLRDE, and a KDE, for a sum of $d = k$ normals with $a_j = 1$, over $[-2, 2]$.

		$\hat{\nu}$					e19				
		$d=2$	$d=3$	$d=5$	$d=10$	$d=20$	$d=2$	$d=3$	$d=5$	$d=10$	$d=20$
CDE-1	MC	0.99	0.98	1.02	1.00	1.02	22.1	21.4	20.8	19.8	19.2
	Lat+s	2.83	2.00	1.85	1.40	1.04	52.3	39.8	32.1	23.6	19.7
	Lat+s+b	2.69	2.11	1.69	1.14	1.05	50.5	41.5	31.1	21.8	20.0
	Sob+LMS	2.62	2.10	1.81	1.04	1.04	49.3	40.7	31.1	21.3	19.7
CDE-avg	MC	1.06	0.92	1.03	1.01	1.01	23.4	22.1	21.6	20.6	19.8
	Lat+s	2.79	1.84	1.33	1.19	1.05	53.3	39.8	32.2	23.0	20.6
	Lat+s+b	2.65	1.90	1.71	1.05	1.08	51.6	41.4	32.3	23.4	21.3
	Sob+LMS	2.60	2.10	1.92	1.02	1.03	49.8	42.0	33.0	22.7	20.5
GLRDE	MC	0.98	0.95	1.03	1.05	1.00	17.0	16.1	15.9	14.9	14.1
	Lat+s	1.51	1.56	1.45	0.94	1.06	28.2	24.9	22.1	17.8	17.2
	Lat+s+b	1.49	1.41	1.05	1.06	1.04	27.3	23.9	20.4	18.8	17.6
	Sob+LMS	1.49	1.33	1.15	0.99	1.16	27.5	24.0	21.0	18.3	17.4
KDE	MC	0.79	0.80	0.76	0.75	0.77	17.0	17.0	16.9	16.9	17.0
	Lat+s	1.08	1.39	0.92	0.97	0.76	25.1	22.4	19.4	18.2	17.4
	Lat+s+b	1.23	0.94	0.72	0.73	0.74	24.1	20.1	18.1	17.3	17.2
	Sob+LMS	1.18	0.98	0.83	0.74	0.77	24.4	20.8	17.9	17.2	17.1

Table 2 Values of $\hat{\nu}$ and e19 with a CDE for selected choices of $\mathcal{G}_{-k}$, for a linear combination of $d = 11$ normals with $a_j^2 = 2^{1-j}$.

	$\hat{\nu}$				e19			
	$k=1$	$k=2$	$k=5$	$k=11$	$k=1$	$k=2$	$k=5$	$k=11$
MC	1.00	1.02	1.01	1.00	22.2	21.0	18.8	15.5
Lat+s	1.43	1.48	1.34	1.04	30.3	28.5	22.8	15.6
Lat+s+b	1.57	1.65	1.28	1.02	33.5	30.8	22.1	15.6
Sob+LMS	1.78	1.56	1.21	1.02	34.1	30.4	21.7	15.7

In our second experiment, we take $a_j^2 = 2^{1-j}$ for $j = 1, \ldots, d$. Now, the choice of k for the CDE makes a difference, and the best choice will obviously be $k = 1$, i.e., hide the term that has the largest variance. Note that with MC, $\mathrm{Var}[X] = 2 - 2^{-d}$, and when we apply CMC by hiding $a_k Z_k$ from the sum, we hide a term of variance $a_k^2 = 2^{1-k}$ and generate a partial sum S_{-k} of variance $2 - 2^{1-k} - 2^{-d}$. Both terms have a normal distribution with mean 0. The results of Example 3 hold with these variances. Table 2 reports the numerical results for $d = 11$ and $k = 1, 2, 5, 11$.

