



Discrete approximations of continuous probability distributions obtained by minimizing Cramér-von Mises-type distances

Alessandro Barbiero¹ · Asmerilda Hitaj²

Received: 6 September 2021 / Revised: 2 August 2022 / Accepted: 7 September 2022
© The Author(s) 2022

Abstract

We consider the problem of approximating a continuous random variable, characterized by a cumulative distribution function (cdf) $F(x)$, by means of k points, $x_1 < x_2 < \dots < x_k$, with probabilities p_i , $i = 1, \dots, k$. For a given k , a criterion for determining the x_i and p_i of the approximating k -point discrete distribution can be the minimization of some distance to the original distribution. Here we consider the weighted Cramér-von Mises distance between the original cdf $F(x)$ and the step-wise cdf $\hat{F}(x)$ of the approximating discrete distribution, characterized by a non-negative weighting function $w(x)$. This problem has been already solved analytically when $w(x)$ corresponds to the probability density function of the continuous random variable, $w(x) = F'(x)$, and when $w(x)$ is a piece-wise constant function, through a numerical iterative procedure based on a homotopy continuation approach. In this paper, we propose and implement a solution to the problem for different choices of the weighting function $w(x)$, highlighting how the results are affected by $w(x)$ itself and by the number of approximating points k , in addition to $F(x)$; although an analytic solution is not usually available, yet the problem can be numerically solved through an iterative method, which alternately updates the two sub-sets of k unknowns, the x_i 's (or a transformation thereof) and the p_i 's, till convergence. The main apparent advantage of these discrete approximations is their universality, since they can be applied to most continuous distributions, whether they possess or not the first moments. In order to shed some light on the proposed approaches, applications to several well-known con-

✉ Alessandro Barbiero
alessandro.barbiero@unimi.it
Asmerilda Hitaj
asmerilda.hitaj@uninsubria.it

¹ Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano, via Conservatorio 7, 20122 Milan, Italy

² Department of Economics, Università degli Studi dell'Insubria, Via Monte Generoso 71, 21100 Varese, Italy

tinuous distributions (among them, the normal and the exponential) and to a practical problem where discretization is a useful tool are also illustrated.

Keywords Cumulative distribution function · Moment matching · Quantile function · Quantization · Statistical distance

Mathematics Subject Classification 62E15 · 62E17

1 Introduction

Using a discrete approximation of a continuous random variable (rv) is a procedure that is often adopted in many problems where uncertainty is present and needs to be taken into account. Substituting a continuous probability density function (pdf) with an approximating probability mass function (pmf), supported on a (possibly) finite number of points, can heavily reduce the computational burden required for determining a numerical solution for the problem at hand, and can produce an approximate solution whose degree of accuracy is still acceptable. This is particularly true when a problem involves several quantities that should be modelled as continuous rvs and the solution depends on some complex function thereof. The exact solution in this case could be obtained by applying some multivariate numerical integration technique whose computational cost may be severe and dramatically increasing with the number of rvs involved. Approximating each continuous rv through a properly chosen discrete rv allows the researcher to avoid numerical integration and resort to enumeration, which is much easier to manage (Luceno 1999).

An interesting application of discrete approximation of continuous distributions can be found in the field of insurance. If one is required to determine the distribution of the total claim amount $S = \sum_{i=1}^N X_i$ corresponding to a random number N of i.i.d. claims whose size X_i is assumed to follow some continuous probability distribution, rather than resorting to integral convolution to determine the exact distribution of S , one can approximate the distribution of the claim size X_i through discretization and then apply Panjer's formula (Panjer 1981), which is a recursive formula that exactly calculates the distribution of the total, holding when the claim size is arithmetic with span size $h > 0$ and the distribution of the number of claims N belongs to the $(a, b, 0)$ class.

In the field of quantitative finance, a similar application is the following. Let $\mathbf{L} = (L_1, \dots, L_d)$ denote a vector of possibly dependent rvs, each one representing a loss on a particular trading desk, portfolio or operating unit within a firm, over a fixed time period. Sometimes we need to aggregate these losses into a single rv, typically the sum $L^+ = \sum_{i=1}^d L_i$, on which one can calculate a measure of the aggregate risk, for example, the Value-at-Risk, which is nothing else than the quantile at a prespecified level $0 < \alpha < 1$ of the distribution of L^+ . However, determining this measure of risk, when the joint distribution of \mathbf{L} is fully specified, requires the computation of the distribution of L^+ , which is not straightforward to derive even if all the rvs are mutually independent. A possible answer is represented by the discretization of the L_i and the construction of a joint pmf approximating the joint distribution of \mathbf{L} , on which the evaluation of the VaR of the aggregating function is much more straightforward (see,

for example, Jamshidian and Zhu (1996), where the authors consider the multivariate Gaussian distribution).

In reliability engineering, the stress-strength model describes a system with random strength which is subject to a random stress during its functioning, so that the system works only when the strength is greater than the stress. The probability that a system correctly works is termed reliability (Johnson 1988). Evaluating the reliability of a system thus requires the knowledge of the probability distributions of both stress and strength; when these latter depend on several stochastic factors, the probability distributions (and then the reliability) are often not analytically tractable and hence some form of approximation is needed. Given a known functional relationship between stress (or strength) and its random subfactors and assuming the subfactors of the stress (or strength) are independent, one feasible approach of approximating the probability distribution of stress (or strength) is through discretization of the subfactors (see, e.g., English et al. 1996).

For an assigned continuous probability distribution and a fixed number of approximating points k , the matter is how to build an “optimal” discrete approximation. Several criteria have been used so far; here is a rough classification into four main categories:

- moment equalization or moment matching: the discrete approximation is the one preserving as many moments as possible of the original distribution. Moments’ matching is carried out through a procedure known as Gaussian quadrature (Golub and Welsch 1969); in its more authentic form, the support points of the discrete approximation and their probabilities are derived simultaneously. This is by far the oldest and most popular discretization technique; as stated by Miller and Rice (1983), “Few people would accept an approximation that did not have roughly the same mean, variance, and skew as the original distribution”. However, its application is limited by the finiteness and existence of a closed-form expression for the first integer moments of the continuous rv: many random distributions commonly used in quantitative finance, for example, do not possess even lower-order moments. Several variants of Gaussian quadrature have been proposed; for example, Tanaka and Toda (2013) suggested that the k support points of the approximating distribution have to be chosen “a priori” and the probabilities have to be derived by maximizing the relative entropy to an assigned “reference distribution”, under the constraint that it matches as many moments as possible. Convergence properties of this maximum entropy method and applications to stochastic processes were later presented in Tanaka and Toda (2015), Farmer and Toda (2017);
- preservation of the distribution function: the discrete approximation preserves, at each support point, the value of the cdf or, alternatively, of the survival function (Roy and Dasgupta 2001). Actually, this technique has been employed for constructing discrete counterparts of continuous distributions, by defining its pmf as $p(i) = F(i) - F(i - 1)$, with i integer; such construction automatically supplies a valid pmf, which preserves the cdf of the continuous distribution at any integer value i (Chakraborty 2015). Straightforward modifications have been proposed in order to handle finite supports consisting of possibly non-integer points;
- minimizing the mean squared error between the assigned continuous rv $X \sim F$ and its approximation \hat{X} , i.e., minimizing $\mathbb{E}_F(X - \hat{X})^2$, where the expected value

is computed with respect to the distribution of X ; this yields the so-called optimal quantization, which is a well-known technique in signal theory (Gray and Neuhoﬀ 1998). The minimization problem can be solved iteratively, by alternately solving two sets of conditions, one expressing each support point (also named “quantum”) as the center of mass of the intervals of a partition of the support of the continuous rv, and the other expressing the endpoint of each of these intervals as the midpoint of the segments between successive quanta (Lloyd 1982);

- minimizing a distance between the two distribution functions: the discrete approximation is obtained as the distribution minimizing some statistical distance between the two (continuous and step-wise) cdf. In Kennan (2006), an analytical solution is found for the optimal k -point discrete approximation when employing a particular class of distances.

