Springer Proceedings in Mathematics & Statistics

Alexander Keller *Editor*

Monte Carlo and Quasi-Monte Carlo Methods

MCQMC 2020, Oxford, United Kingdom, August 10–14

Springer

Springer Proceedings in Mathematics & Statistics

Volume 387

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at https://link.springer.com/bookseries/10533

Alexander Keller
Editor

Monte Carlo and Quasi-Monte Carlo Methods

MCQMC 2020, Oxford, United Kingdom, August 10–14

Springer

Editor
Alexander Keller
NVIDIA
Berlin, Germany

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-98318-5 ISBN 978-3-030-98319-2 (eBook)
https://doi.org/10.1007/978-3-030-98319-2

Mathematics Subject Classification: 11K45, 11K38, 11K36, 65D30, 65C05, 65C10, 11K16, 91G60, 65-04, 62G07, 68Q17, 65R20, 65C40, 62D05, 68T07

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume represents the refereed proceedings of the 14th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing which was held online, August 10–14, 2020. The program of the conference consisted of 97 recorded regular talks featured on the MCQMC presentations YouTube channel. Highlights include the invited plenary talks presented live by Yves Atchadé (Boston University), Jing Dong (Columbia University), Pierre L'Ecuyer (University of Montreal), Mark Jerrum (Queen Mary University London), Peter Kritzer (RICAM Linz), Thomas Müller (NVIDIA), David Pfau (Google DeepMind), Claudia Schillings (University of Mannheim), Mario Ullrich (JKU Linz), and the tutorials by Aretha Teckentrup (Edinburgh) and Fred Hickernell (IIT). While the MCQMC conference regularly attracts between 180 and 240 attendees, more than 600 participants registered for the online version of MCQMC 2020.

The articles in this volume were carefully screened and cover both the theory and the applications of Monte Carlo and quasi-Monte Carlo methods in scientific computing. We thank the anonymous reviewers for their reports and many others who contributed enormously to the excellent quality of the conference presentations and to the high standards for publication in these proceedings by reviewing the abstracts and manuscripts that were submitted.

The next International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2022) will be hosted by the Johannes Kepler University (JKU) and the Johann Radon Institute for Computational and Applied Mathematics (RICAM) in Linz, Austria, in July 2022.

Berlin, Germany
November 2021

Alexander Keller

Acknowledgements

The University of Oxford organizing committee had been looking forward to welcoming you all to Oxford. Unfortunately of course, the coronavirus pandemic has completely ruined all plans. Meeting in person has not been possible; indeed the United Kingdom has been one of the worst affected countries with all of the University of Oxford organizing committee significantly impacted in their jobs.

We appreciate the assistance and sponsoring of the International Centre for Mathematical Sciences (ICMS) in hosting the plenary talks and tutorials. We thank all contributors, who recorded their presentations, and the Berlin organizing committee, who took care of the MCQMC presentations YouTube channel and the conference program booklet. We are very grateful for the support from our sponsors the Mathematical Institute, University of Oxford, the Department of Statistics, University of Oxford, and the Engineering and Physical Sciences Research Council (EPSRC). Last but not least, we thank the scientific committee for arranging an exceptional conference program.

The organizers

Mike Giles
Arnaud Doucet
Alexander Keller

University of Oxford Organizing Committee

Mike Giles	United Kingdom, Oxford
Arnaud Doucet	United Kingdom, Oxford
Abdul-Lateef Haji-Ali	United Kingdom, Heriot-Watt
Jeremy Heng	Singapore, ESSEC
Daniel Paulin	United Kingdom, Edinburgh
Alex Shestopaloff	United Kingdom, Alan Turing Institute
Lukasz Szpruch	United Kingdom, Edinburgh
Aretha Teckentrup	United Kingdom, Edinburgh

Berlin Organizing Committee

Alexander Keller	Germany, NVIDIA
Nikolaus Binder	Germany, NVIDIA
Thomas Müller	Germany, NVIDIA
Merlin Nimier-David	Switzerland, EPFL

Scientific Committee

Christophe Andrieu	United Kingdom, Bristol
Andrea Barth	Germany, Stuttgart
Zdravko Botev	Australia, University of New South Wales
Hector Cancela	Uruguay, University of the Republic
Frédéric Cérou	France, Inria
Nicolas Chopin	France, ENSAE
Ronald Cools	Belgium, KU Leuven
Josef Dick	Australia, University of New South Wales
Arnaud Doucet	United Kingdom, Oxford
Stefan Geiss	Finland, Jyväskylä
Mike Giles	United Kingdom, Oxford
Mark Girolami	United Kingdom, University of Warwick
Paul Glasserman	USA, Columbia
Michael Gnewuch	Germany, Osnabrück
Emmanuel Gobbet	France, Ecole Polytechnique
Takashi Goda	Japan, The University of Tokyo
Arnaud Guyader	France, University Pierre et Marie Curie
Stefan Heinrich	Germany, University Kaiserslautern
Fred J. Hickernell	USA, IIT
Aicke Hinrichs	Austria, JKU Linz

Wenzel Jakob	Switzerland, EPFL
Bert Kappen	Netherlands, Radboud
Alexander Keller	Germany, NVIDIA
Dirk Kroese	Australia, University Queensland
Frances Kuo	Australia, University of New South Wales
Gerhard Larcher	Austria, JKU Linz
Christian Lécot	France, University Savoie
Pierre L'Ecuyer	Canada, University Montréal
Christiane Lemieux	Canada, Waterloo
Gunther Leobacher	Austria, Graz
Faming Liang	USA, University Florida, Gainesville
Makoto Matsumoto	Japan, Hiroshima
Eric Moulines	France, Ecole Polytechnique
Thomas Müller-Gronbach	Germany, Passau
Andreas Neuenkirch	Germany, Mannheim
Harald Niederreiter	Austria, Academy of Sciences
Erich Novak	Germany, FSU, Jena
Dirk Nuyens	Belgium, KU Leuven
Art Owen	USA, Stanford University
Gilles Pagès	France, UPMC
Gareth Peters	United Kingdom, University College London
Friedrich Pillichshammer	Austria, JKU Linz
Mike Pitt	United Kingdom, KCL
Sebastian Reich	Germany, Potsdam
Klaus Ritter	Germany, Kaiserslautern
Gerardo Rubino	France, Inria
Wolfgang Schmid	Austria, Salzburg
Ian Sloan	Australia, University of New South Wales
Lukasz Szpruch	United Kingdom, Edinburgh
Aretha Teckentrup	United Kingdom, Edinburgh
Raul Tempone	Germany, Aachen
Bruno Tuffin	France, INRIA
Grzegorz Wasilkowski	USA, University Kentucky
Henryk Woźniakowski	USA, Columbia University

The MCQMC Conference Series

The MCQMC conference series is a biennial meeting focused on Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods in scientific computing. Its aim is to provide a forum where leading researchers and users can exchange information on the latest theoretical developments and important applications of these methods. The events are held in alternate years with the International Conference on Monte Carlo Methods and Applications (MCM).

In a nutshell, Monte Carlo methods study complex systems by simulations fed by computer-generated pseudorandom numbers. Quasi-Monte Carlo methods replace these random numbers by more uniformly distributed and carefully selected numbers to improve their effectiveness. A large variety of special techniques have been developed and used to make these methods more effective in terms of speed and accuracy. The conference series focuses on the mathematical study of these techniques, their implementation and concrete applications, and their empirical assessment.

The conference was initiated by Harald Niederreiter [6], who co-chaired the first seven conferences. In 2006, Harald Niederreiter announced his wish to step down from the organizational role, and a steering committee was formed to ensure and oversee the continuation of the conference series. Both the steering committee and the locations of the so far 15 conferences are set out below.

If you are interested in hosting a future MCQMC conference at your institution, then please contact any member of the steering committee.

Steering Committee

Alexander Keller (Chair)	Germany, NVIDIA
Josef Dick	Australia, University of New South Wales
Fred J. Hickernell	USA, Illinois Institute of Technology
Pierre L'Ecuyer	Canada, University Montréal
Christiane Lemieux	Canada, University of Waterloo
Art Owen	USA, Stanford University
Friedrich Pillichshammer	Austria, Johannes Kepler University Linz

Conferences

We express our gratitude to Springer-Verlag for publishing the proceedings of the MCQMC conferences.