The MC rates $\hat{\nu}$ agree again with the theory, but here the IV depends very much on the choice of k, and this effect is more significant when k is smaller. For example, for Sobol' points, the IV with $k = 1$ is about 300,000 times smaller than with $k = 11$. The reason is that with $k = 11$, we hide only a variable having a very small variance, so the CDE for one sample is a high narrow peak, and the HK variation of $\tilde{g}'(x, \mathbf{u})$ is very large. For $k = 1$ or 2, we have the opposite and the integrand is much more RQMC-friendly.

B.2. Buckling strength of a steel plate

This is a six-dimensional example, taken from Schields and Zhang (2016). It models the buckling strength of a steel plate by

$$X = \left(\frac{2.1}{\Lambda} - \frac{0.9}{\Lambda^2}\right)\left(1 - \frac{0.75 Y_5}{\Lambda}\right)\left(1 - \frac{2 Y_6 Y_2}{Y_1}\right), \tag{1}$$

where $\Lambda = (Y_1/Y_2)\sqrt{Y_3/Y_4}$, and $Y_1,\ldots,Y_6$ are independent random variables whose distributions are given in Table 3. Each distribution is either normal or lognormal, and the table gives the mean and the coefficient of variation (cv), which is the standard deviation divided by the mean. We estimate the density of X over $[a,b] = [0.5169, 0.6511]$, which contains about 99% of the density (leaving out 0.5% on each side). There is a nonzero probability of having $Y_4 \le 0$, in which case X is undefined, but this probability is extremely small and this has a negligible impact on the density estimator over $[a,b]$, so we just ignore it (alternatively we could truncate the density of Y_4). There are also negligible probabilities that the density estimates below are negative and we ignore this.

Table 3 Distribution of each parameter for the buckling strength model.

parameter	distribution	mean	cv
Y_1	normal	23.808	0.028
Y_2	lognormal	0.525	0.044
Y_3	lognormal	44.2	0.1235
Y_4	normal	28623	0.076
Y_5	normal	0.35	0.05
Y_6	normal	5.25	0.07

For this example, computing the density of X conditional on $\mathcal{G}_{-5}$ or $\mathcal{G}_{-6}$ (i.e., when hiding Y_5 or Y_6) is relatively easy, so we will try and compare these two choices. If we hide one of the variables that appear in Λ, the CDE would be harder to compute (it would require to solve a polynomial equation of degree 4 for each sample), and we do not do it. Let us define

$$V_1 = \frac{2.1}{\Lambda} - \frac{0.9}{\Lambda^2}, \qquad V_2 = 1 - \frac{2Y_6Y_2}{Y_1}, \quad \text{and} \quad V_3 = 1 - \frac{3Y_5}{4\Lambda}.$$

Then we have

$$X \le x \quad \Leftrightarrow \quad Y_5 \ge \left(1 - \frac{x}{V_1V_2}\right)\frac{4\Lambda}{3}$$

and

$$f(x \mid \mathcal{G}_{-5}) = f_5\left(\left(1 - \frac{x}{V_1V_2}\right)\frac{4\Lambda}{3}\right)\frac{4\Lambda}{3V_1V_2} = \phi\left(\frac{(1 - x/(V_1V_2))\,4\Lambda/3 - 0.35}{0.0175}\right)\frac{4\Lambda}{0.0525 \cdot V_1V_2}.$$

Similarly,

$$f(x \mid \mathcal{G}_{-6}) = f_6\left(\left(1 - \frac{x}{V_1V_3}\right)\frac{Y_1}{2Y_2}\right)\frac{Y_1}{2Y_2V_1V_3} = \phi\left(\frac{(1 - x/(V_1V_3))\,Y_1/(2Y_2) - 5.25}{0.3675}\right)\frac{Y_1}{0.735 \cdot Y_2V_1V_3}.$$

For GLRDE using Y_6, let $C = (2.1/\Lambda - 0.9/\Lambda^2)(1 - 0.75Y_5/\Lambda)$. We have $X = h(\mathbf{Y}) = C(1 - 2Y_6Y_2/Y_1)$, $h_6(\mathbf{Y}) = 2CY_2/Y_1$, $h_{66}(\mathbf{Y}) = 0$, $\partial \log f_6(Y_6)/\partial Y_6 = -(Y_6 - \mu_6)/\sigma_6^2$, and $\Psi_6 = Y_1(Y_6 - \mu_6)/(2CY_2\sigma_6^2)$.