In this work, we will concentrate on the last class of discrete approximations. More precisely, we propose the use of three distances: the Cramér-von Mises, the Anderson-Darling, and the Cramér distance between the cdf of the assigned continuous rv and the step-wise cdf of its k -point discrete approximation. They can be all regarded as weighted Cramér-von Mises distances with a proper choice of the weighting function. We will derive for these three cases the optimal solution (i.e., the k -point discrete approximation leading to the minimum distance) either analytically or computationally, providing in the latter case some details about the numerical procedure to be implemented. To the best of our knowledge, these distances have not been used regularly in the existing literature for the discrete approximation of a continuous rv. The aim of this paper is to shed some light on their use and explain their pros and cons.

The rest of the paper is structured as follows. In the next section, we will discuss distances between cdf in general and then present and solve the main research question, that is, finding an optimal k -point discrete approximation to a continuous random distribution through minimization of a statistical distance. The three statistical distances mentioned above will be considered and the corresponding solutions to the problem will be described and compared, also practically referring to some known parametric families of continuous distributions. Section 3 illustrates two applications of discretization to the (approximate) determination of the distribution or of some parameter of an assigned function of independent rvs, which is carried out by using each of the discrete approximations previously described and also other existing techniques. A software implementation in the R programming environment is presented in Sect. 4. Some comments and remarks are provided in the last section.

2 Optimal k -point approximation based on minimization of a distance between distribution functions

2.1 Statement of the problem

Let us consider two cdfs $F(x)$ and $G(x)$, and assume that $F(x)$ possesses a density, say $f(x)$, so that we can write $F(x) = \int_{-\infty}^x f(u)du$ for any real x . Several statistical distances between such two cdfs have been proposed. The most popular is the Kolmogorov-Smirnov (KS) distance, defined as $\sup_{x \in \mathbb{R}} |F(x) - G(x)|$, i.e., the maxi-

mum (or, better, supremum) absolute difference between the two cdfs. The KS distance is commonly employed when one has to test whether an i.i.d. sample (x_1, x_2, \dots, x_n) is consistent with some known cdf F ; in this case, the distance to be computed is between the assigned F and the empirical cdf $\tilde{F}(x)$, defined as $\tilde{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(x_i \leq x)$.

Another wide family of distances between cdfs is the following

$$d_2(F, G) = \int_{-\infty}^{+\infty} [F(x) - G(x)]^2 w(x) dx, \tag{1}$$

where $w(x)$ is some non-negative function on \mathbb{R} .

If we set $w(x) \equiv 1$, then we obtain the second-order squared Cramér distance (Cramér 1928), which is closely related to the so-called energy distance (Rizzo and Székely 2016). If we set $w(x) \equiv f(x)$, then we obtain the so-called Cramér-von Mises distance, which can be also conveniently written as $\int_0^1 [u - G(F^{-1}(u))]^2 du$, provided that $F(x)$ is invertible. If we set $w(x) = f(x)[F(x)(1 - F(x))]^{-1}$, then we obtain the Anderson-Darling distance; with respect to Cramér-von Mises, this distance puts more weight in the tails of the distribution, i.e., where $F(x)$ is close to either zero or one.

Hanebeck and Klumpp (2008), also referring to the less explored multivariate case, remark “how statistical distances are used for both analysis and synthesis purposes. Analysis is concerned with assessing whether a given sample stems from a given continuous distribution. Synthesis is concerned with both density estimation, i.e., calculating a suitable continuous approximation of a given sample, and density discretization, i.e., approximation of a given continuous random vector by a discrete one”. In this work, we focus on the latter aspect of research.

If one is interested in building an optimal k -point discrete approximation of a given continuous cdf F , i.e., a discrete probability distribution somehow resembling the original one, then he/she can find it as the discrete distribution minimizing, over all the k -point discrete distributions, the distance (1), computed between the cdf F and the cdf \hat{F} of the discrete approximation, for a given choice of the weighting function $w(x)$.

The problem can be more formally stated as follows. Suppose we want to approximate the continuous probability distribution F by a discrete probability distribution \hat{F} , consisting of $k > 1$ points $x_1 < x_2 < \dots < x_{k-1} < x_k$, with probabilities $p_i, i = 1, \dots, k$ (obviously, $p_i \geq 0$ for $i = 1, \dots, k$ and $\sum_{i=1}^k p_i = 1$). Let us gather the support points and the probabilities of the discrete distribution in a vector $\eta = (x_1, \dots, x_k, p_1, \dots, p_k)$. Then, the optimal discrete distribution can be defined as the one (univocally identified by $\hat{\eta}$) minimizing $d(F(x), \hat{F}(x; \eta))$: $\hat{\eta} = \arg \min d(F, \hat{F}; \eta)$. For the case $w(x) \equiv f(x)$, corresponding to the Cramér-von Mises distance, an analytical solution is available (Kennan 2006). When $w(x)$ is a piece-wise constant function, the problem was solved numerically by Schremp et al. (2006), through an iterative procedure based on a homotopy continuation approach. For any other possible choice of the weighting function $w(x)$, we can solve the minimization problem (at least) numerically.

We start our analysis from the Cramér-von Mises distance.

2.2 Cramér-von Mises

The Cramér-von Mises distance between continuous cdf is one of the distinguished measures of deviation between distributions (Cramér 1928; von Mises 1931). For a probabilistic interpretation, see Baringhaus and Henze (2017). It is obtained by setting $w(x) \equiv f(x)$ in (1), so that the distance between F and \hat{F} becomes

$$d_{CvM}(F, \hat{F}) = \int_{\mathbb{R}} |F(x) - \hat{F}(x)|^2 dF(x) = \int_0^1 |t - \hat{F}(F^{-1}(t))|^2 dt, \quad (2)$$

where F^{-1} denotes the mathematical inverse of the cdf F of X , which we assume to be strictly increasing over the support of X ; the distance in (2) is a particular case of

$$d_r(F, \hat{F}) = \int_{\mathbb{R}} |F(x) - \hat{F}(x)|^r dF(x) = \int_0^1 |t - \hat{F}(F^{-1}(t))|^r dt, \quad r > 0. \quad (3)$$

The best k -point discrete approximation (that is, the one yielding the minimum value of the distance d_r) has been proved (Kennan 2006) to have, for any $r > 0$, equally-weighted support points x_i (i.e., with probability $1/k$ each) such that $F(x_i) = \frac{2i-1}{2k}$ or, equivalently,

$$x_i = F^{-1}\left(\frac{2i-1}{2k}\right), \quad i = 1, \dots, k. \quad (4)$$

Here we report the proof, fixing and adding some points left out in Kennan (2006).

Let us introduce the quantities q_i and Q_i , defined in the following way: $q_i = F(x_i)$, $i = 1, \dots, k$, setting $q_0 = 0$ and $q_{k+1} = 1$; $Q_i = \hat{F}(x_i)$, $i = 1, \dots, k$, setting $Q_0 = 0$; it follows that $Q_k = 1$. Therefore, the q_i represent the values of the cdf of the continuous rv at the support points x_i of its discrete approximation; the Q_i represent the values of the cdf of the discrete approximating rv at its support points $x_i = F^{-1}(q_i)$, so that $p_i = Q_i - Q_{i-1}$, $i = 1, \dots, k$ (see Fig. 1).