1. University of Nevada in Las Vegas, Nevada, USA, June 1994 [9]
2. University of Salzburg, Austria, July 1996 [8]
3. Claremont Colleges in Claremont, California, USA, June 1998 [10]
4. Hong Kong Baptist University in Hong Kong, China, November 2000 [4]
5. National University of Singapore, Republic of Singapore, November 2002 [7]
6. Palais des Congrès in Juan-les-Pins, France, June 2004 [11]
7. Ulm University, Germany, July 2006 [5]
8. Université de Montréal, Canada, July 2008 [3]
9. University of Warsaw, Poland, August 2010 [13]
10. University of New South Wales, Sydney, Australia, February 2012 [2]
11. Katholieke Universiteit Leuven, Belgium, April 2014 [1]
12. Stanford University, USA, August 2016 [12]
13. University of Rennes, France, July 2018 [14]
14. University of Oxford, United Kingdom, held online, August 2020
15. Johannes Kepler University in Linz, Austria, July 2022

References

1. Cools, R., Nuyens, D. (eds.): Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Leuven, Belgium, April 2014. Springer Proceedings in Mathematics & Statistics. Springer (2016). doi:https://doi.org/10.1007/978-3-319-33507-0
2. Dick, J., Kuo, F.Y., Peters, G.W., Sloan, I.H. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2012. Springer Proceedings in Mathematics & Statistics. Springer (2013). doi:https://doi.org/10.1007/978-3-642-41095-6
3. Ecuyer, P.L., Owen, A.B. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2008. Springer Proceedings in Mathematics & Statistics. Springer (2009). doi:https://doi.org/10.1007/978-3-642-04107-5
4. Fang, K.T., Hickernell, F.J., Niederreiter, H. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2000 Proceedings of a Conference held at Hong Kong Baptist University, Hong Kong SAR, China, November 27 – December 1, 2000. Springer Proceedings in Mathematics & Statistics. Springer (2002). doi:https://doi.org/10.1007/978-3-642-56046-0
5. Keller, A., Heinrich, S., Niederreiter, H. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2006. Springer Proceedings in Mathematics & Statistics. Springer (2008). doi:https://doi.org/10.1007/978-3-540-74496-2
6. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)
7. Niederreiter, H. (ed.): Monte Carlo and Quasi-Monte Carlo Methods 2002, Proceedings of a Conference held at the National University of Singapore, Republic of Singapore, November 25–28, 2002. Springer Proceedings in Mathematics & Statistics. Springer (2004). doi:https://doi.org/10.1007/978-3-642-18743-8
8. Niederreiter, H., Hellekalek, P., Larcher, G., Zinterhof, P. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 1996, Proceedings of a conference at the University of Salzburg, Austria, July 9–12, 1996. Springer Proceedings in Mathematics & Statistics. Springer (1998). doi:https://doi.org/10.1007/978-1-4612-1690-2

9. Niederreiter, H., Shiue, P.J. (eds.): Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Proceedings of a conference at the University of Nevada, Las Vegas, Nevada, USA, June 23–25, 1994. Springer Proceedings in Mathematics & Statistics. Springer (1995). doi:https://doi.org/10.1007/978-1-4612-2552-2
10. Niederreiter, H., Spanier, J. (eds.): Monte Carlo and Quasi-Monte Carlo Methods, Proceedings of a Conference held at the Claremont Graduate University, Claremont, California, USA, June 22–26, 1998. Springer Proceedings in Mathematics & Statistics. Springer (2000). doi:https://doi.org/10.1007/978-3-642-59657-5
11. Niederreiter, H., Talay, D. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2004. Springer Proceedings in Mathematics & Statistics. Springer (2006). doi:https://doi.org/10.1007/3-540-31186-6
12. Owen, A.B., Glynn, P.W. (eds.): Monte Carlo and Quasi-Monte Carlo Methods, MCQMC 2016, Stanford, CA, August 14-19. Springer Proceedings in Mathematics & Statistics. Springer (2018). doi:https://doi.org/10.1007/978-3-319-91436-7
13. Plaskota, L., Woźniakowski, H. (eds.): Monte Carlo and Quasi-Monte Carlo Methods 2010. Springer Proceedings in Mathematics & Statistics. Springer (2012). doi:https://doi.org/10.1007/978-3-642-27440-4
14. Tuffin, B., L'Ecuyer, P. (eds.): Monte Carlo and Quasi-Monte Carlo Methods, MCQMC 2018, Rennes, France, July 1–6. Springer Proceedings in Mathematics & Statistics. Springer (2020). doi:https://doi.org/10.1007/978-3-030-43465-6

Contents

Invited Talks and Tutorials

Density Estimation by Monte Carlo and Quasi-Monte Carlo

Pierre L'Ecuyer and Florian Puchhammer

Abstract Estimating the density of a continuous random variable X has been studied extensively in statistics, in the setting where n independent observations of X are given a priori and one wishes to estimate the density from that. Popular methods include histograms and kernel density estimators. In this review paper, we are interested instead in the situation where the observations are generated by Monte Carlo simulation from a model. Then, one can take advantage of variance reduction methods such as stratification, conditional Monte Carlo, and randomized quasi-Monte Carlo (RQMC), and obtain a more accurate density estimator than with standard Monte Carlo for a given computing budget. We discuss several ways of doing this, proposed in recent papers, with a focus on methods that exploit RQMC. A first idea is to directly combine RQMC with a standard kernel density estimator. Another one is to adapt a simulation-based derivative estimation method such as smoothed perturbation analysis or the likelihood ratio method to obtain a continuous estimator of the cumulative density function (CDF), whose derivative is an unbiased estimator of the density. This can then be combined with RQMC. We summarize recent theoretical results with these approaches and give numerical illustrations of how they improve the convergence of the mean square integrated error.

Keywords Density estimation · Conditional Monte Carlo · Likelihood ratio · Kernel density

P. L'Ecuyer (✉)
Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montreal, Quebec, Canada
e-mail: lecuyer@iro.umontreal.ca

F. Puchhammer
Université de Montréal, Montreal, Quebec, Canada
e-mail: fpuchhammer@bcamath.org

Basque Center for Applied Mathematics, Bilbao, Spain

A. Keller (ed.), *Monte Carlo and Quasi-Monte Carlo Methods*, Springer Proceedings in Mathematics & Statistics 387, https://doi.org/10.1007/978-3-030-98319-2_1

1 Introduction

In September 2015, the first author (PL) had an interesting lunchtime discussion with Art Owen and Fred Hickernell at a workshop on High-Dimensional Numerical Problems, at the Banff International Research Center, in the Canadian Rocky Mountains. It went as follows. In the MCQMC community, we focus largely on studying QMC and RQMC methods to estimate integrals that represent the mathematical expectations of certain random variables. In applications, the output random variable X of interest often represents a random cost or performance measure. But why estimate only the mean (the expectation) $\mathbb{E}[X]$? Data from simulation experiments can provide much more useful information than just an estimator and a confidence interval for $\mathbb{E}[X]$. When the number n of realizations of X is large enough, it permits one to estimate the entire distribution of X. And when X is a continuous random variable, this distribution is best visualized by showing its density. On the other hand, density estimation from a sample of n independent realizations of X is known to be a difficult problem in statistics. The leading density estimation methods, e.g., kernel density estimators (KDEs), only achieve a convergence rate of $O(n^{-4/5})$ for the mean square error (MSE) on the density at a given point, compared to a $O(n^{-1})$ rate for the expectation with MC. The main question raised in our 2015 discussion was: We know that RQMC can improve the $O(n^{-1})$ rate for the mean, but can it also improve the $O(n^{-4/5})$ rate for the density, by how much, and how?

Of course, this question makes sense only when the samples of X are obtained by simulation from a model, and not in the situation where n independent observations of X are given a priori. When the observations are generated from a model, there is room to change the way we generate them and construct the estimator, and in particular we may use RQMC points in place of independent uniform random numbers to generate the observations of X. Following this discussion, PL started exploring empirically what happens when we do this with an ordinary KDE. That is, what happens with the variance and MSE of the KDE estimator when the n observations of X are generated by simulation using a set of n RQMC points in place of n independent points, just like we do when estimating the mean. After much experiments and theoretical work with co-authors, this led to [3]. In that paper, we were able to prove an upper bound for the MSE with KDE+RQMC, but this bound converges at a faster rate than $O(n^{-4/5})$ only when the dimension s is very small. For moderate and large s, the bound converges at a slower rate than for crude Monte Carlo (MC), although the observed MSE was never larger than for MC in our experiments. The reason for the slow rate for the bound is that when increasing n, we need to reduce the bandwidth of the KDE to reduce the square bias and the MSE, but reducing the bandwidth increases rapidly the variation of the estimator as a function of the uniform random numbers, and this hurts the RQMC estimator.

We understood that for RQMC to be effective, we need smoother density estimators. In January 2017, while PL was visiting A. Owen at Stanford University to work on [3] he attended a talk by S. Asmussen who (by pure coincidence) was presenting [1], in which he shows how to obtain an unbiased density estimator for a sum

of independent random variables by conditional Monte Carlo. The conditioning is done by hiding the last variable in the sum and taking the density of the last variable right-shifted by the sum of other variables as a density estimator. We extended this idea to more general simulation models and this gave us what we needed to obtain smooth unbiased and RQMC-friendly density estimators. This led to the conditional density estimator (CDE) studied in [17], also presented in 2018 at a SAMSI workshop on QMC methods in North Carolina and at a RICAM workshop in Austria. The idea of this CDE method is to define a continuous estimator of the CDF $F(x)$ by conditioning, and take its sample derivative with respect to x as a density estimator. Under appropriate conditions, this provides an unbiased density estimator, and when further favorable conditions hold, this estimator can be smooth and RQMC-friendly. In March 2021, while we were finalizing this paper, Mike Fu pointed out that [6] already contains an example in which he uses conditional Monte Carlo to estimate the density of the length of the longest path in a six-link network in which the last link is shared by all paths. His unbiased density estimator is essentially the same as in [1]: it is the density of the length of the last link, right-shifted by the length of the longest path up to that link.