Table 4 summarizes the results. We see again that with a very simple conditioning, the CDE with RQMC performs extremely well and much better than the GLRDE and the KDE. It is also much better to condition on $\mathcal{G}_{-6}$ than on $\mathcal{G}_{-5}$, and combining the two provides no significant improvement. The GLRDE is better than the KDE under RQMC, but not under MC. Figure 1 displays the IV as a function of n in a log-log-scale for the CDE with $\mathcal{G}_{-5}$ and $\mathcal{G}_{-6}$. It unveils a slightly more erratic behavior of the MISE for the shifted lattice rule (Lat+s) than for the other methods; the performance depends on the choice of parameters of the lattice rule and their interaction with the particular integrand.

Table 4 Values of $\hat{\nu}$ and e19 with a CDE for $\mathcal{G}_{-5}$, $\mathcal{G}_{-6}$, their combination, GLRDE, and the KDE, for the buckling strength model.

	$\hat{\nu}$					e19				
	$\mathcal{G}_{-5}$	$\mathcal{G}_{-6}$	comb.	GLRDE	KDE	$\mathcal{G}_{-5}$	$\mathcal{G}_{-6}$	comb.	GLRDE	KDE
MC	1.00	1.00	1.00	0.98	0.76	13.5	15.4	15.4	10.2	11.7
Lat+s	1.89	1.56	1.56	1.29	0.81	20.0	24.9	24.9	16.6	13.7
Lat+s+b	1.46	1.65	1.60	1.19	0.85	17.5	25.1	25.1	15.9	12.7
Sob+LMS	1.40	1.75	1.75	1.16	0.81	17.7	25.5	25.5	15.9	12.4

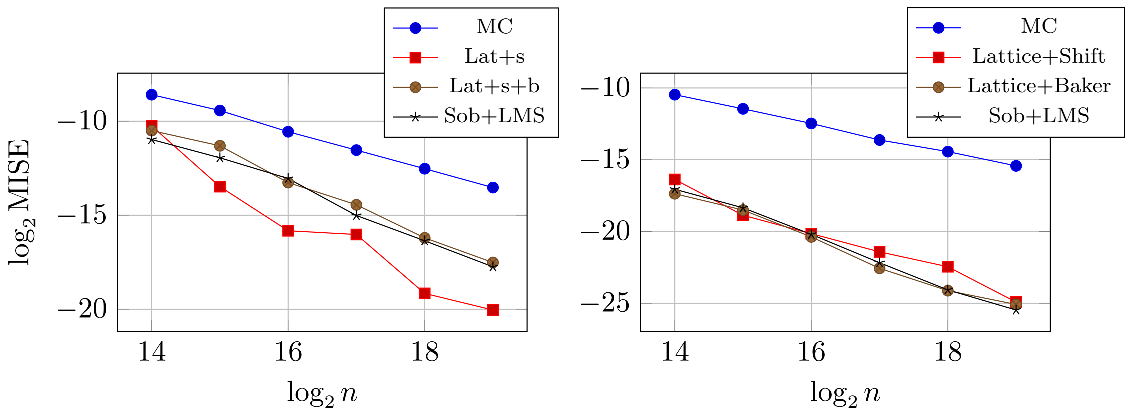

Figure 1 MISE vs n in log-log scale for the $\mathcal{G}=\mathcal{G}_{-5}$ (left) and $\mathcal{G}=\mathcal{G}_{-6}$ (right) for the buckling strength model.

