

Article

Improving Kernel Methods for Density Estimation in Random Differential Equations Problems

Juan Carlos Cortés López *  and Marc Jornet Sanz 

Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València,
46022 Valencia, Spain; marjorsa@doctor.upv.es

* Correspondence: jccortes@imm.upv.es

Received: 30 May 2020; Accepted: 18 June 2020; Published: 18 June 2020

Abstract: Kernel density estimation is a non-parametric method to estimate the probability density function of a random quantity from a finite data sample. The estimator consists of a kernel function and a smoothing parameter called the bandwidth. Despite its undeniable usefulness, the convergence rate may be slow with the number of realizations and the discontinuity and peaked points of the target density may not be correctly captured. In this work, we analyze the applicability of a parametric method based on Monte Carlo simulation for the density estimation of certain random variable transformations. This approach has important applications in the setting of differential equations with input random parameters.

Keywords: probability density estimation; Monte Carlo simulation; parametric method; random differential equation

MSC: 62G07; 65C05; 34F05

1. Introduction

Given an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \rightarrow \mathbb{R}$, its probability density function f_X is defined as the Borel measurable and non-negative function that satisfies $\mathbb{P}[X \in B] = \int_B f_X(x) dx$ for any Borel set $B \subseteq \mathbb{R}$ [1] Ch. 2. In other words, $f_X(x) = \frac{d\mathbb{P}_X(x)}{dx}$ is the Radon–Nikodym derivative of the probability law $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ with respect to the Lebesgue measure. When the density function exists, the random variable (its probability law) is said to be absolutely continuous. Examples of absolutely continuous distributions are Uniform, Triangular, Normal, Gamma, etc.

When the probability law of X is unknown in explicit form but realizations of X can be drawn, a kernel estimation may be used to reconstruct the probability density function of X . It is a non-parametric method that uses a finite data sample from X . If x_1, \dots, x_M are independent realizations of X , then the kernel density estimator takes the form $\hat{f}_X^M(x) = \frac{1}{Mb} \sum_{i=1}^M K\left(\frac{x-x_i}{b}\right)$, where K is the kernel function and b is the bandwidth, see [2].

Kernel density estimation may present slow convergence rate with the number of realizations and may smooth out discontinuity and non-differentiability points of the target density. For certain random variable transformations $X = g(U, V)$, where U and V are independent random quantities and g is a deterministic real function, we will see that there is an alternative estimation method of parametric nature that is based on Monte Carlo simulation. Such approach has been used for certain random differential equations problems, see [3–7]. However, a detailed comparison with kernel density estimation, both theoretically and numerically, has not been performed yet. This detailed comparison is the novelty of the present paper. We will demonstrate that the parametric method is more efficient and correctly captures density features. Numerical experiments will illustrate these improvements.

As it shall be seen later (see expression (2)), our approach relies upon the representation of the probability density via the expectation of a transformation derived through the application of the so-called Random Variable Transformation (RVT) method [1] (also termed “change of variable formula”, see [8]). This method depends on the computation of invertible functions with Jacobian determinant. We point out that, when the number of random model parameters is large, it is convenient to apply competitive techniques to carry out computations. In this regard, we here highlight recent contributions that address this important issue in the machine learning setting. In [8], the authors applied real-valued non-volume preserving (real NVP) transformations to estimate the density when reconstructing natural images on datasets through sampling, log-likelihood evaluation, and latent variable manipulations. Recently, in [9], the real NVP was improved by combining the so called flow-based generative models. In the machine learning context, these are methods based upon the application of the “change of variable formula”—together with Knothe–Rosenblatt rearrangement.

2. Method

Let $X = g(U, V)$ be a random variable, where U is a random variable, V is a random variable/vector, and g is a deterministic real function. The aim is to estimate the probability density function of X .

A kernel density estimator takes the form

$$\hat{f}_X^M(x) = \frac{1}{Mb} \sum_{i=1}^M K\left(\frac{x - g(u_i, v_i)}{b}\right), \quad (1)$$

where u_1, \dots, u_M and v_1, \dots, v_M are independent realizations of U and V , respectively.