At the Eleventh International Conference on Monte Carlo Methods and Applications (MCM), in July 2017, the authors of [10] presented a different approach that can provide an unbiased density estimator for a sum of random variables as in [1], except that the variables can be dependent. This approach can be generalized to obtain a continuous CDF estimator and then an unbiased density estimator, via the likelihood ratio (LR) simulation-based derivative estimation method [7, 11] and a clever change of variable, and by taking again the sample derivative of this CDF estimator. This likelihood ratio density estimator (LRDE) is discussed in Sect. 6 and also in [16]. We also explain how it can be combined with RQMC.

A generalized version of the LR gradient estimator method, named GLR, was proposed in [23] to handle situations in which neither the usual LR estimator nor the direct sample derivative apply, because of discontinuities. In [18], the authors sketch out how this GLR method could be used to obtain an unbiased density estimator. Their general formulas are not easy to understand and implement, but more convenient formulas for these GLR density estimators are given in Theorem 1 of [22]. A modified version of the GLR named GLR-U was developed recently in [24] to handle large classes of situations that could not be handled easily by the original GLR from [23]. The model of [24] is expressed explicitly in terms of independent uniform random variables over $(0, 1)$. Density estimators can also be obtained by this method.

All these LR and GLR methods use a multivariate change of variable of some sort. They provide unbiased density estimators that are often not smooth with respect to the underlying uniforms, so their direct combination with RQMC does not always bring much gain. However, it is often possible to smooth out the LR, GLR, or GLR-U density estimator by conditioning just before applying RQMC.

The aim of this paper is to provide an overview of these recent developments on density estimation for simulation models, by MC and RQMC. We summarize the main theoretical results and give numerical illustrations on how the estimators behave, using simple examples.

The remainder is organized as follows. In Sects. 2 and 3, we recall basic facts about one-dimensional density estimation and RQMC sampling. In Sect. 4, we summarize what happens when we directly combine a KDE with RQMC. In Sect. 5, we discuss the CDE and its combination with RQMC. In Sect. 6, we examine the LR and GLR density estimators. Section 8 gives numerical illustrations. We wrap up with a conclusion in Sect. 9.

2 Basic Density Estimation

Let X be a continuous real-valued random variable with CDF F and density f. The goal is to estimate the density f over a finite interval $[a, b]$, from a sample $X_1, \ldots, X_n$ of n realizations of X (not necessarily independent). This problem has been studied at length in statistics for the case where $X_1, \ldots, X_n$ are independent [26]. To measure the quality of an arbitrary density estimator $\hat{f}_n$ based on this sample, we will use the *mean integrated square error* (MISE), which is the integral of the MSE over the interval $[a, b]$:

$$\text{MISE} = \text{MISE}(\hat{f}_n) = \int_a^b \mathbb{E}[\hat{f}_n(x) - f(x)]^2 \mathrm{d}x = \text{IV} + \text{ISB} \tag{1}$$

where

$$\text{IV} = \int_a^b \mathbb{E}(\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)])^2 \mathrm{d}x \quad \text{and} \quad \text{ISB} = \int_a^b (\mathbb{E}[\hat{f}_n(x)] - f(x))^2 \mathrm{d}x$$

are the *integrated variance* (IV) and the *integrated square bias* (ISB), respectively.

Two popular types of density estimators are histograms and KDEs. To define a *histogram*, one can partition $[a, b]$ into m intervals of length $h = (b - a)/m$ and put

$$\hat{f}_n(x) = \frac{n_j}{nh} \quad \text{for } x \in I_j = [a + (j-1)h, a + jh), \quad j = 1, \ldots, m,$$

where n_j is the number of observations X_i that fall in interval I_j. To define a KDE [21, 26], select a *kernel* k (usually a unimodal symmetric density centered at 0) and a *bandwidth* $h > 0$ (an horizontal stretching factor for the kernel), and put

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right).$$

These two density estimators are biased. Asymptotically, when $n \to \infty$ and $h \to 0$ jointly, in the case of independent samples $X_1, \ldots, X_n$, the IV and ISB behave as

Table 1 Constants involved in the convergence rates of the MISE for histograms and KDEs

	C	B	α	h^*	MISE
Histogram	1	$R(f')/12$	2	$(nR(f')/6)^{-1/3}$	$O(n^{-2/3})$
KDE	$\mu_0(k^2)$	$(\mu_2(k))^2 R(f'')/4$	4	$\left(\frac{\mu_0(k^2)}{(\mu_2(k))^2 R(f'')n}\right)^{1/5}$	$O(n^{-4/5})$

$$\text{MISE} \;=\; \text{IV} + \text{ISB} \;\sim\; C/(nh) + Bh^{\alpha}$$

where C, B, and α depend on the method. The asymptotically optimal h is then

$$h^* = (C/(B\alpha n))^{1/(\alpha+1)}$$

and it gives MISE $\sim K n^{-\alpha/(1+\alpha)}$ for some constant K. Table 1 gives expressions for C, B, α, h^*, and $\alpha/(1+\alpha)$, for histograms and KDEs, with independent samples. It uses the following definitions, for any $g : \mathbb{R} \to \mathbb{R}$:

$$R(g) = \int_a^b (g(x))^2 \mathrm{d}x \quad \text{and} \quad \mu_r(g) = \int_{-\infty}^{\infty} x^r g(x) \mathrm{d}x \quad \text{for } r = 0 \text{ and } 2.$$

Note that these expressions hold under the simplifying assumption that h must be the same all over $[a, b]$. One may often do better by varying the bandwidth over $[a, b]$, but this is more complicated. To estimate h^* in practice, one can estimate $R(f')$ and $R(f'')$ by using a KDE to estimate f' and f'' (very roughly). This type of crude (plugin) estimate is often good enough. In the following, we will see how to improve on these MISE rates and values in a simulation setting, by reducing the variance. In general, using RQMC points instead of MC does not change the bias.

3 RQMC

We recall here some basic principles of RQMC used in the forthcoming sections. For more extensive coverages, see [4, 12, 13, 20], for example. Suppose we want to estimate $\mu = \mathbb{E}[g(\boldsymbol{U})]$ where $\boldsymbol{U} = (U_1, \ldots, U_s)$ has the uniform distribution over the s-dimensional unit cube $(0, 1)^s$ and $g : (0, 1)^s \to \mathbb{R}$. With standard MC, we draw n independent random points $\boldsymbol{U}_i$ uniformly over $(0, 1)^s$ and we estimate the expectation by the average

$$\hat{\mu}_{n,\text{mc}} = \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{U}_i). \tag{2}$$

With RQMC, we replace the independent random points $\boldsymbol{U}_i$ by a set of *dependent* random points $\tilde{P}_n = \{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n\} \subset (0, 1)^s$ such that (1) the point set $\tilde{P}_n$ covers

the unit hypercube very evenly (in a sense that must be precisely defined) with probability 1; and (2) each point $\boldsymbol{U}_i$ has the uniform distribution over $(0, 1)^s$. Then we estimate the expectation by the same average as in (2), which we now denote $\hat{\mu}_{n,\text{rqmc}}$. For various spaces $\mathcal{H}$ of functions g, usually Hilbert or Banach spaces, we have inequalities of the form

$$\text{Var}[\hat{\mu}_{n,\text{rqmc}}] \leq \mathcal{D}^2(P_n) \cdot \mathcal{V}^2(g) \tag{3}$$

where $\mathcal{D}(P_n)$ measures the *discrepancy* of P_n (with respect to the uniform distribution) and $\mathcal{V}(g)$ measures the variation of the function g. For many of these function spaces, we also know explicitly how to construct RQMC point sets for which $\mathcal{D}(P_n) = O(n^{-\alpha/2}(\log n)^{s-1})$ for some $\alpha > 1$ [4, 9, 14]. This leads to

$$\text{Var}[\hat{\mu}_{n,\text{rqmc}}] = O(n^{-\alpha}(\log n)^{2(s-1)})$$

when $\mathcal{V}(f) < \infty$. A classical case is the standard Koksma-Hlawka inequality, for which $\alpha = 2$, $\mathcal{D}(P_n) = \mathcal{D}^*(P_n)$ is the star discrepancy, and $\mathcal{V}(g)$ is the variation in the sense of Hardy and Krause, defined by

$$\mathcal{V}(g) = \mathcal{V}_{\text{HK}}(g) = \sum_{\emptyset \neq \mathfrak{v} \subseteq \{1,\ldots,s\}} \int_{(0,1)^{|\mathfrak{v}|}} \left| \frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} g(\boldsymbol{u}_{\mathfrak{v}}, \mathbf{1}) \right| \mathrm{d}\boldsymbol{u}_{\mathfrak{v}}, \tag{4}$$

where $\boldsymbol{u}_{\mathfrak{v}}$ is the vector of coordinates whose indices belong to $\mathfrak{v}$, $|\mathfrak{v}|$ is the cardinality of $\mathfrak{v}$, and under the assumption that this expression is well defined. The main construction methods for P_n are lattice rules and digital nets.