The distance (3) can be thus rewritten as

$$d_r(F, \hat{F}) = \sum_{i=0}^k \int_{q_i}^{q_{i+1}} |t - Q_i|^r dt,$$

and the first-order condition on the Q_i gives

$$|q_i - Q_i|^r - |q_{i+1} - Q_i|^r = 0,$$

from which $Q_i = (q_i + q_{i+1})/2$; the first-order condition on the q_i provides

$$-|q_i - Q_i|^r + |q_i - Q_{i-1}|^r = 0,$$

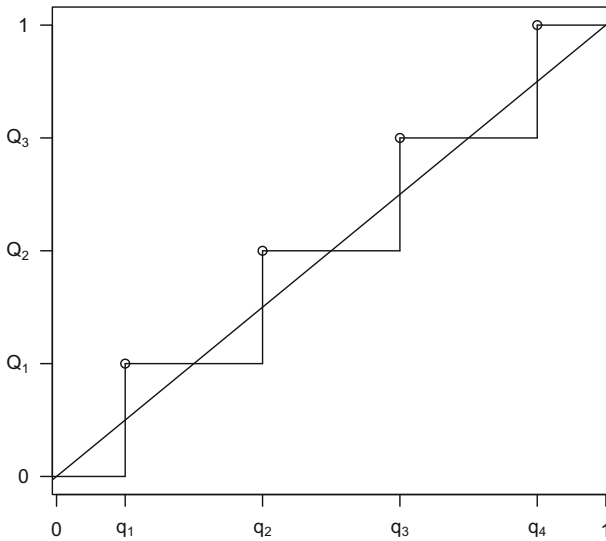


Fig. 1 Meaning of q_i and Q_i when constructing a k -point discrete approximation based on the minimization of the Cramér-von Mises distance (here $k = 4$). Since $q_i = F(x_i)$ and $Q_i = \hat{F}(x_i)$, it is possible to alternatively represent any k -point discrete approximation as a set of k points (q_i, Q_i) belonging to the unit square

from which $q_i = (Q_{i-1} + Q_i)/2$. Combining these two last results together, we obtain

$$Q_i = \frac{1}{2} \left(\frac{Q_{i-1} + Q_i}{2} + \frac{Q_i + Q_{i+1}}{2} \right)$$

that is, $Q_i - Q_{i-1} = Q_{i+1} - Q_i$, and being $Q_0 = 0$ and $Q_k = 1$, we derive $Q_i = i/k$ and $p_i = 1/k$, for $i = 1, \dots, k - 1$. By substituting this expression in that for the q_i , we obtain

$$q_i = \frac{1}{2} \left(\frac{i-1}{k} + \frac{i}{k} \right) = \frac{2i-1}{2k},$$

and then $x_i = F^{-1}((2i-1)/k)$ for $i = 1, \dots, k$. Note the particular form of the optimal solution: the specific continuous distribution enters the equation of the support points through the inverse of its cdf F , which is applied to the q_i , which are “distribution-free” quantities; the probabilities p_i are independent from F as well, being all constant.

Another interesting property of this solution can be pointed out. Letting \mathcal{F}_k be the set of discrete distributions with k support points, we have just proved that for each $r > 1$ the k -point discrete approximation $\hat{F} \in \mathcal{F}_k$ that satisfies

$$\inf_{G \in \mathcal{F}_k} d_r^{1/r}(F, G) = \inf_{G \in \mathcal{F}_k} \left(\int_{\mathbb{R}} |F(x) - G(x)|^r dF(x) \right)^{1/r}$$

$$\begin{aligned}
&= \inf_{G \in \mathcal{F}_k} \left(\int_0^1 |t - G(F^{-1}(t))|^r dt \right)^{1/r} \\
&= \left(\int_0^1 |t - \hat{F}(F^{-1}(t))|^r dt \right)^{1/r} = \left(\int |F(x) - \hat{F}(x)|^r dF(x) \right)^{1/r} \\
&= d_r^{1/r}(F, \hat{F})
\end{aligned}$$

is the discrete uniform distribution on the set of point $x_i = F^{-1}((2i - 1)/(2k))$, $i = 1, \dots, k$. So, being for each $G \in \mathcal{F}_k$

$$\begin{aligned}
\lim_{r \rightarrow \infty} \left(\int_0^1 |t - G(F^{-1}(t))|^r dt \right)^{1/r} &= \sup_{0 < t < 1} |t - G(F^{-1}(t))| = \sup_{x \in \mathbb{R}} |F(x) - G(x)| \\
&= d_{KS}(F, G),
\end{aligned}$$

recalling the relationship between the L^r norm and the supremum norm (see e.g. Stein and Shakarchi 2011), one obtains

$$d_{KS}(F, G) = \lim_{r \rightarrow \infty} d_r^{1/r}(F, G) \geq \lim_{r \rightarrow \infty} d_r^{1/r}(F, \hat{F}) = d_{KS}(F, \hat{F})$$

and deduces from this that

$$\inf_{G \in \mathcal{F}_k} d_{KS}(F, G) = d_{KS}(F, \hat{F}),$$

that is, for a fixed integer k , the best approximating distribution obtained by minimizing d_r is the same we would obtain by minimizing d_{KS} .

Another peculiarity of the optimal solution or, better, of the statistical distance used as a criterion for finding an optimal solution, is its all-encompassing applicability, since it does not require the existence and finiteness of any integer moment of the original continuous distribution: the distance (3), since it can be rewritten as an integral over the unit interval of a quantity which is finite, can be always computed and always possesses a global minimum, corresponding to the solution derived above. Therefore, it is possible to derive such a discrete approximation for Student's t , say, with any value of the degree-of-freedom parameter, and other heavy-tailed distributions, which would not be possible in general if using the moment-matching or quantization techniques (comprising the method by Drezner and Zerom (2016), which can be considered as a compromise between the two techniques): the finiteness of the first $2k - 1$ moments is required by the former, the finiteness of the first two moments by the latter. It can be also shown (Barbiero and Hitaj 2022) that $\lim_{k \rightarrow \infty} \hat{F}_k(x) = F(x) \forall x \in \mathbb{R}$, i.e., the approximating k -point discrete rv converges in distribution to the original rv.

The distance (3) can be further rewritten as the sum of $k + 1$ integrals,

$$d_r(F, \hat{F}) = \int_0^{q_1} t^r dt + \sum_{i=1}^{k-1} \int_{q_i}^{q_{i+1}} |t - Q_i|^r dt + \int_{q_k}^1 |1 - t|^r dt$$

$$= \frac{q_1^{r+1}}{r+1} + \sum_{i=1}^{k-1} \left(\frac{|q_{i+1} - Q_i|^{r+1}}{r+1} + \frac{|Q_i - q_i|^{r+1}}{r+1} \right) + \frac{|1 - q_k|^{r+1}}{r+1},$$

and its minimum value is thus equal to

$$\begin{aligned} \min d_r &= \left(\frac{1}{2k}\right)^{r+1} \frac{1}{r+1} + 2(k-1) \left(\frac{1}{2k}\right)^{r+1} \frac{1}{r+1} + \left(\frac{1}{2k}\right)^{r+1} \frac{1}{r+1} \\ &= \frac{(2k)^{-r}}{r+1}, \end{aligned} \tag{5}$$

since for the optimal solution $q_1 = q_{i+1} - Q_i = Q_i - q_i = 1 - q_k = 1/(2k)$, for each $i = 1, \dots, k - 1$. Furthermore, as one can expect, the minimum distance is a decreasing function of k for any fixed $r > 0$: by increasing the number of points, the optimal approximating discrete distribution gets closer (in terms of Cramér-von Mises distance) to the continuous one.

Table 1 displays, just for illustrative purposes, k -point approximations ($k = 5; 6; 7$) of a standard normal and an exponential distribution (with unit rate parameter). For each k and for both distributions, values of expectation and variance (simply indicated as μ and σ^2) are reported, in order to be easily compared with the analogous values of the continuous distribution.

For the standard normal distribution, variance and kurtosis of the k -point discrete approximation monotonically converge to the corresponding value of the original continuous distribution rather slowly: kurtosis, in particular, when $k = 100$, is 2.834. This result could have been expected: although the cdf of this discrete approximation converges to the cdf of the original continuous rv, this does not mean that for a finite k the two functions are so similar: for the normal case, recall that the pdf of the continuous rv is displayed through the classical bell-shaped curve, whereas the pmf of the discrete approximation has uniform probabilities, and this overall translates into a mismatch of (even) moments. Recall that moment equalization would produce discrete distributions with a very large range and very small probabilities assigned to the extreme support points (Barbiero and Hitaj 2022).