Let us see an alternative method when U and V are independent. Suppose that U has an explicit density function f_U . Suppose also that $g(\cdot, V)$ is invertible for all V , where $h(\cdot, V)$ is the inverse. Then,

$$f_X(x) = \mathbb{E}[f_U(h(x, V)) |\partial_1 h(x, V)|], \quad (2)$$

where ∂_1 is the partial derivative with respect to the first variable and \mathbb{E} is the expectation operator. Notice that we are not requiring V to have a probability density.

Indeed, if $\mathbb{P}_V = \mathbb{P} \circ V^{-1}$ denotes the probability law of V , then the following chain of equalities holds:

$$\begin{aligned} \mathbb{P}[g(U, V) \leq x] &= \int_{\mathbb{R}} \mathbb{P}[g(U, V) \leq x | V = v] \mathbb{P}_V(dv) \\ &= \int_{\mathbb{R}} \mathbb{P}[g(U, v) \leq x | V = v] \mathbb{P}_V(dv) \\ &= \int_{\mathbb{R}} \mathbb{P}[g(U, v) \leq x] \mathbb{P}_V(dv) \\ &= \int_{\mathbb{R}} \int_{-\infty}^x f_{g(U, v)}(u) du \mathbb{P}_V(dv) \\ &= \int_{\mathbb{R}} \int_{-\infty}^x f_U(h(u, v)) |\partial_1 h(u, v)| du \mathbb{P}_V(dv) \\ &= \int_{-\infty}^x \int_{\mathbb{R}} f_U(h(u, v)) |\partial_1 h(u, v)| \mathbb{P}_V(dv) du \\ &= \int_{-\infty}^x \mathbb{E}[f_U(h(u, V)) |\partial_1 h(u, V)|] du. \end{aligned}$$

We have used the independence, the transformed density function [1, Section 6.2.1], and Fubini's theorem.

To estimate the expectation from (2) when the probability law of V is unknown in closed form but realizations of it can be drawn, Monte Carlo simulation may be conducted. For example, the crude Monte Carlo method estimates $f_X(x)$ by using

$$\hat{f}_X^M(x) = \frac{1}{M} \sum_{i=1}^M f_U(h(x, v_i)) |\partial_1 h(x, v_i)|, \quad (3)$$

where v_1, \dots, v_M are independent realizations of V [10, pp. 53–54].

Notice that the structure and the complexity of this estimator are very similar to a kernel one. For (1), one needs realizations $u_1, \dots, u_M, v_1, \dots, v_M$, while (3) only uses v_1, \dots, v_M . Hence, the complexity of drawing realizations of V affects (1) and (3) in exactly the same way. On the other hand, (1) evaluates K and g , while (3) evaluates f_U and h .

By the central limit theorem, the root mean square error of the Monte Carlo estimate (3) of the mean is $\sigma/\sqrt{M} = \mathcal{O}(M^{-1/2})$, where σ is the standard deviation of the random quantity within the expectation (2) (which is assumed to be finite). Variance reduction methods may be applied, such as antithetic or control variates, see [11]. By contrast, the root mean square error of the kernel estimate (1) is $\mathcal{O}(M^{-r})$, $r < 1/2$ [2].

The density estimation of $f_X(x)$ thus becomes a parametric problem, as we are estimating the expectation parameter from a population distribution. This is in contrast to kernel density estimation, which is non-parametric because it reconstructs a distribution. Moreover, the method presented here acts pointwise, so discontinuities and non-differentiability points are correctly captured, without smoothing them out.

3. Some Numerical Experiments

In this section, we compare the new parametric method with kernel density estimation numerically. We use the software Mathematica® (Wolfram Research, Inc, Mathematica, Version 12.0, Champaign, IL, USA, 2019).