B.3. Density of the failure time of a system

We consider a d-component system in which each component starts in the operating mode (state 1) and fails (jumps to state 0) at a certain random time, to stay there forever. Let Y_j be the failure time of component j for $j=1,\ldots,d$. For $t\geq 0$, let $W_j(t)=\mathbb{I}[Y_j>t]$ be the state of component j and $\mathbf{W}(t)=(W_1(t),\ldots,W_d(t))^{\mathrm{t}}$ the system state, at time t. The system is in the failed mode at time t if and only if $\Phi(\mathbf{W}(t))=0$, where $\Phi:\{0,1\}^d\to\{0,1\}$ is called the *structure function*. Let $X=\inf\{t\geq 0:\Phi(\mathbf{W}(t))=0\}$ be the random time when the system fails. We want to estimate the density of X. A straightforward way of simulating a realization of X is to generate the component lifetimes $Y_j=\inf\{t\geq 0:W_j(t)=0\}$ for $j=1,\ldots,d$, and then compute X from that.

As in Section 4.3, the GLRDE method of Section 2.5 does not work for this example, because $h_j(\mathbf{Y})\neq 0$ only when $X=Y_j$, and there is no j for which this is certain to happen.

If the Y_j are independent and exponential, one can construct a CMC estimator of the cdf $F(x)=\mathbb{P}[X\leq x]$ as follows (Gertsbakh and Shpungin 2010, Botev et al. 2013). Generate all the Y_j's and sort them in increasing order. Then, erase their values and retain only their order, which is a permutation π of $\{1,\ldots,d\}$. Compute the critical number $C=C(\pi)$, defined as the number of component failures required for the system to fail (that is, the system fails at the Cth component failure, for the given π). Note that C can also be computed by starting with all components failed and resurrecting them one by one in reverse order of their failure, until the system becomes operational. Computing C using this reverse order is often more efficient (Botev et al. 2016). Then compute the conditional cdf $\mathbb{P}[X\leq x\mid\pi]$, where X is the time of the Cth component failure. This is an unbiased estimator of $F(x)$ with smaller variance than the indicator $\mathbb{I}[X\leq x]$. It can also be shown that in an asymptotic regime in which the component failure rates converge to 0 so that $1-F(x)\to 0$, the relative variance of this CMC estimator of $1-F(x)$ remains bounded whereas it goes to infinity with the conventional estimator $\mathbb{I}[X>x]$; i.e., the CMC estimator has bounded relative error (Botev et al. 2013, 2016). This X is a sum of C independent exponentials, so it has a hypoexponential distribution, whose cdf has an explicit formula that can be written in terms of a matrix exponential, and developed explicitly as a sum of products in terms of the rates of the exponential lifetimes, as explained below. By taking the derivative of the conditional cdf formula with respect to x, one obtains the conditional density.

More specifically, let component j have an exponential lifetime with rate $\lambda_j>0$, for $j=1,\dots,d$. For a given realization, let $\pi(j)$ be the jth component that fails and let $C(\pi)=c$ for the given π, let A_1 be the time until the first failure, and let A_j be the time between the $(j-1)$th and jth failures, for $j>1$. Conditional on π, we have $X=A_1+\cdots+A_c$ where the A_j's are independent and A_j is exponential with rate Λ_j for all $j\ge 1$, with $\Lambda_1=\lambda_1+\cdots+\lambda_d$, and $\Lambda_j=\Lambda_{j-1}-\lambda_{\pi(j-1)}$ for all $j\ge 2$. The conditional distribution of X is then hypoexponential with cdf

$$\mathbb{P}[X\le x\mid\pi]=\mathbb{P}[A_1+\cdots+A_c\le x\mid\pi]=1-\sum_{j=1}^{c}p_j e^{-\Lambda_j x},$$

where

$$p_j=\prod_{k=1,k\ne j}^{c}\frac{\Lambda_k}{\Lambda_k-\Lambda_j}.$$

See Gertsbakh and Shpungin (2010), Appendix A, and Botev et al. (2016), for example. Taking the derivative with respect to x gives the CDE

$$f(x\mid\pi)=\sum_{j=1}^{c}\Lambda_j p_j e^{-\Lambda_j x},$$

in which c, the Λ_j and the p_j depend on π. This conditional density is well defined and computable everywhere in $[0,\infty)$. There are instability issues for computing p_j when $\Lambda_k-\Lambda_j$ is close to 0 for some $k\ne j$, but this can be addressed by a stable numerical algorithm of Higham (2009).