For the discrete approximation of the exponential distribution, expected value, variance, skewness, and kurtosis tend monotonically and asymptotically to the value of the parent distribution (1, 1, 2, and 9, respectively), but for finite k , it underestimates kurtosis to a large extent; when $k = 7$, the value of kurtosis for the discretized exponential distribution is 2.779 (its expected value is 0.951, its variance 0.686 and its skewness 0.963); when $k = 100$, it is 6.662 (its expected value is 0.997, its variance 0.960 and its skewness 1.759).

2.3 Anderson-Darling

What happens if we consider a weighted Cramér-von Mises distance, for example the Anderson-Darling distance, characterized by the weighting function $w(x) \equiv f(x)/[F(x)(1 - F(x))]$? We may expect that the best approximating distribution has now unequal probabilities or that the support points are no longer quantiles of

Table 1 k -point discrete approximations of two continuous random distributions based on the minimization of the Cramér-von Mises distance; μ and σ^2 indicate expectation and variance of the approximation

(a) Standard normal $X \sim N(0, 1)$					
$k = 5$		$k = 6$		$k = 7$	
x_i	p_i	x_i	p_i	x_i	p_i
-1.282	0.2	-1.383	0.167	-1.465	0.143
-0.524	0.2	-0.674	0.167	-0.792	0.143
0.000	0.2	-0.210	0.167	-0.366	0.143
0.524	0.2	0.210	0.167	0.000	0.143
1.282	0.2	0.674	0.167	0.366	0.143
		1.383	0.167	0.792	0.143
				1.465	0.143
$\mu = 0, \sigma^2 = 0.767$		$\mu = 0, \sigma^2 = 0.804$		$\mu = 0, \sigma^2 = 0.831$	
(b) Exponential with unit rate $X \sim Exp(\lambda = 1)(\mathbb{E}(X) = 1, Var(X) = 1)$					
0.105	0.2	0.087	0.167	0.074	0.143
0.357	0.2	0.288	0.167	0.241	0.143
0.693	0.2	0.539	0.167	0.442	0.143
1.204	0.2	0.875	0.167	0.693	0.143
2.303	0.2	1.386	0.167	1.030	0.143
		2.485	0.167	1.540	0.143
				2.639	0.143
$\mu = 0.932, \sigma^2 = 0.605$		$\mu = 0.943, \sigma^2 = 0.651$		$\mu = 0.951, \sigma^2 = 0.686$	

equally-spaced orders of the original distribution, as it occurs with Cramér-von Mises distance.

Minimizing the Anderson-Darling distance is equivalent to minimizing the following quantity, which consists of the sum of $k + 1$ contributions:

$$\begin{aligned}
 d_{AD}(F, \hat{F}) &= \int_{\mathbb{R}} |F(x) - \hat{F}(x)|^2 [F(x)(1 - F(x))]^{-1} dF(x) \\
 &= \int_0^1 \frac{|t - \hat{F}(F^{-1}(t))|^2}{t(1 - t)} dt = \sum_{i=0}^k \int_{q_i}^{q_{i+1}} \frac{(t - Q_i)^2}{t(1 - t)} dt,
 \end{aligned}$$

with respect to the q_i and Q_i . The first-order condition for Q_i is

$$\begin{aligned}
 \int_{q_i}^{q_{i+1}} \frac{d}{dQ_i} \frac{(t - Q_i)^2}{t(1 - t)} dt &= \int_{q_i}^{q_{i+1}} -\frac{2(t - Q_i)}{t(1 - t)} dt \\
 &= \int_{q_i}^{q_{i+1}} -\frac{2}{1 - t} + \frac{2Q_i}{t(1 - t)} dt
 \end{aligned}$$

$$\begin{aligned}
 &= 2 \log(1 - t) + 2Q_i \log \left(\frac{t}{1 - t} \right) \Big|_{q_i}^{q_{i+1}} \\
 &= 2 \log \left(\frac{1 - q_{i+1}}{1 - q_i} \right) \\
 &\quad + 2Q_i \log \left(\frac{q_{i+1}(1 - q_i)}{q_i(1 - q_{i+1})} \right) = 0,
 \end{aligned}$$

for $i = 1, \dots, k$, from which

$$Q_i = \log \left(\frac{1 - q_i}{1 - q_{i+1}} \right) / \log \left(\frac{q_{i+1}(1 - q_i)}{q_i(1 - q_{i+1})} \right). \tag{6}$$

The first-order condition for q_i implies

$$\frac{|q_i - Q_{i-1}|^2}{q_i(1 - q_i)} - \frac{|q_i - Q_i|^2}{q_i(1 - q_i)} = 0, \quad i = 1, \dots, k,$$

from which we obtain again

$$q_i = (Q_{i-1} + Q_i)/2, \quad i = 1, \dots, k. \tag{7}$$

The solution derived by Eqs.(6) and (7) cannot be expressed in an analytic closed form for each q_i and Q_i . However, a simple iterative algorithm can be implemented with the aim of recovering their values numerically; as initial guess values for q_i and Q_i , we adopt their optimal values found by minimizing the Cramér-von Mises distance. In a similar fashion as done in Pavlikov and Uryasev (2018), one can think of alternatively updating the values of the probabilities p_i and the values of the discrete points x_i (or, better, their probability transforms q_i) till convergence. Note that in Pavlikov and Uryasev (2018) all the p_i are initially set equal to $1/k$, as for the analytical solution based on the Cramér-von Mises distance. The algorithm works as follows:

1. Set $t = 0$; for $i = 1, \dots, k$, set $p_i^{(0)} = 1/k$, $Q_i^{(0)} = i/k$
2. Set $\epsilon^{(0)} = 1$ (or any large positive value) and $\epsilon_{\max} = 10^{-6}$ (or any arbitrarily small positive value, to be used for checking convergence of the solution)
3. While $\epsilon^{(t)} > \epsilon_{\max}$:
 - (a) Update the iteration index $t \leftarrow t + 1$
 - (b) Update the q_i according to (7): $q_i^{(t)} = \frac{Q_{i-1}^{(t-1)} + Q_i^{(t-1)}}{2}$
 - (c) Update the Q_i according to (6): $Q_i^{(t)} = \log \left(\frac{1 - q_i^{(t)}}{1 - q_{i+1}^{(t)}} \right) / \log \left(\frac{q_{i+1}^{(t)}(1 - q_i^{(t)})}{q_i^{(t)}(1 - q_{i+1}^{(t)})} \right)$
 - (d) Derive the updated probabilities for the discrete rv: $p_i^{(t)} = Q_i^{(t)} - Q_{i-1}^{(t)}$
 - (e) Calculate the maximum absolute deviation between two consecutive iterations in terms of p_i : $\epsilon^{(t)} = \max_{i=1}^k |p_i^{(t)} - p_i^{(t-1)}|$. Alternatively, other distances can be used to compare the probability vectors obtained in two consecutive iterations, as the Euclidean distance $\sqrt{\sum_{i=1}^k (p_i^{(t)} - p_i^{(t-1)})^2}$

Table 2 5-, 6- and 7-point discrete approximations based on the minimization of the Anderson-Darling distance

$k = 5$		$k = 6$		$k = 7$	
q_i	p_i	q_i	p_i	q_i	p_i
0.0799	0.1598	0.0624	0.1249	0.0505	0.1011
0.2702	0.2208	0.2131	0.1764	0.1734	0.1448
0.5000	0.2388	0.4007	0.1987	0.3294	0.1671
0.7298	0.2208	0.5993	0.1987	0.5	0.1741
0.9201	0.1598	0.7869	0.1764	0.6706	0.1671
		0.9376	0.1249	0.8266	0.1448
				0.9495	0.1011

The solution is expressed in terms of q_i (probability transformation of x_i , $q_i = F(x_i)$) and p_i

4. Return $q_i^{(t)}$, $Q_i^{(t)}$ and $p_i^{(t)}$

The number of iterations required clearly depends on the threshold ϵ_{\max} and on the number of points k : diminishing ϵ_{\max} or increasing k leads to a larger number of iterations; for plausible values of ϵ (say 10^{-8}) and k (say smaller than one hundred), the computation times are in the order of fractions of a second.