Example 1. Let $X = U + V$, where $U \sim \text{Normal}(0, 1)$ and $V \sim \text{Uniform}(-1, 1)$ are independent random variables. Obviously, the exact density function of X is known via convolution: $f_X(x) = \int_{-\infty}^{\infty} f_U(u - v) f_V(v) dv$. However, this example allows for comparing the new parametric method with kernel density estimation by calculating exact errors. With $g(U, V) = U + V$ and $h(U, V) = U - V$, expression (2) becomes $f_X(x) = \mathbb{E}[f_U(x - V)]$. From M realizations, we conduct a kernel density estimation (Gaussian kernel with Silverman's rule to determine the bandwidth, $b = 0.9 \times \min\{\hat{\sigma}, \text{IQR}/1.34\} \times M^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of the sample and IQR is the interquartile range of the sample), crude Monte Carlo simulation on $\mathbb{E}[f_U(x - V)]$, antithetic variates method ($M/2$ realizations of V and the sample is completed by changing signs), and control variates method (the control variable here is V). Let $\hat{f}_X^M(x)$ be the density estimate of $f_X(x)$. We consider the error measure $\delta^M = \int_{-\infty}^{\infty} (f_X(x) - \hat{f}_X^M(x))^2 dx$. Numerically, this integral is computed by fixing a large interval $[-R, R]$ and performing Gauss–Legendre quadrature on it. As δ^M is random, we better consider $\epsilon^M = (\mathbb{E}[\delta^M])^{1/2}$. This expectation is estimated by repeating the density estimate several times (we did so 20 times). Figure 1 reports the estimated error $\hat{\epsilon}^M$ in log-log scale, for different orders M growing geometrically. Observe that the parametric approach is more efficient than the kernel method. Variance reduction methods also allow for lowering the error. The lines corresponding to the three parametric methods have slope $-1/2$ approximately, due to the well-known decrease of the Monte Carlo error.

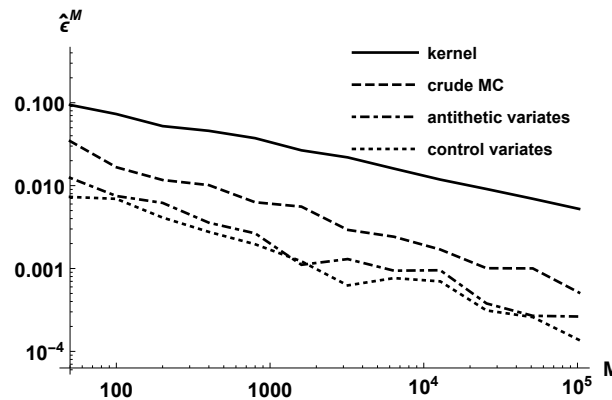


Figure 1. Density estimation using a kernel method, parametric crude Monte Carlo (MC) simulation, antithetic variates and control variates methods. Estimated error $\hat{\epsilon}^M$ in log-log scale. This figure corresponds to Example 1.

For example, to achieve an error $\epsilon^M < 0.01$, the kernel density estimation requires about 25,000 realizations, while the parametric crude Monte Carlo estimation needs 400 realizations. The antithetic method decreases the number of required realizations to 50–100, and the control variates approach does so even more, to less than 50.

In Table 1, we present the timings in seconds for achieving root mean square errors less than 0.01 and 0.001. We work at the density location $x = 1$. Observe that the kernel density estimation is the least efficient method, especially as we require smaller errors.

Table 1. Time in seconds to achieve root mean square error less than 0.01 and 0.001. We employ a kernel method, parametric crude Monte Carlo (MC) simulation, antithetic variates and control variates methods. This table corresponds to Example 1.

Time	Kernel	Crude MC	Antithetic Variates	Control Variates
error 0.01	0.063	0.00089	0.00023	0.00015
error 0.001	4.2	0.13	0.012	0.0057

Example 2. Let $X = UV$, where $U \sim \text{Normal}(0,1)$ and $V = P + \sum_{i=1}^{12} Q_i$, being $P \sim \text{Poisson}(1)$ and $Q_i \sim \text{Triangular}(0.1, 0.105)$. All of the random variables are assumed to be independent. With $g(U, V) = UV$ and $h(U, V) = U/V$, expression (2) is $f_X(x) = \mathbb{E}[f_U(\frac{x}{V})\frac{1}{V}]$. Notice that, despite V being discrete, X has a density. This expectation cannot be computed via quadratures, due to the large dimension of the random space. We employ our parametric method using crude Monte Carlo simulation. In Figure 2, first panel, we depict the estimated density function $\hat{f}_X^M(x)$ using M realizations. For $M = 10,000$ and $M = 30,000$, we perceive good agreement between the approximations. This is in contrast to the second panel of Figure 2, where the kernel density estimation does not show convergence yet.