All of this can be generalized easily to a model in which the lifetimes are dependent, with the dependence modeled by a Marshall-Olkin copula (Botev et al. 2016). In that model, the Y_j represent the occurrence times of shocks that can take down one or more components simultaneously.

It is interesting to note that although $f(x\mid\pi)$ is an unbiased estimator of the density $f(x)$ at any x, this estimator is a function of the permutation π only, so it takes its values in a finite set, which means that the corresponding $\tilde{g}(\mathbf{u})$ is a piecewise constant function, which is not RQMC-friendly. Therefore, we do not expect RQMC to bring a very large gain.

Table 5 Values of $\hat{\nu}$ and e19 with the CDE, for the network reliability example.

	$\hat{\nu}$	e19
MC	1.00	19.9
Lat+s	1.22	23.9
Lat+s+b	1.19	23.8
Sob+LMS	1.33	23.9

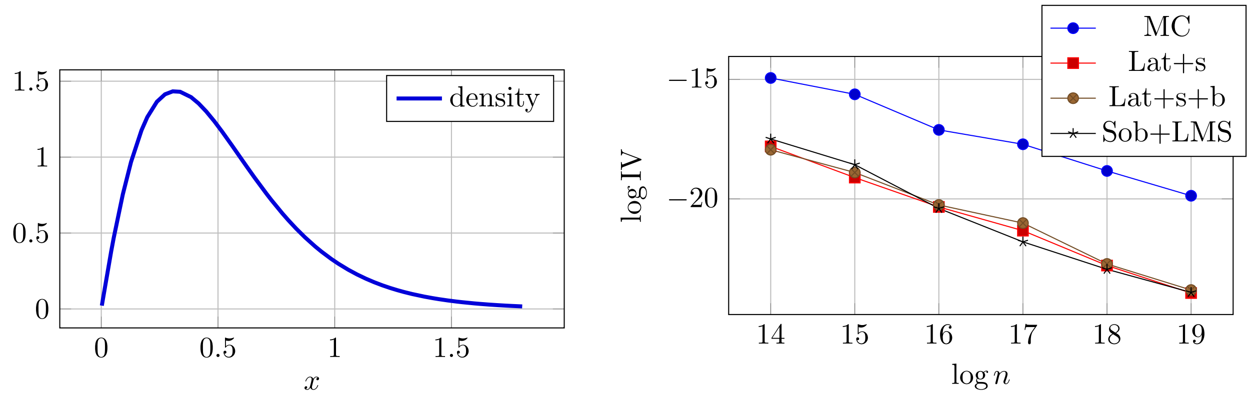

Figure 2 **Density (left) and $\log \mathrm{IV}$ as a function of $\log n$ (right) for the network failure time.**

For a numerical illustration, we take the same graph as in Section 4.3. For $j = 1, \ldots, 13$, Y_j is exponential with rate λ_j and the Y_j are independent. The system fails as soon as there is no path going from the source to the sink. For simplicity, here we take $\lambda_j = 1$ for all j, although taking different λ_j's brings no significant additional difficulty. We estimate the density over the interval $(a, b] = (0, 1.829]$, which cuts off roughly 1% of the probability on the right side. Table 5 and Figure 2 give the results. The density of X estimated with $n = 2^{20}$ random samples is shown on the left and the IV plots are on the right. Despite the discontinuity of $\tilde{g}$, RQMC outperforms MC in terms of the IV by a factor of about $2^4 = 16$ for $n = 2^{19}$, and also by improving the empirical rate $\hat{\nu}$ to about -1.2 for lattices and even better with Sobol' points. The Sobol' points used here were constructed using LatNet Builder (Marion et al. 2020) with a CBC search based on the t-value of all projections up to order 6, with order-dependent weights $\gamma_k = 0.8^k$ for projections of order k.