By numerical inspection, the Anderson-Darling weighting function leads to an inverted U-shaped trend for the p_i , which turn out to be symmetrical around the central value or values, according to whether k is odd or even, i.e., $p_j = p_{k-j+1}$, $j = 1, \dots, k$; see Table 2. As for the optimal q_i values, they present a form of symmetry around the central value(s), such that a continuous symmetrical distribution remains symmetrical after discretization (see Tables 2 and 3a, referring to the discretization of a standard normal rv).

If we consider the exponential distribution with unit rate parameter and we derive the k -point discrete approximation by minimizing the Anderson-Darling distance with $k = 7$ (see Table 3b), its expected value turns to be equal to 0.961, its variance 0.743, its skewness 1.147, and its kurtosis 3.424. Although not so close to the values of the underlying continuous distribution, the discrepancies are smaller if compared to those resulting from Cramér-von Mises approximation (see the previous subsection).

We note that using a distance $d(F, \hat{F}) = \int_{\mathbb{R}} (F(x) - \hat{F}(x))^2 w(F(x)) dF(x)$, with any other possible positive weighting function $w(x)$, even asymmetrical, supplies (although in general only numerically) values of q_i and Q_i that do not depend on the specific cdf $F(x)$; the original distance can in fact be rewritten as $\sum_{i=0}^k \int_{q_i}^{q_{i+1}} (t - Q_i)^2 w(t) dt$ and the first-order condition on the Q_i becomes $\int_{q_i}^{q_{i+1}} (t - Q_i) w(t) dt = 0$ and thus the Q_i that solve this condition can be expressed as a function, dependent on $w(x)$, of the q_i and q_{i+1} only. The simplest and most apparent case is provided by the “unweighted” Cramér-von Mises distance discussed in Sect. 2.2, where such values are available through simple analytical formulas involving just the index i and the number of support points k .

Table 3 k -point discrete approximation of two continuous random distributions based on the minimization of the Anderson-Darling distance

(a) Standard normal					
$k = 5$		$k = 6$		$k = 7$	
x_i	p_i	x_i	p_i	x_i	p_i
-1.406	0.160	-1.535	0.125	-1.640	0.101
-0.612	0.221	-0.796	0.176	-0.941	0.145
0.000	0.239	-0.252	0.199	-0.442	0.167
0.612	0.221	0.252	0.199	0.000	0.174
1.406	0.160	0.796	0.176	0.442	0.167
		1.535	0.125	0.941	0.145
				1.640	0.101
(b) Exponential with unit rate					
0.083	0.160	0.064	0.125	0.052	0.101
0.315	0.221	0.240	0.176	0.190	0.145
0.693	0.239	0.512	0.199	0.400	0.167
1.309	0.221	0.915	0.199	0.693	0.174
2.527	0.160	1.546	0.176	1.110	0.167
		2.774	0.125	1.752	0.145
				2.985	0.101

2.4 Cramér distance

We now consider the Cramér distance between the cdf of a continuous rv and that of its discrete approximation:

$$d_C(F, \hat{F}) = \int_{-\infty}^{+\infty} (F(x) - \hat{F}(x))^2 dx = \sum_{i=0}^k \int_{F^{-1}(q_i)}^{F^{-1}(q_{i+1})} (F(x) - Q_i)^2 dx,$$

which can be also rewritten as

$$d_C(F, \hat{F}) = \int_0^1 \frac{(t - \hat{F}(F^{-1}(t)))^2}{F'(F^{-1}(t))} dt = \sum_{i=0}^k \int_{q_i}^{q_{i+1}} \frac{(t - Q_i)^2}{F'(F^{-1}(t))} dt. \tag{8}$$

This last expression highlights how, differently from the two distances previously examined, the Cramér distance may not be always computed: for some cdf F , the first and/or the last term of the sum, in fact, may be not finite: although the two corresponding integrals are computed over a limited interval, the integrand function may be not finite (when t tends to 0 or t tends to 1, the denominator $F'(F^{-1}(t))$ may tend to zero; and in the former case, for example, it may tend to zero as fast as t^3 or faster and then the improper integral would not converge). We will see an example later.

Assuming that the Cramér distance can be computed, the first-order condition on the q_i leads again to

$$q_i = \frac{Q_i + Q_{i-1}}{2}, i = 1, \dots, k;$$

while the first-order condition on the Q_i leads to

$$\begin{aligned} \int_{q_i}^{q_{i+1}} \frac{d}{dQ_i} \frac{(t - Q_i)^2}{f(F^{-1}(t))} dt &= \int_{q_i}^{q_{i+1}} \frac{-2(t - Q_i)}{f(F^{-1}(t))} dt \\ &= -2 \int_{F^{-1}(q_i)}^{F^{-1}(q_{i+1})} (F(x) - Q_i) dx = 0 \end{aligned}$$

which becomes

$$\int_{F^{-1}(q_i)}^{F^{-1}(q_{i+1})} F(x) dx = Q_i [F^{-1}(q_{i+1}) - F^{-1}(q_i)], \tag{9}$$

for $i = 1, \dots, k$. This condition therefore depends on the specific distribution F of the continuous rv X .

We will now analytically develop condition (9) for several choices of F , by examining some well known parametric families. We will see that in general Eq. (9), combined with the condition on the q_i , does not lead to a closed-form solution of the discrete approximation, which must be solved numerically, for instance resorting to the iterative procedure sketched out in Sect. 2.3.

2.4.1 Normal

In case of a standard normal rv, with cdf $F(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$, for which the following equality holds (see, for example, Owen (1980), formula 1000):

$$\int \Phi(x) dx = x\Phi(x) + \phi(x) + \text{constant},$$

where $\phi(x) = \Phi'(x)$, the first-order condition on the Q_i becomes

$$\begin{aligned} Q_i &= \frac{\Phi^{-1}(q_{i+1}) \cdot q_{i+1} + \phi(\Phi^{-1}(q_{i+1})) - \Phi^{-1}(q_i) \cdot q_i - \phi(\Phi^{-1}(q_i))}{\Phi^{-1}(q_{i+1}) - \Phi^{-1}(q_i)}, \quad i \\ &= 1, \dots, k, \end{aligned} \tag{10}$$

which, along with the first-order condition on the q_i , allows us to determine the optimal solution numerically, following analogous steps to those sketched out in Sect. 2.3.

If instead of considering a standard normal rv, we focus on a generic normal rv with parameters μ and σ^2 , with cdf $F(x) = \Phi(\frac{x-\mu}{\sigma})$, the first-order condition on the Q_i

Table 4 5-, 6- and 7-point discrete approximation of a standard normal based on the minimization of the Cramér distance

$k = 5$		$k = 6$		$k = 7$	
x_i	p_i	x_i	p_i	x_i	p_i
-1.377	0.1685	-1.500	0.1337	-1.600	0.1095
-0.592	0.2167	-0.768	0.1751	-0.907	0.1454
0.000	0.2296	-0.242	0.1912	-0.424	0.1617
0.592	0.2167	0.242	0.1912	0.000	0.1666
1.377	0.1685	0.768	0.1751	0.424	0.1617
		1.500	0.1337	0.907	0.1454
				1.600	0.1095

would not change; both the left and right members of (9), in fact, are simply multiplied by σ .

Table 4 displays the values (x_i, p_i) of the best k -point approximation of a standard normal rv, for $k = 5; 6; 7$. We note that, as for the discrete approximation based on the minimization of the Anderson-Darling distance, the support points are symmetrical around zero and the probabilities satisfy $p_j = p_{k-j+1}, j = 1, \dots, k$.