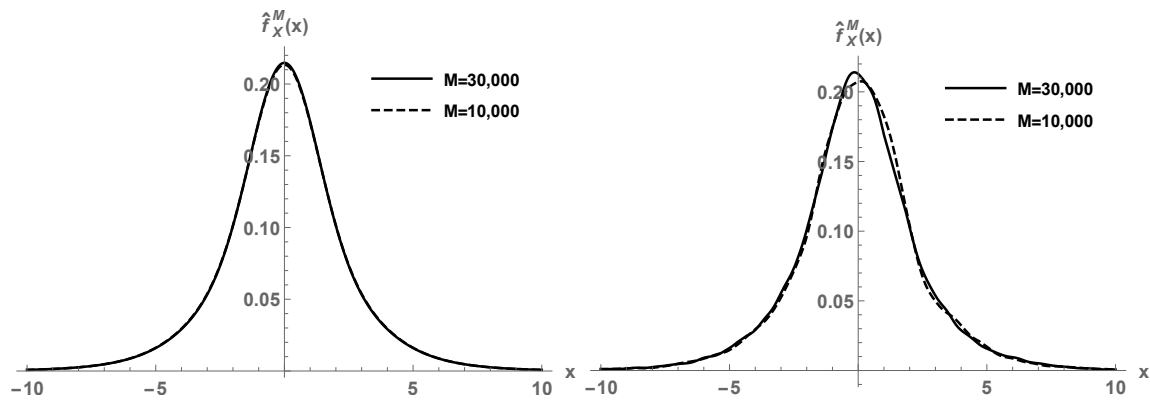


Figure 2. Density estimation $\hat{f}_X^M(x)$ using parametric crude Monte Carlo simulation (first panel) and kernel density estimation (second panel) with M realizations. This figure corresponds to Example 2.

Example 3. Consider the same setting of Example 2 but $U \sim \text{Uniform}(-1,1)$ (now f_U is not continuous on \mathbb{R}). In Figure 3, first plot, we show the estimated density function $\hat{f}_X^M(x)$ by the parametric crude Monte Carlo method using M realizations. The discontinuity points of the target density f_X are correctly captured. Indeed, the method acts pointwise in x , so any feature of f_X will be correctly identified, no matter how rare it is. By contrast, the kernel density estimation smooths out the discontinuities, see the second panel of the figure.

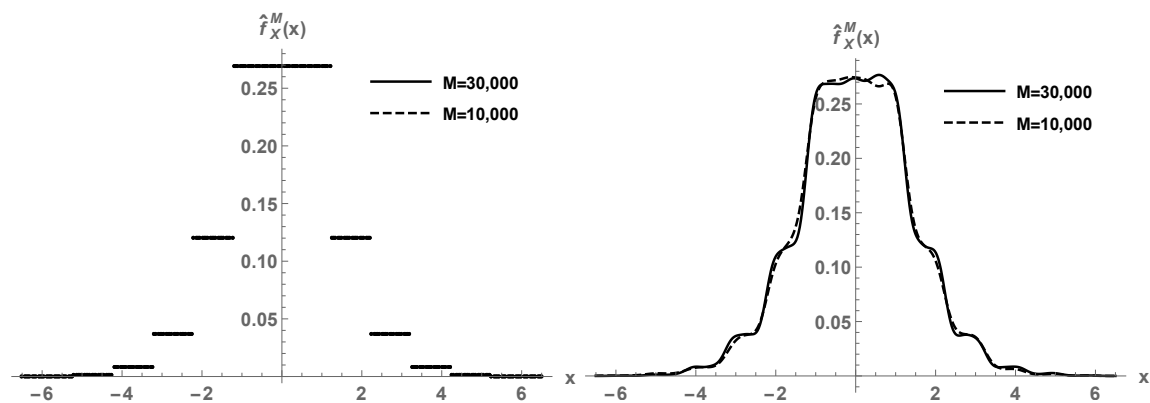


Figure 3. Density estimation $\hat{f}_X^M(x)$ using parametric crude Monte Carlo simulation (first panel) and kernel density estimation (second panel) with M realizations. This figure corresponds to Example 3.