In the context of density estimation, the average in (2) is replaced by the density estimator $\hat{f}_n(x)$ at a given point x. If our density estimator can be written as an average of the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}(x, \boldsymbol{U}_i) \tag{5}$$

where $\tilde{g}$ is a sufficiently smooth function of its second argument, then we can apply the RQMC theory just described to this density estimator by replacing the function $g(\cdot)$ by $\tilde{g}(x, \cdot)$. We look at this in the next few sections.

4 Kernel Density Estimators with RQMC

The KDE at a given point $x \in [a, b]$ is

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} k\left(\frac{x - g(\boldsymbol{U}_i)}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}(x, \boldsymbol{U}_i).$$

We assume that the kernel k is a smooth probability density, symmetric about 0, and at least s times differentiable everywhere. With RQMC points $\boldsymbol{U}_i$, this is an RQMC estimator of $\mathbb{E}[\tilde{g}(x, \boldsymbol{U})] = \mathbb{E}[\hat{f}_n(x)]$. RQMC does not change the bias of this density estimator, but it may reduce $\mathrm{Var}[\hat{f}_n(x)]$, which would reduce in turn the IV and the MISE.

To prove RQMC variance bounds via (3), we need to bound the variation $\mathcal{V}(\tilde{g})$. This was done in [3] for the classical Hardy-Krause variation (4), which is bounded if and only if all the partial derivatives

$$\frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} \tilde{g}(x, \boldsymbol{u}) = \frac{1}{h} \frac{\partial^{|\mathfrak{v}|}}{\partial \boldsymbol{u}_{\mathfrak{v}}} k\left(\frac{x - g(\boldsymbol{u})}{h}\right)$$

exist and are uniformly bounded. The derivatives with respect to k are easily bounded for instance if k is a standard normal density (the Gaussian kernel). However, when expanding the derivatives via the chain rule, we obtain terms in h^{-j} for $j = 2, \ldots, |\mathfrak{v}| + 1$. The dominant term asymptotically is the term for $|\mathfrak{v}| = s$, and it grows in general as $h^{-s-1} \left| k^{(s)}\left((x - g(\boldsymbol{u}))/h\right) \prod_{j=1}^{s} g_{\{j\}}(\boldsymbol{u}) \right| = O(h^{-s-1})$ when $h \to 0$, where $g_{\{j\}}$ is the derivative of g with respect to its jth coordinate. We can bring it down to $O(h^{-s})$ via a change of variables, which leads to the following result proved in [3]:

Proposition 1 *Let $g : [0, 1]^s \to \mathbb{R}$ be* piecewise monotone *in each coordinate u_j when the other coordinates are fixed. Assume that all first-order partial derivatives of g are continuous and that $\|g_{\mathfrak{w}_1} g_{\mathfrak{w}_2} \cdots g_{\mathfrak{w}_\ell}\|_1 < \infty$ for all selections of non-empty, mutually disjoint index sets $\mathfrak{w}_1, \ldots, \mathfrak{w}_\ell \subseteq \{1, \ldots, s\}$, where $g_{\mathfrak{w}}$ is the derivative of g with respect to all the coordinates in the index set $\mathfrak{w}$.*

Then the Hardy-Krause variation of $\tilde{g}(x, \cdot)$ for any fixed $x \in [a, b]$ satisfies

$$\mathcal{V}_{\mathrm{HK}}(\tilde{g}(x, \cdot)) \le c h^{-s} + O(h^{-s+1})$$

for some constant $c > 0$ given in [3], and with RQMC point sets having a star discrepancy $\mathcal{D}^*(P_n) = O(n^{-1+\epsilon})$ for all $\epsilon > 0$ when $n \to \infty$, we obtain

$$\mathrm{IV} = O(n^{-2+\epsilon} h^{-2s}) \quad \textit{for all } \epsilon > 0.$$

RQMC does not change the bias, so the ISB has exactly the same expression as for MC. By picking h to minimize the MISE bound, we get $\mathrm{MISE} = O(n^{-4/(2+s)+\epsilon})$.

This rate for the MISE is worse than the MC rate when $s \ge 4$. The factor h^{-2s} in the IV bound really hurts. On the other hand, this is only an upper bound, not the actual IV. Proposition 4.4 of [3] also shows via a different analysis that for the KDE, there exist RQMC constructions for which the asymptotic decrease rate of the IV is not worse than for MC.

5 Conditional Density Estimation with RQMC

To estimate the density $f(x) = F'(x)$, one may think of simply taking the sample derivative of an unbiased estimator of the CDF $F(x)$. The simplest unbiased estimator of this CDF is the *empirical CDF*

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[X_i \le x].$$

However $\mathrm{d}\hat{F}_n(x)/\mathrm{d}x = 0$ almost everywhere, so this *cannot* be a useful density estimator! We need a smoother CDF estimator, which should be at least continuous in x.

One effective way of smoothing an estimator and often make it continuous is to replace it by its conditional expectation given partial (filtered) information. This is *conditional Monte Carlo* (CMC) [2]. That is, one replaces the indicator $\mathbb{I}[X_i \le x]$ in the expression of $\hat{F}_n(x)$ above by the conditional CDF $F(x \mid \mathcal{G}) = \mathbb{P}[X_i \le x \mid \mathcal{G}]$, where $\mathcal{G}$ is a sigma-field that contains not enough information to reveal X but enough to compute $F(x \mid \mathcal{G})$, then one takes the sample derivative. We call it the *conditional density estimator* (CDE). For more details about the CMC method in general and the choice of $\mathcal{G}$ in specific cases, see for example [2, 5, 15]. For examples in the context of density estimation, see [17] and the examples in Sect. 8. We assume here that we can compute the conditional density either directly or numerically by an iterative algorithm. The following proposition, proved in [17], gives sufficient conditions for this CDE to be an unbiased density estimator with finite variance.

Proposition 2 *Suppose that for all realizations of* $\mathcal{G}$, $F(x \mid \mathcal{G})$ *is a* continuous *function of* x *over* $[a, b]$, differentiable *except perhaps over a denumerable set of points* $D(\mathcal{G}) \subset [a, b]$, *and for which* $f(x \mid \mathcal{G}) = F'(x \mid \mathcal{G}) = \mathrm{d}F(x \mid \mathcal{G})/\mathrm{d}x$ *(when it exists) is bounded uniformly in* x *by a random variable* Γ *such that* $\mathbb{E}[\Gamma^2] \le K_\gamma < \infty$. *Then, for all* $x \in [a, b]$, $\mathbb{E}[f(x \mid \mathcal{G})] = f(x)$ *and* $\mathrm{Var}[f(x \mid \mathcal{G})] < K_\gamma$. *Moreover, if* $\mathcal{G} \subset \tilde{\mathcal{G}}$ *both satisfy the assumptions of this proposition, then* $\mathrm{Var}[f(x \mid \mathcal{G})] \le \mathrm{Var}[f(x \mid \tilde{\mathcal{G}})]$.

For a sample of size n, the CDE becomes

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

where $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(n)}$ are n "realizations" of $\mathcal{G}$. When the n realizations are independent we have $\mathrm{Var}[\hat{f}_{\mathrm{cde},n}(x)] \le K_\gamma/n = O(n^{-1})$.

To combine the CDE with RQMC, we want to write $f(x \mid \mathcal{G}) = \tilde{g}(x, \boldsymbol{u})$ for some function $\tilde{g} : [a, b] \times [0, 1)^s \to \mathbb{R}$. This function $\tilde{g}(x, \cdot)$ will be used in (5). The combined CDE+RQMC estimator is then defined by

$$\hat{f}_{\text{cde-rqmc},n}(x) = \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(x, \boldsymbol{U}_i). \tag{6}$$

where $\{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n\}$ is an RQMC point set.

If $\tilde{g}(x, \cdot)$ has *bounded variation*, then we can get an $O(n^{-2+\epsilon})$ *rate for the MISE*, and sometimes better. This holds in several examples that we tried. If $\tilde{g}(x, \cdot)$ has *unbounded variation*, RQMC may still reduce the IV, but there is no guarantee.