2.4.2 Exponential

In case of an exponential rv, with cdf $F(x) = 1 - e^{-\lambda x}, x > 0$, and quantile function $F^{-1}(u) = -\frac{\log(1-u)}{\lambda}, 0 < u < 1$, we have that the first order condition on the Q_i can be written as

$$\int_{F^{-1}(q_i)}^{F^{-1}(q_{i+1})} (1 - e^{-\lambda x}) dx = Q_i \left[-\frac{\log(1 - q_{i+1})}{\lambda} + \frac{\log(1 - q_i)}{\lambda} \right] \implies$$

$$\left[x + \frac{e^{-\lambda x}}{\lambda} \right]_{-\frac{\log(1-q_i)}{\lambda}}^{-\frac{\log(1-q_{i+1})}{\lambda}} = \frac{Q_i}{\lambda} \log \frac{1 - q_i}{1 - q_{i+1}} \implies$$

$$\frac{1}{\lambda} [\log(1 - q_i) - \log(1 - q_{i+1}) + 1 - q_{i+1} - (1 - q_i)] = \frac{Q_i}{\lambda} \log \frac{1 - q_i}{1 - q_{i+1}},$$

from which

$$Q_i = 1 - \frac{q_{i+1} - q_i}{\log \frac{1 - q_i}{1 - q_{i+1}}}. \tag{11}$$

Table 5 displays the values (x_i, p_i) of the best k -point approximation of an exponential rv with unit parameter for some values of k . It is very important to notice that empirical inspection shows that this type of approximation is characterized by a decreasing pmf, which thus resembles the decreasing trend of the pdf; on the contrary, the discrete approximation derived from the minimization of the Anderson-Darling distance possesses an increasing-decreasing pmf; the discrete approximation based on the minimization of Cramér-von Mises distance possesses a constant pmf. If we

Table 5 5-, 6- and 7-point optimal discrete approximations of an exponential with unit rate parameter based on the minimization of Cramér distance

$k = 5$		$k = 6$		$k = 7$	
x_i	p_i	x_i	p_i	x_i	p_i
0.147	0.273	0.123	0.231	0.106	0.200
0.502	0.243	0.410	0.210	0.346	0.185
0.968	0.208	0.765	0.187	0.634	0.168
1.646	0.166	1.231	0.160	0.989	0.149
2.903	0.110	1.909	0.128	1.455	0.128
		3.166	0.084	2.133	0.102
				3.389	0.067

consider the exponential distribution with unit rate parameter and its optimal 7-point discrete approximation, the latter has an expected value equal to 0.972, a variance equal to 0.804, a skewness 1.313, and a kurtosis 4.089. Such values are closer to the analogous ones for the exponential distribution if compared to the discrete approximations obtained from the minimization of Cramér-von Mises and Anderson-Darling distances.

2.4.3 Cauchy

In case of a standard Cauchy rv, with cdf $F(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$, $x \in \mathbb{R}$, we have that the first-order condition on the Q_i provides

$$\int_{\tan[\pi(q_i-1/2)]}^{\tan[\pi(q_{i+1}-1/2)]} \left(\frac{1}{\pi} \arctan x + \frac{1}{2} \right) dx = Q_i \{ \tan[\pi(q_{i+1} - 1/2)] - \tan[\pi(q_i - 1/2)] \}$$

and then

$$\left[\frac{1}{\pi} \left(x \arctan x - \frac{1}{2} \log(1 + x^2) \right) + \frac{x}{2} \right]_{\tan[\pi(q_i-1/2)]}^{\tan[\pi(q_{i+1}-1/2)]} = Q_i \{ \tan[\pi(q_{i+1} - 1/2)] - \tan[\pi(q_i - 1/2)] \},$$

from which

$$Q_i = \frac{q_{i+1} \tan[\pi(q_{i+1} - 1/2)] - q_i \tan[\pi(q_i - 1/2)] - \frac{1}{2\pi} \log \frac{1 + \tan^2[\pi(q_{i+1}-1/2)]}{1 + \tan^2[\pi(q_i-1/2)]}}{\tan[\pi(q_{i+1} - 1/2)] - \tan[\pi(q_i - 1/2)]}.$$

We highlight that it is possible to find a k -point discrete approximation for the Cauchy distribution by minimizing the Cramér distance (as well as Cramér-von Mises or Anderson-Darling), although such a distribution does not possess any positive integer moment and then any discretization technique based on some form of moment-equalization – among others, moment matching, which requires equalization of the

first $2k - 1$ moments, but even the technique presented by Drezner and Zerom (2016) – would be not applicable at all.

2.4.4 Logistic

If we consider the standard logistic distribution with cdf $F(x) = (1 + e^{-x})^{-1}$, $x \in \mathbb{R}$, then its inverse cdf is $F^{-1}(u) = \log \frac{u}{1-u}$, $0 < u < 1$, and the first-order condition on the Q_i can be written as:

$$\int_{\log \frac{q_i}{1-q_i}}^{\log \frac{q_{i+1}}{1-q_{i+1}}} \frac{1}{1 + e^{-x}} dx = Q_i \left[\log \frac{q_{i+1}}{1 - q_{i+1}} - \log \frac{q_i}{1 - q_i} \right]$$

from which

$$\begin{aligned} \left[\log(1 + e^x) \right]_{\log \frac{q_i}{1-q_i}}^{\log \frac{q_{i+1}}{1-q_{i+1}}} &= \log \frac{1}{1 - q_{i+1}} - \log \frac{1}{1 - q_i} \\ &= Q_i \left[\log \frac{q_{i+1}}{1 - q_{i+1}} - \log \frac{q_i}{1 - q_i} \right] \end{aligned}$$

and then

$$Q_i = \frac{\log \frac{1-q_i}{1-q_{i+1}}}{\log \frac{q_{i+1}(1-q_i)}{(1-q_{i+1})q_i}} = \frac{\log \frac{1-q_i}{1-q_{i+1}}}{\log \frac{q_{i+1}}{q_i} + \log \frac{1-q_i}{1-q_{i+1}}}.$$

We note that the latter condition is the same as condition (6), obtained for the optimal k -point approximation based on the minimization of Anderson-Darling distance. This occurs since for the logistic distribution the equality $f(x)[F(x)(1 - F(x))]^{-1} = 1$ holds for any $x \in \mathbb{R}$, and hence the Cramér distance coincides with the Anderson-Darling distance.

Similarly to what happens with the normal distribution, if we consider a non-standard logistic distribution with location parameter μ and scale parameter σ , with cdf $F(x; \mu, \sigma) = \frac{1}{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}}$ and quantile function $F^{-1}(u; \mu, \sigma) = \mu + \sigma \log \frac{u}{1-u}$, it is straightforward to see that the first-order condition on Q_i would remain the same.

2.4.5 Lomax or Pareto

For the Lomax distribution with scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$, the expression of the cdf is $F(x) = 1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha}$, $x > 0$; the quantile function is $F^{-1}(u) = \lambda[(1 - u)^{-1/\alpha} - 1]$. The first-order condition on Q_i , if $\alpha \neq 1$, is

$$\begin{aligned} &\left[x - \frac{\lambda}{1 - \alpha} \left(1 + \frac{x}{\lambda}\right)^{1-\alpha} \right]_{\lambda[(1-q_i)^{-1/\alpha}-1]}^{\lambda[(1-q_{i+1})^{-1/\alpha}-1]} \\ &= Q_i \left\{ \lambda[(1 - q_{i+1})^{-1/\alpha} - 1] - \lambda[(1 - q_i)^{-1/\alpha} - 1] \right\}, \end{aligned}$$

from which

$$Q_i = 1 - \frac{1}{1 - \alpha} \frac{(1 - q_{i+1})^{(\alpha-1)/\alpha} - (1 - q_i)^{(\alpha-1)/\alpha}}{(1 - q_{i+1})^{-1/\alpha} - (1 - q_i)^{-1/\alpha}}. \tag{12}$$