4. Application to Random Differential Equations

A random differential equation problem considers some of the terms in the system (input coefficients, forcing term, initial conditions, etc.) as random variables or stochastic processes [1]. This is due to errors in measurements of data when trying to model a physical phenomenon, which introduces a variability in the parameters estimation. The solution to the system is then a stochastic process. Its deterministic trajectories are not the main concern. Instead, the study of the statistical content of the solution is the main goal. An important aim is to compute its probability density function. The key idea is that the solution is a transformation of the input random parameters; therefore, the probability density may be estimated as described in this paper whenever the solution is given in closed form. In the notation used in this paper, U and V would denote the input random parameters of the system, and g would be the transformation mapping that relates the output with the inputs. A specific input parameter is selected as U , with respect to which the mapping g is easily invertible to obtain h .

We consider the Ricatti random differential equation

$$\begin{cases} \dot{X}(t) = -AX(t)^2, & t \in [0, \infty), \\ X(0) = X_0, \end{cases} \quad (4)$$

where $A > 0$ and X_0 are independent random variables. The solution to (4) is given by

$$X(t) = \frac{X_0}{1 + AX_0 t}.$$

By taking $U = X_0$, the density function of $X(t)$ is given by

$$f_{X(t)}(x) = \mathbb{E} \left[f_{X_0} \left(\frac{x}{1 - Axt} \right) \frac{1}{(1 - Axt)^2} \right].$$

From M realizations of A , say $A^{(1)}, \dots, A^{(M)}$, the expectation is estimated via crude Monte Carlo simulation:

$$f_{X(t)}(x) \approx \frac{1}{M} \sum_{i=1}^M f_{X_0} \left(\frac{x}{1 - A^{(i)}xt} \right) \frac{1}{(1 - A^{(i)}xt)^2}. \quad (5)$$

As seen in the previous sections, this procedure is more efficient and certain than kernel density estimation, expressed as

$$f_{X(t)}(x) \approx \frac{1}{Mb} \sum_{i=1}^M K \left(\frac{x - X_0^{(i)} / (1 + A^{(i)}X_0^{(i)}t)}{b} \right),$$

where the superscript i denotes the i -th realization, $i = 1, \dots, M$.

For a specific numerical example, let us fix $A \sim \text{Uniform}(-1, 1)$ and $X_0 \sim \text{Uniform}(0.1, 0.13)$. Figure 4, first panel, plots (5) for a certain number of realizations M for A , for time $t = 0.3$. A comparison is conducted with kernel density estimation with a Gaussian kernel (second panel). It must be appreciated that the parametric Monte Carlo method (5) correctly captures the non-differentiability points of the target density.

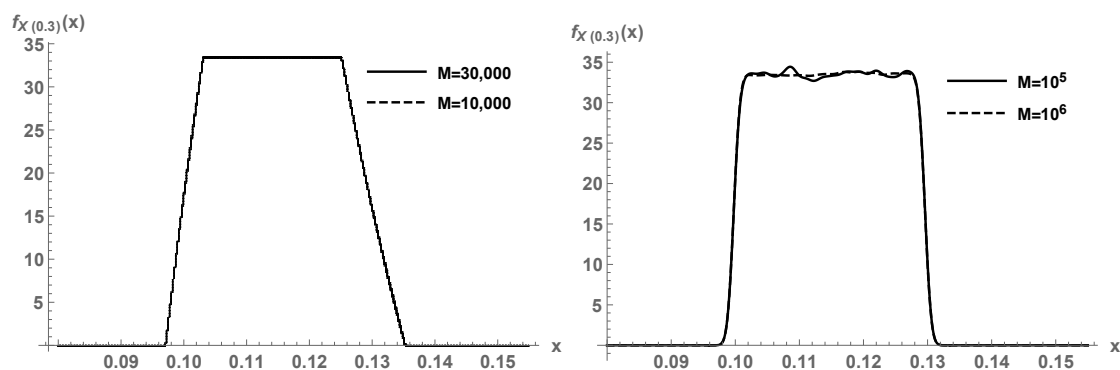


Figure 4. First panel: Estimated density function $f_{X(0.3)}(x)$ with parametric Monte Carlo for $M = 10,000$ and $M = 30,000$. Second panel: Kernel density estimate with Gaussian kernel for $M = 10^5$ and $M = 10^6$.