6 Likelihood Ratio Density Estimators

There are situations where a CDE as in Sect. 5 might be too difficult to obtain. An alternative can be a *likelihood ratio density estimator* (LRDE), defined as follows. Suppose that $X = h(\boldsymbol{Y})$ where $\boldsymbol{Y}$ has known density $f_{\boldsymbol{Y}}$ over $\mathbb{R}^s$, and we know how to generate it and compute $X = h(\boldsymbol{Y})$. For simplicity, let $x > 0$ (in case we are really interested in some $x \le 0$, we can simply add a constant to the function h). We have

$$F(x) = \mathbb{P}[h(\boldsymbol{Y}) \le x] = \int_{\mathbb{R}^s} \mathbb{I}[h(\boldsymbol{y})/x \le 1] f_{\boldsymbol{Y}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}.$$

We want to change this integrand into a continuous function of x, so we can take the derivative with respect to x inside the integral. One way to do this is to make a change of variable $\boldsymbol{y} \mapsto \boldsymbol{z} = \boldsymbol{z}(x)$ of the form $\boldsymbol{y} = \varphi_x(\boldsymbol{z})$, with Jacobian $|J_x(\boldsymbol{z})|$, so that $\tilde{h}(\boldsymbol{z}) = h(\varphi_x(\boldsymbol{z}))/x$ no longer depends on x for any given $\boldsymbol{z}$. We can then rewrite

$$F(x) = \int_{\mathbb{R}^s} \mathbb{I}[\tilde{h}(\boldsymbol{z}) \le 1] f_{\boldsymbol{Y}}(\varphi_x(\boldsymbol{z}))|J_x(\boldsymbol{z})| \mathrm{d}\boldsymbol{z}.$$

In a small open neighborhood of a given $x_0 \in [a, b]$, we have

$$F(x) = \int_{\mathbb{R}^s} \mathbb{I}[\tilde{h}(\boldsymbol{z}) \le 1] L(\boldsymbol{z}; x, x_0) f_{\boldsymbol{Y}}(\varphi_{x_0}(\boldsymbol{z}))|J_{x_0}(\boldsymbol{z})| \mathrm{d}\boldsymbol{z}$$

where

$$L(\boldsymbol{z}; x, x_0) = \frac{f_{\boldsymbol{Y}}(\varphi_x(\boldsymbol{z}))|J_x(\boldsymbol{z})|}{f_{\boldsymbol{Y}}(\varphi_{x_0}(\boldsymbol{z}))|J_{x_0}(\boldsymbol{z})|}$$

is the *likelihood ratio* between the density of $\boldsymbol{z}$ at x and at x_0. Under appropriate conditions:

$$
\begin{aligned}
f(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \int_{\mathbb{R}^s} \mathbb{I}[\tilde{h}(z) \le 1] L(z; x, x_0) f_{\boldsymbol{Y}}(\varphi_{x_0}(z)) |J_{x_0}(z)| \mathrm{d}z \\
&= \int_{\mathbb{R}^d} \mathbb{I}[\tilde{h}(z) \le 1] \left(\frac{\mathrm{d}}{\mathrm{d}x} L(z; x, x_0) \right) \frac{f_{\boldsymbol{Y}}(\varphi_x(z)) |J_x(z)|}{L(z; x, x_0)} \mathrm{d}z \\
&= \int_{\mathbb{R}^d} \mathbb{I}[\tilde{h}(z) \le 1] \left(\frac{\mathrm{d}}{\mathrm{d}x} \ln L(z; x, x_0) \right) f_{\boldsymbol{Y}}(\varphi_x(z)) |J_x(z)| \mathrm{d}z \\
&= \int_{\mathbb{R}^s} \mathbb{I}[h(\boldsymbol{y}) \le x] S(\boldsymbol{y}, x) f_{\boldsymbol{Y}}(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}
\end{aligned}
$$

where

$$
S(\boldsymbol{y}, x) = \frac{\mathrm{d} \ln L(z; x, x_0)}{\mathrm{d}x} = (\nabla(\ln f_{\boldsymbol{Y}})(\boldsymbol{y})) \cdot (\nabla_x \varphi_x(z)) + \frac{\mathrm{d} \ln |J_x(z)|}{\mathrm{d}x}
$$

is the *score function* associated with L. This gives the unbiased LRDE

$$
\hat{f}_{\text{lrde}}(x) = \mathbb{I}[h(\boldsymbol{Y}) \le x] \, S(\boldsymbol{Y}, x) \tag{7}
$$

where $\boldsymbol{Y} \sim f_{\boldsymbol{Y}}$. Here, $\boldsymbol{Y}$ can have a multivariate distribution for which conditioning is hard whereas $S(\boldsymbol{Y}, x)$ may be easier to compute.

This LR approach has been widely used to estimate the derivative of $\mathbb{E}[h(\boldsymbol{Y})]$ with respect to a parameter of the distribution of $\boldsymbol{Y}$ [2, 7, 8, 11]. Laub et al. [10] obtained (via a different argument) the estimator (7) for the special case where $h(\boldsymbol{Y})$ is a sum of random variables. The following is proved in [16].

Proposition 3 *Suppose that with probability one over realizations of $\boldsymbol{Y} = \varphi_x(\boldsymbol{Z})$, $f_{\boldsymbol{Y}}(\varphi_x(\boldsymbol{Z}))|J_x(\boldsymbol{Z})|$ is continuous in x over $[a, b]$ and is differentiable in x except perhaps at a countable set of points $D(\boldsymbol{Y}) \subset [a, b]$. Suppose that there is also a random variable Γ defined over the same probability space as $\boldsymbol{Y}$, such that $\mathbb{E}[\Gamma^2] < \infty$, and for which*

$$
\sup_{x \in [a,b] \setminus D(\boldsymbol{Y})} |\mathbb{I}[h(\boldsymbol{Y}) \le x] \, S(\boldsymbol{Y}, x)| \le \Gamma.
$$

Then, $\hat{f}_{\text{lrde}}(x) = \mathbb{I}[h(\boldsymbol{Y}) \le x] \, S(\boldsymbol{Y}, x)$ is an unbiased estimator of $f(x)$ at almost all $x \in [a, b]$, with variance bounded uniformly by $\mathbb{E}[\Gamma^2]$.

Note that the unbiased LRDE in (7) is usually discontinuous in the underlying uniforms, because of the indicator function, so it is not a smooth RQMC-friendly estimator. One can think of making it continuous by taking its conditional expectation. On the other hand, when we can find a conditioning that makes the indicator continuous, then we may be able to apply the CDE instead and this is usually more effective, according to our experiments. The LRDE is nevertheless useful for the situations in which a CDE is difficult to obtain.

7 Generalized Likelihood Ratio Estimators

Peng et al. [23] proposed a *generalized likelihood ratio* (GLR) method that generalizes the LR derivative estimation approach. Peng et al. [22] gave an adaptation of this method to density estimation. It goes as follows. Let $X = h(\boldsymbol{Y}) = h(Y_1, \ldots, Y_s)$ for some random variables $Y_1, \ldots, Y_s$, and assume that X is a continuous random variable with (unknown) density f. Let $A(x, \epsilon) = \{\boldsymbol{y} \in \mathbb{R}^s : x - \epsilon \le h(\boldsymbol{y}) \le x + \epsilon\}$, which is the inverse image of an ϵ-neighborhood of x by h. Suppose there is an $\epsilon_0 > 0$ such that

$$\lim_{\epsilon \to 0} \sup_{x \in [a-\epsilon_0, b+\epsilon_0]} \lambda(A(x, \epsilon)) = 0,$$

where λ is the Lebesgue measure on $\mathbb{R}^s$. Select some index $j \in \{1, \ldots, s\}$ for which Y_j is a continuous random variable with CDF F_j and density f_j, and is independent of $\{Y_k,\ k \ne j\}$. Let $h_{(j)}(\boldsymbol{y}) := \partial h(\boldsymbol{y})/\partial y_j$, $h_{(jj)}(\boldsymbol{y}) := \partial^2 h(\boldsymbol{y})/\partial y_j^2$, and

$$\Psi_j(\boldsymbol{y}) = \frac{\partial \ln f_j(y_j)/\partial y_j - h_{(jj)}(\boldsymbol{y})/h_{(j)}(\boldsymbol{y})}{h_{(j)}(\boldsymbol{y})},$$

where all these derivatives are assumed to exist. Suppose that there are functions $v_\ell : \mathbb{R} \to \mathbb{R}$ for $\ell = 1, \ldots, s$ such that $|h_{(j)}(\boldsymbol{y})|^{-1} \le \prod_{\ell=1}^d v_\ell(y_\ell)$ and

$$\lim_{y \to \pm\infty} v_j(y) f_j(y) = 0 \ \text{ and } \ \mathbb{E}[v_j(Y_j)] < \infty.$$

Finally, suppose also that $\mathbb{E}[\mathbb{I}[X \le x]\Psi_j^2(\boldsymbol{Y})] < \infty$. Under all these conditions, a simple modification of the proof of Theorem 1 in [22] yields the following:

Proposition 4 *With the assumptions just given, $D_j(x, \boldsymbol{Y}) = \mathbb{I}[X \le x]\Psi_j(\boldsymbol{Y})$ is an unbiased and finite-variance estimator of the density $f(x)$ at x.*

When the conditions hold for all $j = 1, \ldots, s$, as assumed in [22], this gives s unbiased estimators $D_1(x, \boldsymbol{Y}), \ldots, D_s(x, \boldsymbol{Y})$. Instead of selecting only one of them, we can take a linear combination $D(x, \boldsymbol{Y}) = w_1 D_1(x, \boldsymbol{Y}) + \cdots + w_s D_s(x, \boldsymbol{Y})$ where $w_1 + \cdots + w_s = 1$. This is exactly equivalent to taking, say $D_1(x, \boldsymbol{Y})$ as the base estimator and the $C_j = D_j(x, \boldsymbol{Y}) - D_1(x, \boldsymbol{Y})$ as mean-zero control variates, for $j = 2, \ldots, s$, because one has $D(x, \boldsymbol{Y}) = D_1(x, \boldsymbol{Y}) + w_2 C_2 \cdots + w_s C_s$. Therefore, standard control variate theory [2] can be used to optimize the coefficients w_j. When the conditions are satisfied only for certain values of j, then one can take the linear combination only for these values. It may also happen that the assumptions are satisfied for no j, in which case this method does not apply.