B.4. Estimating a quantile with a confidence interval

For $0 < q < 1$, the q-quantile of the distribution of X is defined as $\xi_q = F^{-1}(q) = \inf\{x : F(x) \geq q\}$. Given n i.i.d. observations of X, a standard (consistent) estimator of ξ_q is the q-quantile of the empirical distribution, defined as $\hat{\xi}_{q,n} = X_{(\lceil nq \rceil)}$, where $X_{(1)}, \ldots, X_{(n)}$ are the n observations sorted in increasing order (the order statistics). We assume that the density $f(x)$ is positive and continuously differentiable in a neighborhood of ξ_q. Then we have the central limit theorem (CLT):

$$\sqrt{n}(\hat{\xi}_{q,n} - \xi_q)/\sigma_\xi \Rightarrow \mathcal{N}(0,1) \qquad \text{for } n \to \infty,$$

where $\sigma_\xi^2 = q(1-q)/f^2(\xi_q)$ (Serfling 1980). This provides a way to compute a confidence interval on ξ_q, but requires the estimation of $f(\xi_q)$, which is generally difficult. Some approaches for doing this include finite differences with the empirical cdf, batching, and sectioning (Asmussen and Glynn 2007, Nakayama 2014a,b).

In our setting, one can do better by taking the q-quantile $\hat{\xi}_{\mathrm{cmc},q,n}$ of the conditional cdf

$$\hat{F}_{\mathrm{cmc},n}(x) = \frac{1}{n} \sum_{i=1}^{n} F(x \mid \mathcal{G}^{(i)}).$$

That is, $\hat{\xi}_{\mathrm{cmc},q,n} = \inf\{x : \hat{F}_{\mathrm{cmc},n}(x) \geq q\}$. This idea was already suggested by Nakayama (2014b), who pointed out that this estimator obeys a CLT just like $\hat{\xi}_{q,n}$, but with the variance constant σ_ξ^2 replaced by $\sigma_{\mathrm{cmc},\xi}^2 = \mathrm{Var}[F(\xi_q \mid \mathcal{G})]/f^2(\xi_q) \leq \sigma_\xi^2$. This is an improvement on the quantile estimator itself. Our CDE approach also provides an improved estimator of the density $f(\xi_q)$ which appears in the variance expression. We estimate $f(\xi_q)$ by $\hat{f}_{\text{cde},n}(\hat{\xi}_{\text{cmc},q,n})$. This provides a more accurate confidence interval of ξ_q.

Further improvements on the variances of both the quantile and density estimators can be obtained by using RQMC to generate the realizations $\mathcal{G}^{(i)}$. In particular, if $\tilde{g}(\xi_q, \mathbf{u}) = F(\xi_q \mid \mathcal{G})$ is a sufficiently smooth function of $\mathbf{u}$, $\text{Var}[\hat{\xi}_{\text{cmc},q,n}]$ can converge at a faster rate than $\mathcal{O}(n^{-1})$. When using RQMC with n_r randomizations to estimate a quantile, the quantile estimator will be the empirical quantile of all the $n_r \times n$ observations.

A related quantity is the *expected shortfall*, defined as $c_q = \mathbb{E}[X \mid X > \xi_q] = \xi_q - \mathbb{E}[(\xi_q - X)^+]/q$ which is often estimated by its empirical version (Hong et al. 2014)

$$\hat{c}_{q,n} = \hat{\xi}_{q,n} - \frac{1}{nq}\sum_{i=1}^{n}(\hat{\xi}_{q,n} - X_i)^+.$$

This estimator obeys the CLT $\sqrt{n}(\hat{c}_{q,n} - c_q)/\sigma_c \Rightarrow \mathcal{N}(0,1)$ for $n \to \infty$, where $\sigma_c^2 = \text{Var}[(\xi_q - X)^+]/q^2$, if this variance is finite (Hong et al. 2014). By improving the quantile estimator, CDE+RQMC can also improve the expected shortfall estimator a well as the estimator of the variance constant σ_c^2 and the quality of confidence intervals on c_q. We leave this as a topic for future work.