In particular, if $\alpha = 2$, formula (12) reduces to the following expression

$$Q_i = 1 - \sqrt{(1 - q_i)(1 - q_{i+1})}.$$

If $\alpha = 1$, the first-order condition on the Q_i can be written as

$$\lambda Q_i [(1 - q_{i+1})^{-1} - (1 - q_i)^{-1}] = \left[x - \lambda \log \left(1 + \frac{x}{\lambda} \right) \right]_{\lambda[(1 - q_i)^{-1} - 1]}^{\lambda[(1 - q_{i+1})^{-1} - 1]},$$

from which

$$Q_i = 1 - \frac{\log \frac{1 - q_i}{1 - q_{i+1}}}{(1 - q_{i+1})^{-1} - (1 - q_i)^{-1}}$$

If $\alpha \leq 1/2$, it can be proved that the Cramér distance cannot be computed; in fact, for any feasible value of λ and k , since the integrand function of the last integral in (8), $\int_{q_k}^1 (1 - t)^2 / f(F^{-1}(t)) dt$, turns out to be proportional to $(1 - t)^{1-1/\alpha}$, then, being $\alpha \leq 1/2$, the (improper) integral is not finite for any feasible value of q_k . We highlight that the Lomax distribution does not possess any integer moment for $\alpha \leq 1$, so the proposed procedure is still able to produce a k -point discrete approximation even if the distribution does not possess a finite expectation, for $1/2 < \alpha \leq 1$.

We would obtain the same results if we considered the Pareto rv with cdf $F(x) = 1 - (\frac{\lambda}{x})^\alpha$, $x > \lambda$, by simply substituting x with $\lambda + x$.

In general, it is not possible to obtain a closed-form optimal solution, unless for small values of k . For example, if $\lambda = 1$ and $\alpha = 2$, for $k = 2$, the conditions on the q_i are $q_1 = Q_1/2$ and $q_2 = (1 + Q_1)/2$; the condition on Q_1 is $Q_1 = 1 - \sqrt{(1 - q_1)(1 - q_2)}$. Substituting the first two expressions for q_1 and q_2 into the last condition, we obtain, after a few simple passages, that $1 - Q_1 = \sqrt{\frac{1}{2} + \frac{Q_1^2}{4} - \frac{3Q_1}{4}}$ and then a second-order equation in Q_1 whose unique feasible solution is $Q_1 = \frac{2}{3}$. Consequently, $q_1 = \frac{1}{3}$ and $q_2 = \frac{5}{6}$; the best discrete approximation consists of the points $x_1 = \sqrt{\frac{3}{2}} - 1$ and $x_2 = \sqrt{6} - 1$ with probabilities $p_1 = 2/3$ and $p_2 = 1/3$.

2.4.6 Power function

The cdf of a power function rv is $F(x) = (x/b)^c$, $0 < x < b$, $b > 0$, $c > 0$. The corresponding quantile function is $F^{-1}(u) = bu^{1/c}$, $0 < u < 1$.

The first-order condition on Q_i turns into

$$\left[\frac{x^{c+1}}{b^c(c+1)} \right]_{bq_i^{1/c}}^{bq_{i+1}^{1/c}} = Q_i \left[bq_{i+1}^{1/c} - bq_i^{1/c} \right]$$

from which

$$Q_i = \frac{1}{c + 1} \cdot \frac{q_{i+1}^{(c+1)/c} - q_i^{(c+1)/c}}{q_{i+1}^{1/c} - q_i^{1/c}}.$$

If $c = 1$ (corresponding to the case of a uniform distribution) then $Q_i = (q_i + q_{i+1})/2$ and we obviously obtain the same solution as in Sect. 2.2. If $c = 2$, then we have

$$\begin{aligned} Q_i &= \frac{1}{3} \frac{q_{i+1}^{3/2} - q_i^{3/2}}{q_{i+1}^{1/2} - q_i^{1/2}} = \frac{1}{3} \frac{(q_{i+1}^{3/2} - q_i^{3/2})(q_{i+1}^{1/2} + q_i^{1/2})}{q_{i+1} - q_i} \\ &= \frac{1}{3} \frac{(q_{i+1} - q_i)(q_i + q_{i+1} + \sqrt{q_i q_{i+1}})}{q_{i+1} - q_i} = \frac{1}{3} (q_i + q_{i+1} + \sqrt{q_i q_{i+1}}); \end{aligned}$$

thus the cumulative probability Q_i of the optimal solution corresponds to the arithmetic mean of the two consecutive probabilities q_i and q_{i+1} and their geometric mean.

It can be numerically shown that for values of c larger than 1 (when the pdf is increasing), the k -point discrete approximation has an increasing pmf; on the contrary, for $0 < c < 1$ (when the pdf is decreasing), the k -point discrete approximation has a decreasing pmf.

In general, it is not possible to derive the optimal solution in a closed-form, unless for small values of k . For example, if $b = 1$ and $c = 2$, for $k = 2$, the conditions on the q_i are $q_1 = Q_1/2$ and $q_2 = (1 + Q_1)/2$; the condition on Q_1 is $Q_1 = \frac{1}{3}(q_1 + q_2 + \sqrt{q_1 q_2})$. Substituting the first two expressions for q_1 and q_2 into the last equation, we obtain, after a few simple passages, that $4Q_1 - 1 = \sqrt{Q_1(1 + Q_1)}$ and then obtain a second-order equation in Q_1 whose unique feasible solution is $Q_1 = \frac{9 + \sqrt{21}}{30}$. Consequently, $q_1 = \frac{9 + \sqrt{21}}{60}$ and $q_2 = \frac{39 + \sqrt{21}}{60}$; the best discrete approximation consists of the points $x_1 = \sqrt{\frac{9 + \sqrt{21}}{60}}$ and $x_2 = \sqrt{\frac{39 + \sqrt{21}}{60}}$ with probabilities $p_1 = \frac{9 + \sqrt{21}}{30}$ and $p_2 = \frac{21 - \sqrt{21}}{30}$.

2.5 Remarks

2.5.1 The case $k = 1$

We have always implicitly assumed so far that the number of points k by which one approximates the assigned continuous random distribution is greater than 1. It is immediate to realize that if one wants to approximate the distribution of the rv X through one value only, then, except for the cases where the selected statistical distance cannot be computed (see Sect. 2.4.5) the “optimal” approximating value x_1 is the median of F , $F^{-1}(0.5)$, since the only effective condition to be satisfied (for all the three distances examined) would be $q_1 = (Q_0 + Q_1)/2 = 1/2$, being $Q_0 = 0$ and $Q_1 = 1$. We note that optimal quantization (Lloyd 1982) would return the expected value $\mathbb{E}_F(X)$, as far as it exists and it is finite, as the optimal approximating value of F .

2.5.2 Location-scale transformation

Let us consider a continuous rv $X \sim F_X$ and its optimal k -point approximation derived from the minimization of the Cramér-von Mises or Anderson-Darling distance, $x_i = F_X^{-1}(q_i^*)$, $p_i^* = Q_i^* - Q_{i-1}^*$. We know that the optimal q_i^* and Q_i^* are only functions of k and of the selected distance, but do not depend on the specific F_X . Then consider the location-scale transformation $Y = a + bX$, $a \in \mathbb{R}$, $b > 0$. Since for such a transformation we have $F_Y(y) = F_X(\frac{y-a}{b})$ for any real y and $F_Y^{-1}(u) = a + bF_X^{-1}(u)$, $0 < u < 1$, the optimal k -point approximation of Y (derived by using the same distance as for X) is represented by $y_i = F_Y^{-1}(q_i^*) = a + bF_X^{-1}(q_i^*) = a + bx_i$ and the same p_i^* as before, which means that the location-scale transformation applies also to the discretized rv.

This property does not hold in general for the discrete approximation based on the minimization of the Cramér distance: the optimal values q_i^* and Q_i^* now depend on the specific form of the cdf, and F_Y may not belong to the same family as F_X , unless the cdf F_X is a location-scale family of distributions itself: in this case, the property still holds, as the condition (9) does not change if we consider a location-scale transformation of the cdf (recall the examples with the normal and logistic distributions in Sect. 2.4).