The GLR setting of [23] is more general. It permits one to estimate the derivative of $\mathbb{E}[\varphi(g(\boldsymbol{Y}; \theta))]$ with respect to some parameter θ, where $g(\cdot; \theta) : \mathbb{R}^s \to \mathbb{R}^s$ is continuous and one-to-one for the values of θ in the region of interest, so it corresponds to a multivariate change of variable in that region. The authors provide a general form of the unbiased derivative estimator (see also [25]). The general for-

mula is rather complicated and it can be found in the papers. One can use it in principle to estimate the density of X by taking $\theta = x$ and selecting a g for which $\varphi(g(\boldsymbol{Y}; \theta)) \equiv \mathbb{I}[X \le x] = \mathbb{I}[h(\boldsymbol{Y}) - x \le 0]$ and for which the assumptions of [23] are satisfied, when this is possible.

Peng et al. [24] extended the range of applicability of GLR by developing GLR-U, a version of GLR in which the base model is expressed directly in terms of the underlying uniform random numbers. That is, $\boldsymbol{Y}$ takes the form of a vector $\boldsymbol{U}$ which has the uniform distribution over the unit hypercube $(0, 1)^s$. This new setting covers a larger class of models than in [23], including situations where the random variables are generated by inversion, by the rejection method, or via Archimedean copulas, for example. We outline how to use this method to estimate the density of X over $[a, b]$.

The first step is to find a nonempty subset of the input variables $\Upsilon \subseteq \{1, \ldots, s\}$, which we will assume (without loss of generality) to be $\Upsilon = \{1, \ldots, d\}$ for $1 \le d \le s$, together with a function $g(\cdot; x) = g_1(\cdot; x), \ldots, g_d(\cdot; x)) : (0, 1)^s \to \mathbb{R}^d$ for which $\varphi(g(\boldsymbol{U}; x)) \equiv \mathbb{I}[X \le x]$ for all $x \in [a, b]$ and which satisfies the following assumptions. For any $\boldsymbol{u} \in (0, 1)^s$, we decompose $\boldsymbol{u} = (\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)})$ where $\boldsymbol{u}^{(1)}$ contains the first d coordinates and $\boldsymbol{u}^{(2)}$ the other ones. When $\boldsymbol{u}^{(2)}$ is fixed, $g(\cdot; x)$ becomes a function of $\boldsymbol{u}^{(1)}$ only, which we denote by $\tilde{g}(\cdot; \boldsymbol{u}^{(2)}, x)$. An important condition is that this function $\tilde{g}$ must be continuous and correspond to a multivariate change of variable, whose Jacobian $J_g(\boldsymbol{u}; x)$ is a $d \times d$ invertible matrix whose element (i, j) is $\partial g_i(\boldsymbol{u}; x)/\partial u_j$. For any $\boldsymbol{u} = (u_1, \ldots, u_s) \in (0, 1)^s$ and $j = 1, \ldots, d$, let $\overline{\boldsymbol{u}}_j$ and $\underline{\boldsymbol{u}}_j$ be the vector $\boldsymbol{u}$ in the limit when $u_j \to 1$ from the left and the limit when $u_j \to 0$ from the right (see [24, 25]). Define

$$r_j(\boldsymbol{u}; x) = -(J_g^{-1}(\boldsymbol{u}; x))^{\mathsf{t}} \cdot \boldsymbol{e}_j$$

and

$$v(\boldsymbol{u}; x) = -\sum_{j=1}^{d} \boldsymbol{e}_j^{\mathsf{t}} \cdot (J_g^{-1}(\boldsymbol{u}; x)) \left(\frac{\mathrm{d} J_g(\boldsymbol{u}; x)}{\mathrm{d} u_j} \right) (J_g^{-1}(\boldsymbol{u}; x)) \cdot \mathbf{1}$$

where $\boldsymbol{e}_j$ is the jth unit vector, $\mathbf{1}$ is a column vector of ones, and the derivative of $J_g(\boldsymbol{u}; x)$ is element-wise. Then, under some mild regularity conditions, we have:

Proposition 5 *The following is an unbiased density estimator at all $x \in [a, b]$:*

$$G(\boldsymbol{U}, x) = \mathbb{I}[X \le x]\, v(\boldsymbol{U}; x) + \sum_{j=1}^{s} \Big[\varphi(g(\overline{\boldsymbol{U}}_j; x)) r_j(\overline{\boldsymbol{U}}_j; x) - \varphi(g(\underline{\boldsymbol{U}}_j; x)) r_j(\underline{\boldsymbol{U}}_j; x) \Big]. \tag{8}$$

Peng et al. [24, 25] show how to apply this method in the special case where X is the maximum of several variables, each one being the sum of certain Y_j's that are generated by inversion from the U_j's. This may correspond to the length of the longest path between a source node to a destination node in a directed network, for

example. It works in the same way if the maximum is replaced by a minimum, and we will use it in Sect. 8. The number d of selected input variables in this case should be equal to the number of independent paths.

8 Numerical Illustrations

We illustrate the applicability and performance of the various density estimators discussed here on a small shortest path example defined below. We run the simulations with MC and RQMC. For RQMC, we use Sobol' nets with direction numbers taken from [19], and randomized by a left matrix scramble followed by a digital shift. Each RQMC experiment is repeated $m = 100$ times independently. The performance is assessed via the estimated MISE for $n = 2^{20}$ points. For RQMC, we also estimate the convergence rate as follows: we assume that MISE $\approx n^{-\beta}$ for some constant $\beta > 0$ and we estimate β by $\hat{\beta}$ using linear regression in log scale, based on observations obtained with $n = 2^{13}, 2^{14}, \ldots, 2^{20}$. For MC, the rates are known theoretically to be $\beta = 0.8$ for the KDE and $\beta = 1$ for the other methods. For the experiments with the KDE, we select the bandwidth with the same methodology as in [3].

We consider an acyclic directed network as in Fig. 1, with s arcs. For $j = 1, \ldots, s$, arc j has random length Y_j with continuous cdf F_j and density f_j, and the Y_j are assumed independent. We generate Y_j by inversion via $Y_j = F_j^{-1}(U_j)$ where $U_j \sim U(0, 1)$. We want to estimate the density of the length X of the shortest path from the source to the sink.

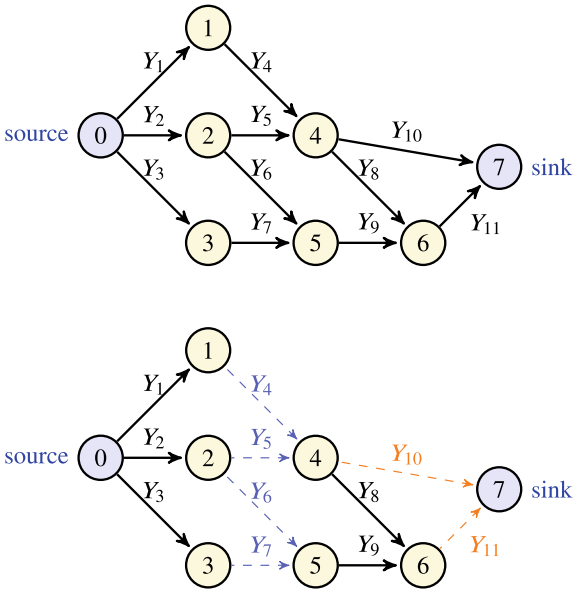

Fig. 1 Upper panel: a directed network with 11 links. Lower panel: two selected minimal cuts $\mathcal{L}_1 = \{4, 5, 6, 7\}$ (in light blue) and $\mathcal{L}_2 = \{10, 11\}$ (in orange) for this network