For another example, let us consider the damped pendulum differential equation with uncertainties:

$$\begin{cases} \ddot{X}(t) + 2\omega_0\zeta\dot{X}(t) + \omega_0^2X(t) = Y(t), & t \in [0, T], \\ X(0) = X_0, \\ \dot{X}(0) = X_1, \end{cases} \quad (6)$$

where $T > 0$, $\omega_0 \neq 0$, $\xi \neq 0$ and $\xi^2 < 1$ (underdamped case) are constant, the initial position X_0 and the initial velocity X_1 are random variables, and the source/forcing term $Y(t)$ is a stochastic process [1, Example 7.2]. The solution is given by

$$X(t) = \left(\frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t) \right) X_0 + \frac{e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\omega_1} X_1 + \int_0^t p(t-s) Y(s) ds, \quad (7)$$

where $\omega_1 = \omega_0 \sqrt{1-\xi^2} \neq 0$ and $p(t) = \frac{1}{\omega_1} e^{-\xi \omega_0 t} \sin(\omega_1 t)$. In [4], the density function of the response $X(t)$ is expressed in terms of an expectation (2), which is estimated by means of parametric crude Monte Carlo simulation due to the large dimension of the random space. By taking $U = X_0$, one of the formulas derived is

$$\begin{aligned} f_{X(t)}(x) &= \mathbb{E} \left[f_{X_0} \left(\frac{x - \frac{e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\omega_1} X_1 - \int_0^t p(t-s) Y(s) ds}{\frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t)} \right) \right] \frac{1}{\left| \frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t) \right|} \\ &\approx \frac{1}{M} \sum_{i=1}^M f_{X_0} \left(\frac{x - \frac{e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\omega_1} X_1^{(i)} - \int_0^t p(t-s) Y^{(i)}(s) ds}{\frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t)} \right) \frac{1}{\left| \frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t) \right|}, \quad (8) \end{aligned}$$

where the superscript i denotes the i -th realization, $i = 1, \dots, M$. Compare (8) with a kernel density estimation

$$f_{X(t)}(x) \approx \frac{1}{Mb} \sum_{i=1}^M K \left(\frac{x - \left(\frac{\xi e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\sqrt{1-\xi^2}} + e^{-\omega_0 \xi t} \cos(\omega_1 t) \right) X_0^{(i)} - \frac{e^{-\omega_0 \xi t} \sin(\omega_1 t)}{\omega_1} X_1^{(i)} - \int_0^t p(t-s) Y^{(i)}(s) ds}{b} \right).$$

The expressions and their complexities are very similar, but the convergence of (8) is faster with M , as justified in the previous sections.

Let us see a numerical example. We take the upper time $T = 1$, the damping ratio $\xi = 1/2$ and the natural frequency $\omega_0 = \pi/2$. Consider $X_0 \sim \text{Exponential}(3)$, $X_1 \sim \text{Binomial}(7, 0.31)$ and $Y(t) = \sum_{j=1}^{\infty} \frac{\sqrt{2}}{j\pi} \sin(tj\pi) \xi_j$. The series is understood in $L^2([0, 1] \times \Omega)$ and $\{\xi_j\}_{j=1}^{\infty}$ is a sequence of independent random variables with $\text{Uniform}(-\sqrt{3}, \sqrt{3})$ distribution. This is a Karhunen–Loève expansion [10, p. 47]. It is assumed that X_0 , X_1 , and Y are independent. By applying (8) with the sum of $Y(t)$ truncated to a finite-term series, we estimate the density function of $X(t)$. For example, in Figure 5, first panel, $f_{X(0.5)}(x)$ is plotted for orders of truncation $N = 4, 5$, and 6. Overlapping of the graphs is clearly appreciated. The strong oscillations of the density are perfectly captured because the method acts pointwise. In the second panel of the figure, the kernel density estimate with Gaussian kernel is plotted, with 10^6 realizations. Although there is apparent agreement between both panels, our parametric Monte Carlo method captures all the peaks while the kernel method smooths them out.