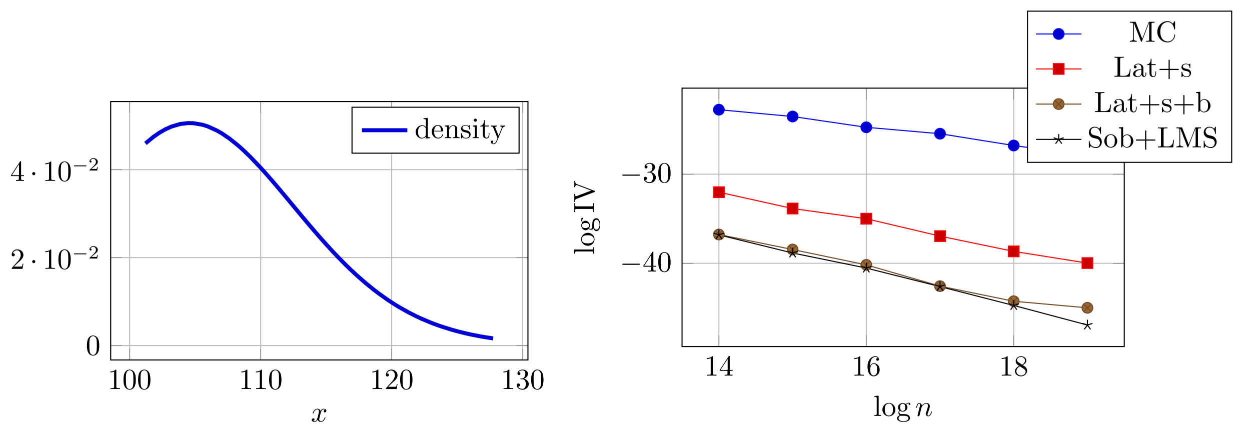

Figure 3 **Estimated density (left) and $\log \text{IV}$ as a function of $\log n$ (left) for the Asian option.**

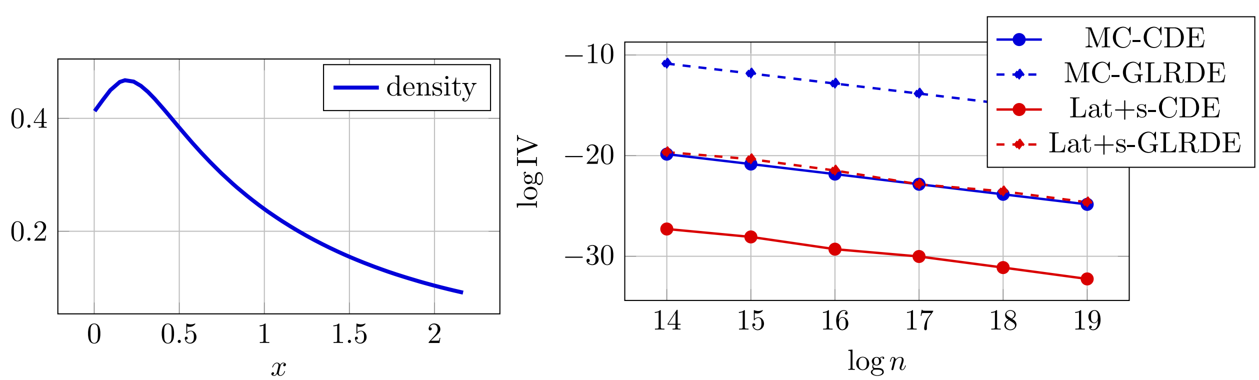

Figure 4 **Estimated density (left) and $\log \text{IV}$ as a function of $\log n$ (right) for the single queue over a finite-horizon.**