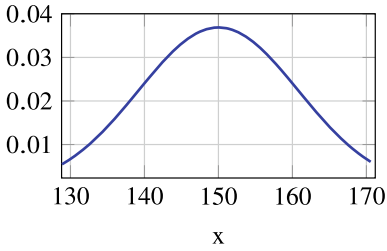

Fig. 2 Estimated density for the shortest path example

In the network of Fig. 1, there are six different directed paths from the source to the sink, each one being defined by a sequence of arcs. They are $\mathcal{P}_1 = \{1, 4, 10\}$, $\mathcal{P}_2 = \{1, 4, 8, 11\}, \mathcal{P}_3 = \{2, 5, 10\}, \mathcal{P}_4 = \{2, 5, 8, 11\}, \mathcal{P}_5 = \{2, 6, 9, 11\}$, and $\mathcal{P}_6 = \{3, 7, 9, 11\}$. The length of path p is $L_p = \sum_{j \in \mathcal{P}_p} Y_j$ and the length of the shortest path is

$$X = h(\boldsymbol{Y}) = \min_{1 \le p \le 6} L_p = \min_{1 \le p \le 6} \sum_{j \in \mathcal{P}_p} F_j^{-1}(U_j). \tag{9}$$

For our experiments, we assume that Y_j is normal with mean $\mu_j = 10j$ and standard deviation $\sigma_j = j$ (to make things simple). The probability of negative arc lengths is negligible. (To be mathematically cleaner, we can truncate the normal density to $[0, \infty)$, but it makes no visible difference in the numerical results.) We estimate the density of X over $[a, b] = [128.8, 171.2]$, which covers about 95% of the density. This density is shown in Fig. 2. It is close to a normal distribution, which is not surprising because all the Y_j are normal.

For the CDE, we select a directed minimal cut $\mathcal{L}$ between the source and the sink, and we condition on $\mathcal{G} = \{Y_j,\ j \notin \mathcal{L}\}$, similarly as for the SAN example in [17]. If $P_j + Y_j$ is the length of the shortest path that goes through arc j for $j \in \mathcal{L}$, then conditional on $\mathcal{G}$, each P_j is known and the conditional cdf of X is

$$F(x \mid \mathcal{G}) = \mathbb{P}\left[X \le x \mid \{P_j : j \in \mathcal{L}\}\right] = 1 - \prod_{j \in \mathcal{L}} (1 - F_j(x - P_j)). \tag{10}$$

If the Y_j's for $j \in \mathcal{L}$ are continuous variables, then the conditional density

$$f(x \mid \mathcal{G}) = \frac{\mathrm{d}}{\mathrm{d}x} F(x \mid \mathcal{G}) = \sum_{j \in \mathcal{L}} f_j(x - P_j) \prod_{l \in \mathcal{L}, l \ne j} (1 - F_l(x - P_l))$$

is an unbiased density estimator. In our numerical experiments, we try the two cuts $\mathcal{L}_1$ and $\mathcal{L}_2$ shown on the lower panel of Fig. 1.

For the LRDE we notice that $h(\boldsymbol{Y})$ is the minimum over the lengths of six possible paths. These lengths, in turn, are simple sums of several of the Y_j, so we have $h(c\boldsymbol{Y}) = ch(\boldsymbol{Y})$ for any constant $c > 0$. Therefore, with the change of variables $\varphi_x(\boldsymbol{z}) = x\,\boldsymbol{z}$ one obtains that $h(\varphi_x(\boldsymbol{z}))/x = h(\varphi_x(\boldsymbol{z})/x) = h(\boldsymbol{z})$ is independent of x.

For $x > 0$ this leads to the LRDE

$$\hat{f}_{\text{lrde}}(x) = \mathbb{I}[h(\boldsymbol{Y}) \le x]\; x^{-1} \left(-\sum_{j=1}^{s} (Y_j - \mu_j) Y_j \sigma_j^{-2} + s \right). \tag{11}$$

For GLR, the estimator in Proposition 4 does not apply to this example, because for the function h given in (9), for any choice of j, the required derivatives do not always exist. For the GLR-U, we want to find a subset of indices and a function g that satisfy the required conditions. In particular, $\tilde{g}$ must be a one-to-one continuous map between the selected inputs U_j and a selected subset of the path lengths L_p, so that the latter subset is sufficient to determine X and the Jacobian $J_g(\cdot; x)$ of this mapping is invertible. Note that the six path lengths are not independent: we have $L_1 + L_4 = L_2 + L_3$. But after removing one of these four paths, there is no linear relationship between any of the five L_p's that remain. Then we must select five input variables U_j for which the mapping g between those selected U_j's and the five L_p's is one-to-one when the other U_j's are fixed. There are several possibilities for the selection of these five indexes j for the inputs, each one leading to a different estimator. We will try two of them in our experiments, namely $\mathcal{J}_1 = \{1, 2, 3, 6, 8\}$ and $\mathcal{J}_2 = \{5, 7, 8, 10, 11\}$. Assuming that we remove the path $\mathcal{P}_4$ and select $\mathcal{J}_1$, we obtain $g(\boldsymbol{U}) = (g_1(\boldsymbol{U}), \ldots, g_5(\boldsymbol{U}))^t$ where $g_p(\boldsymbol{U}) = L_p$ for $p = 1, 2, 3$ and $g_p(\boldsymbol{U}) = L_{p+1}$ for $p = 4, 5$, and the Jacobian is computed by interpreting the $g_p(\boldsymbol{U})$ as functions of U_1, U_2, U_3, U_6, U_8 alone, with the other U_j's fixed. The GLR-U density estimator in (8) turns out to be

$$G_1(\boldsymbol{U}; x) = -\mathbb{I}[h(\boldsymbol{Y}) \le x] \sum_{j=1}^{3} j^{-1} \Phi^{-1}(U_j) = -\mathbb{I}[h(\boldsymbol{Y}) \le x] \sum_{j=1}^{3} j^{-2} (Y_j - 10j) \tag{12}$$

for $\mathcal{J}_1$ and

$$G_1(\boldsymbol{U}; x) = -\mathbb{I}[h(\boldsymbol{Y}) \le x] \sum_{j=10}^{11} j^{-1} \Phi^{-1}(U_j) = -\mathbb{I}[h(\boldsymbol{Y}) \le x] \sum_{j=10}^{11} j^{-2} (Y_j - 10j) \tag{13}$$

for $\mathcal{J}_2$, where Φ denotes the standard normal cdf.

Table 2 summarizes our numerical results for this example, for all the methods. It reports $-\log_2(\text{MISE})$ for $n = 2^{20}$ as well as the convergence rate exponent β for MC and its (noisy) estimate $\hat{\beta}$ for RQMC.

We find that the CDE combined with RQMC outperforms all other methods by a wide margin. Compared with the KDE with MC (the traditional approach), it reduces the MISE for $n = 2^{20}$ by a factor of about $2^{25} \approx 32$ millions. The orange cut $\mathcal{L}_2$ does better than the blue cut $\mathcal{L}_1$, especially for plain MC. This could appear surprising, because $\mathcal{L}_2$ has fewer arcs, but the explanation is that the two arcs of $\mathcal{L}_2$ have a much larger variance, so it pays off to hide them. Generally speaking, we want to select a

Table 2 Estimated values of $-\log_2(\text{MISE})$ with $n = 2^{20}$ points and estimated MISE rate $\hat{\beta}$ for various methods, for the shortest path example. The 21.3 entry (for example) means that for the KDE with MC and $n = 2^{20}$ points, we have MISE $\approx 2^{-21.3}$

Method	MC		RQMC	
	$-\log_2(\text{MISE})$	β	$-\log_2(\text{MISE})$	$\hat{\beta}$
KDE	21.3	0.8	25.7	0.96
CDE (blue cut)	24.7	1.0	45.6	2.12
CDE (orange cut)	29.1	1.0	46.5	1.66
LRDE	20.2	1.0	27.8	1.38
GLR-U in (12)	15.4	1.0	23.2	1.29
GLR-U in (13)	21.5	1.0	29.6	1.35

conditioning that hides (or integrate out) variables that capture as much variance as possible. (For the blue cut, the noise in the linear regression model and the estimate $\hat{\beta}$ appears quite significant.)

We also observe a significant difference of performance between the two choices of input variables for GLR-U. With $\mathcal{I}_2$, the performance is better than for the KDE, whereas for $\mathcal{I}_1$ it is worse. This shows that the choice of input variables may have a significant impact on the performance in general. Note that $\mathcal{I}_2$ contains input variables that have much more variance than $\mathcal{I}_1$. By comparing (12) and (13), we can see why the second estimator has less variance: the terms in the sum that multiplies the indicator have larger constants in the denominator, and therefore a smaller variance. In some sense, the GLR-U estimator integrates out part of the variance contained in the selected input variables, so it makes sense to select a subset of input variables that captures more of the variance.

The LRDE has a larger MISE than the KDE with $n = 2^{20}$ MC samples, but it beats the KDE when using RQMC. It also performs better than GLR-U for one choice of inputs and worse for the other choice.