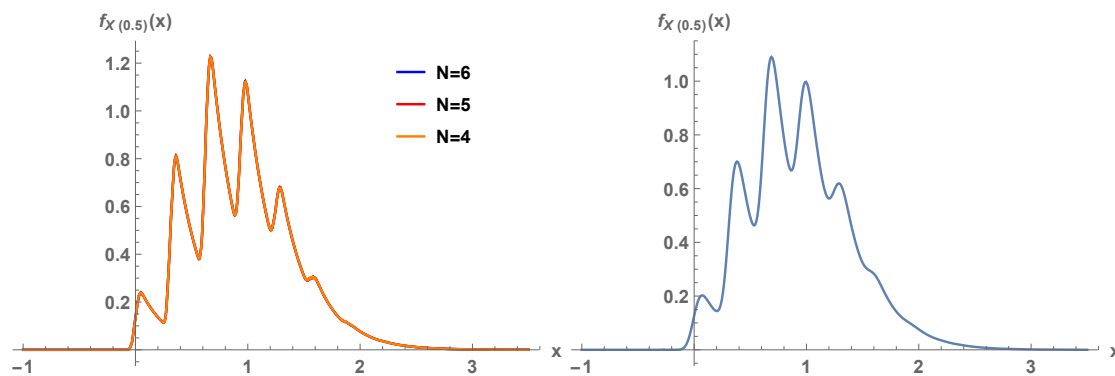


Figure 5. First panel: Estimated density function $f_{X(0.5)}(x)$ with parametric Monte Carlo for orders of truncation $N = 4, 5$ and 6 (see [4]). Second panel: Kernel density estimate with Gaussian kernel for $N = 6$.

Further applications of the methodology for random differential equations may be consulted in [3–7]. The present paper forms the theoretical and computational foundations of the methodology used in those recent contributions.

5. Conclusions

In this paper, we have been concerned with the density estimation of random variables. We have proposed an alternative to kernel density estimation for certain random variable transformations. The alternative is based on the estimation of an expectation parameter via Monte Carlo methods; therefore, it is of parametric nature and improves kernel methods in terms of efficiency. Furthermore, the method captures density features and does not smooth out discontinuity and non-differentiability points of the target density.

As shown here, the solution to some random differential equation problems is an explicit transformation of the input random parameters. The methodology proposed in this paper may be employed to estimate the density function of the closed-form stochastic solution parametrically, instead of relying on kernel estimation.

Author Contributions: Investigation, M.J.S.; Methodology and M.J.S.; Software, M.J.S.; Supervision, J.C.C.L.; Validation, J.C.C.L.; Visualization, J.C.C.L. and M.J.S.; Writing—original draft, M.J.S.; Writing—review and editing, J.C.C.L. and M.J.S.. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER UE) grant MTM2017–89664–P.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Soong, T.T. *Random Differential Equations in Science and Engineering*; Academic Press: New York, NY, USA, 1973.
2. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
3. Calatayud, J.; Cortés, J.-C.; Díaz, J.A.; Jornet, M. Constructing reliable approximations of the probability density function to the random heat PDE via a finite difference scheme. *Appl. Numer. Math.* **2020**, *151*, 413–424.
4. Calatayud, J.; Cortés, J.-C.; Jornet, M. The damped pendulum random differential equation: A comprehensive stochastic analysis via the computation of the probability density function. *Physica A* **2018**, *512*, 261–279.
5. Calatayud, J.; Cortés, J.-C.; Díaz, J.A.; Jornet, M. Density function of random differential equations via finite difference schemes: a theoretical analysis of a random diffusion-reaction Poisson-type problem. *Stochastics* **2019**, doi:10.1080/17442508.2019.1645849.
6. Calatayud, J.; Cortés, J.-C.; Dorini, F.A.; Jornet, M. Extending the study on the linear advection equation subject to stochastic velocity field and initial condition. *Math. Comput. Simul.* **2020**, *172*, 159–174.

7. Jornet, M.; Calatayud, J.; Le Maître, O.P.; Cortés, J.-C. Second order linear differential equations with analytic uncertainties: Stochastic analysis via the computation of the probability density function. *J. Comput. Appl. Math.* **2020**, *374*, 112770.
8. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using Real NVP. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
9. Tang, K.; Wan, X.; Liao, Q. Deep density estimation via invertible block- triangular mapping. *Theor. Appl. Mech. Lett.* **2020**, *10*, 143–148.
10. Xiu, D. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*; Princeton University Press: Princeton, NJ, USA, 2010.
11. Botev, Z.; Ridder, A. Variance Reduction. *Wiley StatsRef: Statistics Reference Online* **2017**, 1–6, doi:10.1002/9781118445112.stat07975.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).