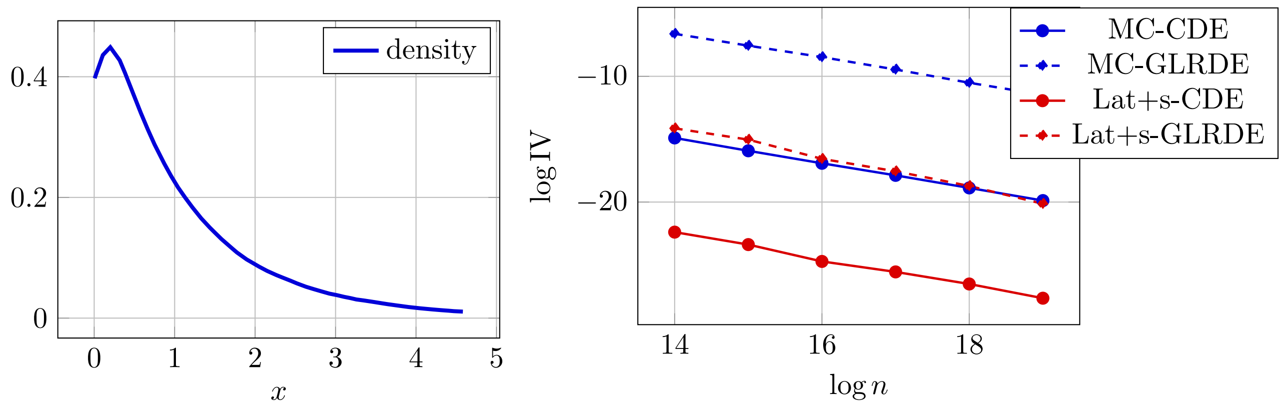

Figure 5 **Estimated density (left) and $\log \mathrm{IV}$ as a function of $\log n$ (right) for the single queue in steady-state.**

Appendix C: Some additional figures

Figure 3 shows the estimated density in $[a,b]$ (left panel) and the IV as a function of n in log-log scale for the bridge CDE in Example 4.6.

Figure 4 shows the estimated density in $[a,b]$ (left panel) and the IV as a function of n in log-log scale for the finite-horizon queueing system in Example 4.4. Figure 5 does the same for the infinite-horizon case.

References

Asmussen S, Glynn PW (2007) *Stochastic Simulation* (New York: Springer-Verlag).

Botev ZI, L'Ecuyer P, Rubino G, Simard R, Tuffin B (2013) Static network reliability estimation via generalized splitting. *INFORMS Journal on Computing* 25(1):56–71.

Botev ZI, L'Ecuyer P, Tuffin B (2016) Static network reliability estimation under the Marshall-Olkin copula. *ACM Transactions on Modeling and Computer Simulation* 26(2):Article 14, 28 pages.

Gertsbakh IB, Shpungin Y (2010) *Models of Network Reliability* (Boca Raton, FL: CRC Press).

Higham NJ (2009) The scaling and squaring method for the matrix exponential revisited. *SIAM Review* 51(4):747–764.

Hong LJ, Hu Z, Liu G (2014) Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Transactions on Modeling and Computer Simulation* 24(4):Article 22.

Marion P, Godin M, L'Ecuyer P (2020) An algorithm to compute the t-value of a digital net and of its projections. *Journal of Computational and Applied Mathematics* 371, URL `http://www.iro.umontreal.ca/~lecuyer/myftp/papers/tvalue.pdf`.

Nakayama MK (2014a) Confidence intervals for quantiles using sectioning when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation* 24(4):Article 9.

Nakayama MK (2014b) Quantile estimation when applying conditional monte carlo. *2014 International Conference on Simulation and Modeling Methodologies, Technologies, and Applications (SIMULTECH)*, 280–285 (IEEE).

Schields MD, Zhang J (2016) The generalization of latin hypercube sampling. *Reliability Engineering and System Safety* 148:96–108.

Serfling RJ (1980) *Approximation Theorems for Mathematical Statistics* (New York, NY: Wiley).