With the same network, we now consider a slightly different problem. We assume that the Y_j's are random link capacities instead of random lengths, and we want to estimate the density of the maximum flow that can be sent from the source to the sink. This maximum flow $h(\boldsymbol{Y})$ is equal to the capacity of the minimal directed cut having the smallest capacity. Here, we assume that Y_j is normal with mean $\mu_j = 10$ and standard deviation $\sigma_j = 1$ for $j < 10$ and normal with mean $\mu_j = 20$ and standard deviation $\sigma_j = 4$ for $j = 10$ and 11. For the CDE, if we take $\mathcal{G}$ as in the previous case, the distribution of X conditional on $\mathcal{G}$ typically has a probability mass at some point. For instance, if $\mathcal{L} = \mathcal{L}_1$, then after the conditioning, $Y_{10} + Y_{11}$ is known and there is a positive probability that this is the value of the maximum flow. As a result, the conditional cdf is sometimes discontinuous and the CDE is no longer an unbiased density estimator. This motivates the use of LRDE for this example.

Similarly as in the previous example, $h(\boldsymbol{Y})$ is the minimum over several simple sums of Y_j's, so multiplying all Y_j's by a positive constant multiplies the maximum

Table 3 Values of the $\log_2(\text{MISE})$ estimated with $n = 2^{20}$ points and the estimated MISE rate $\hat{\beta}$ for various methods for the maximum flow example

Method	MC	RQMC	
	$-\log_2(\text{MISE})$	$-\log_2(\text{MISE})$	$\hat{\beta}$
KDE	18.3	19.7	0.86
LRDE	18.7	23.6	1.23
GLR-U	17.7	23.2	1.26

flow $h(\boldsymbol{Y})$ by the same constant. Therefore, the change of variables $\varphi_x(z) = xz$ can be used again and provides the exact same LRDE as in (11), but with the modified h, μ_j, and σ_j.

For GLR-U, the construction is similar as for the previous example, except that we select a subset of minimal cuts with independent capacities instead of a subset of paths. There are hundreds of thousands of ways of selecting the subset of minimal cuts. We tried a few of them and obtained the best results by selecting the set of cuts: {{10, 11}, {1, 2, 7}, {1, 2, 9}, {1, 2, 11}, {1, 5, 9}, {2, 3, 4}, {4, 5, 11}, {8, 9, 10}, {2, 3, 8, 10}, {6, 7, 8, 10}} and then hiding Y_{11}. This gives the estimator

$$G(\boldsymbol{U}; x) = -\mathbb{I}[h(\boldsymbol{Y}) \le x]\,((Y_1 - 10) + (Y_4 - 10) + (Y_{10} - 20)/16)\,.$$

Numerical results for the KDE, LRDE, and GLR-U for this example are given in Table 3. In terms of MISE, under MC, the LRDE performs better than GLR-U and slightly better than the KDE, but not much. However, RQMC improves the MISE for $n = 2^{20}$ by a factor of about 30 for the LRDE, a bit more for GLR-U, and about 3 for the KDE. The combination of LRDE or GLR-U with RQMC also improves the convergence rate $\hat{\beta}$.

9 Conclusion

We discussed and compared several recent developments regarding density estimation for simulation models, with Monte Carlo and quasi-Monte Carlo methods. Most of these methods provide unbiased density estimators and some of them are also RQMC-friendly, in which case their MISE can converge at a faster rate than the canonical rate of $O(1/n)$ as a function of the sample size n. For the classical density estimators in statistics, in contrast, the MISE converges at a slower rate than $O(1/n)$. In our numerical example (and several other experiments not reported here), the CDE combined with RQMC was by far the best performer. However, for some types of problems it may be difficult to apply, and then one can rely on one of the alternatives. In future work, these density estimators should be adapted, implemented, and compared for a larger variety of Monte Carlo applications for which density estimates are useful.

Acknowledgements This work has been supported by a NSERC Discovery Grant and an IVADO Grant to P. L'Ecuyer. F. Puchhammer was also supported by Spanish and Basque governments fundings through BCAM (ERDF, ESF, SEV-2017-0718, PID2019-108111RB-I00, PID2019-104927GB-C22, BERC 2018e2021, EXP. 2019/00432, ELKARTEK KK-2020/00049), and the computing infrastructure of i2BASQUE academic network and IZO-SGI SGIker (UPV).

References

1. Asmussen, S.: Conditional Monte Carlo for sums, with applications to insurance and finance. Ann. Actuar. Sci. **12**(2), 455–478 (2018)
2. Asmussen, S., Glynn, P.W.: Stochastic Simulation. Springer, New York (2007)
3. Ben Abdellah, A., L'Ecuyer, P., Owen, A., Puchhammer, F.: Density estimation by randomized Quasi-Monte Carlo. SIAM J. Uncertain. Quantif. **9**(1), 280–301 (2021)
4. Dick, J., Pillichshammer, F.: Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration. Cambridge University Press, Cambridge, U.K. (2010)
5. Fu, M., Hu, J.Q.: Conditional Monte Carlo: Gradient Estimation and Optimization Applications. Kluwer Academic, Boston (1997)
6. Fu, M.C.: Sensitivity analysis in Monte Carlo simulation of stochastic activity networks. In: Alt, F.B., Fu, M.C., Golden, B.L. (eds.) Perspectives in Operations Research, Operations Research/Computer Science Interfaces Series, pp. 351–366. Springer, Boston (2006)
7. Glynn, P.W.: Likelihood ratio gradient estimation: an overview. In: Proceedings of the 1987 Winter Simulation Conference, pp. 366–375. IEEE Press, Piscataway, NJ (1987)
8. Glynn, P.W., L'Ecuyer, P.: Likelihood ratio gradient estimation for regenerative stochastic recursions. Adv. Appl. Probab. **27**, 1019–1053 (1995)
9. Goda, T., Suzuki, K.: Recent Advances in Higher Order Quasi-Monte Carlo Methods, pp. 69–102. De Gruyter (2019)
10. Laub, P.J., Salomone, R., Botev, Z.I.: Monte Carlo estimation of the density of the sum of dependent random variables. Mathematics and Computers in Simulation **161**, 23–31 (2019)
11. L'Ecuyer, P.: A unified view of the IPA, SF, and LR gradient estimation techniques. Manag. Sci. **36**(11), 1364–1383 (1990)
12. L'Ecuyer, P.: Quasi-Monte Carlo methods with applications in finance. Financ. Stoch. **13**(3), 307–349 (2009)
13. L'Ecuyer, P.: Randomized Quasi-Monte Carlo: an introduction for practitioners. In: Glynn, P.W., Owen, A.B. (eds.) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC 2016, pp. 29–52. Springer, Berlin (2018)
14. L'Ecuyer, P., Munger, D.: Algorithm 958: lattice builder: a general software tool for constructing rank-1 lattice rules. ACM Trans. Math. Softw. **42**(2), Article 15 (2016)
15. L'Ecuyer, P., Perron, G.: On the convergence rates of IPA and FDC derivative estimators. Oper. Res. **42**(4), 643–656 (1994)
16. Puchhammer, F., L'Ecuyer, P.: Likelihood Ratio Density Estimation for Simulation Models (2022). Submitted
17. L'Ecuyer, P., Puchhammer, F., Ben Abdellah, A.: Monte Carlo and Quasi-Monte Carlo density estimation via conditioning. INFORMS J. Comput. (2022). To appear. See https://doi.org/10.1287/ijoc.2021.1135
18. Lei, L., Peng, Y., Fu, M.C., Hu, J.Q.: Applications of generalized likelihood ratio method to distribution sensitivities and steady-state simulation. Discret. Event Dyn. Syst. **28**(1), 109–125 (2018)
19. Lemieux, C., Cieslak, M., Luttmer, K.: RandQMC User's Guide: A Package for Randomized Quasi-Monte Carlo Methods in C (2004). Software user's guide, available at http://www.math.uwaterloo.ca/~clemieux/randqmc.html

20. Niederreiter, H.: Random number generation and Quasi-Monte Carlo methods. In: SIAM CBMS-NSF CBMS Regional Conference Series in Mathematics, vol.63. SIAM (1992)
21. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**(3), 1065–1076 (1962)
22. Peng, Y., Fu, M.C., Heidergott, B., Lam, H.: Maximum likelihood estimation by Monte Carlo simulation: towards data-driven stochastic modeling. Oper. Res. **68**(6), 1896–1912 (2020)
23. Peng, Y., Fu, M.C., Hu, J.Q., Heidergott, B.: A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. Oper. Res. **66**(2), 487–499 (2018)
24. Peng, Y., Fu, M.C., Hu, J.Q., L'Ecuyer, P., Tuffin, B.: Generalized likelihood ratio method for stochastic models with uniform random numbers as inputs (2021). Submitted manuscript
25. Peng, Y., Fu, M.C., Hu, J.Q., L'Ecuyer, P., Tuffin, B.: Variance reduction for generalized likelihood ratio method by conditional Monte Carlo and randomized Quasi-Monte Carlo. J. Manage. Sci. Eng. (2022). To appear
26. Scott, D.W.: Multivariate Density Estimation, 2nd edn. Wiley, New York (2015)