3 Example of application

Often a researcher in the statistical field is required to determine the distribution (or parameters) of some complex function of several (independent and continuous) rvs $T = t(X_1, X_2, \dots, X_d)$. Multidimensional integration techniques should be employed, but often, due to the complexity of t and to the high dimensionality d , they are either cumbersome to apply or not applicable at all. One can then resort to Monte Carlo simulation, i.e., simulating (independently) a huge number N of pseudo-random values from X_1, X_2, \dots, X_d and then calculating the corresponding N values of the transformation t , which can be regarded as a random sample drawn from T . An alternative is to find an approximate solution via approximation-by-discretization and enumeration. This consists of substituting each X_i with a discrete approximation \hat{X}_i and then determine the corresponding pmf of $\hat{T} = t(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_d)$ by “enumeration”, based on the joint pmf of $(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_d)$ over the Cartesian product of the single supports: $\mathcal{S}(\hat{X}_1) \times \dots \times \mathcal{S}(\hat{X}_d)$.

In order to illustrate and compare the discretization techniques proposed in Sect. 2, in the following subsection we will consider a case where it is possible to derive the exact distribution of T analytically, which allows us to evaluate the statistical performance (in terms of degree of approximation, to be measured through some index) of each technique; in Sect. 3.2, we will analyze a more complicated case where the exact distribution of T is not analytically computable, but a parameter of interest can be recovered numerically.

Table 6 KS distance between the true and the approximated distribution function of the sum of three independent exponential rvs

Method, k	5	6	7	8	9	10	11
Cramér-von Mises	0.0828	0.0682	0.0544	0.0487	0.0433	0.0392	0.0349
Anderson-Darling	0.0785	0.0588	0.0432	0.0350	0.0297	0.0255	0.0207
Cramér	0.0551	0.0401	0.0302	0.0241	0.0201	0.0162	0.0145

Approximation is obtained by applying one of the univariate discrete approximations of Sect. 2 to each rv independently and then reconstructing the distribution of the sum by “enumeration”

3.1 Sum of exponential random variables

Let $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$, $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$, ..., $X_d \sim \text{Gamma}(\alpha_d, \lambda)$, independent to each other; it is well-known that the sum $S = \sum_{i=1}^d X_i$ is $\text{Gamma}(\sum_{i=1}^d \alpha_i, \lambda)$. However, let us approximate each X_i through one of the discrete approximations we illustrated, and then reconstruct the approximated distribution of the sum, \hat{F}_S . We can then evaluate the degree of approximation by using a measure of discrepancy between F_S and \hat{F}_S ; we can use the KS distance, rather than one of the statistical distances employed for the univariate approximation. Here we consider the sum of $d = 3$ exponential rvs; $\lambda_i = 1/2$, $\alpha_i = 1$, $i = 1, 2, 3$, for which we know that $S \sim \text{Gamma}(3, 1/2)$. Table (6), for different values of k (from 5 to 11), reports the KS distance between the exact and the approximated cdf of S , obtained according to the three univariate discrete approximations for the X_i . It is easy to see how the KS distance decreases with k for all the methods: by increasing the number of approximating points for the X_i , it is legitimate to expect that the discrepancy between the true and the approximate cdf decreases, independently from the measure used. Moreover, we notice that the discrete approximation based on Cramér distance overperforms the other two approximations, and its relative level of accuracy quickly improves with k : for $k = 11$, the KS distance provided by the approximation based on the minimization of Cramér distance is much less than one half of the KS distance provided by the approximation based on the minimization of the Cramér-von Mises distance. As we already remarked, the discrete approximation of the exponential distribution based on the minimization of the Cramér distance seems to resemble the shape of the continuous pdf better than the discrete approximations based on the Cramér-von Mises and Anderson-Darling distances, and this also positively reflects when calculating the approximate distribution of the sum of three i.i.d. exponential rvs.

The graph in Fig. 2 displays the curve of the true cdf of the sum, and the three stepwise cdf of the three discrete approximations. It is quite apparent that the latter struggle in resembling the continuous curve in its right tail: it is visible to an unaided eye, in fact, for relatively high values of x (say, between 10 and 15) the three approximations show their maximum absolute error (which corresponds to the value of KS distance).

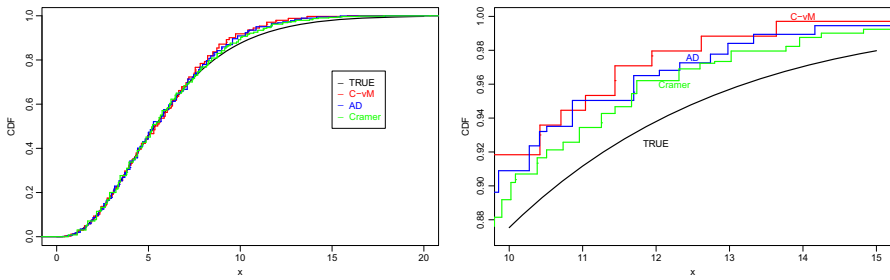


Fig. 2 Comparison between the true distribution function of the sum of exponential rvs and its three approximations (based on 7-point univariate discrete approximations); in the right panel, we zoom in on the right tail of the distribution

3.2 Reliability parameter for a hollow rectangular tube

In Sect. 1, we mentioned the problem of recovering the so-called reliability parameter $R = P(X > Y)$ for a stress-strength model, where X and Y are the strength and stress rvs, respectively, possibly depending on several stochastic sub-factors. Let us consider this example. The functional form of shear stress of a hollow rectangular tube is $Y = \frac{M}{2t(W-t)(H-t)}$, where M is the applied torque, t is the wall thickness, W is the width, and H the height of the tube. Let X , M , t , W , and H be mutually independent rvs. Consider the following parametric set-up: $M \sim \mathcal{N}(\mu_M = 1500, \sigma_M = 150)$, $t \sim \mathcal{N}(\mu_t = 0.2, \sigma_t = 0.005)$, $W \sim \mathcal{N}(\mu_W = 2, \sigma_W = 0.02)$, $H \sim \mathcal{N}(\mu_H = 3, \sigma_H = 0.03)$ (with \mathcal{N} denoting the normal distribution) which accord with that in Roy and Dasgupta (2001). Let the standard deviation of the normal strength X be 60, and assume that the mean of X can vary from 520 to 970 units by steps of 10, thus generating an array of $n = 46$ scenarios.

Although the exact evaluation of R is unfeasible here, as suggested by a referee, its value can be obtained numerically as follows. Denoting by $\phi_{\mu,\sigma}$ and $\Phi_{\mu,\sigma}$ the pdf and cdf, respectively, of a normal rv with mean μ and standard deviation σ , putting

$$f(t, m, w, h) = \left[1 - \Phi_{\mu_X, \sigma_X} \left(\frac{m}{2t(w-t)(h-t)} \right) \right] \cdot \phi_{\mu_t, \sigma_t}(t) \phi_{\mu_M, \sigma_M}(m) \phi_{\mu_W, \sigma_W}(w) \phi_{\mu_H, \sigma_H}(h)$$

for $(t, m, w, h) \in \mathbb{R}^4$, the reliability parameter R can be expressed as

$$R = \int \int \int \int_{\mathbb{R}^4} f(t, m, w, h) dt dm dw dh,$$

and can be approximated by

$$R_Q = \int \int \int \int_Q f(t, m, w, h) dt dm dw dh,$$

where Q is the hyper-rectangle defined as $Q = I_t \times I_M \times I_W \times I_H$, with $I_t = [\mu_t - \gamma_t \sigma_t, \mu_t + \gamma_t \sigma_t]$, $I_M = [\mu_M - \gamma_M \sigma_M, \mu_M + \gamma_M \sigma_M]$, $I_W = [\mu_W - \gamma_W \sigma_W, \mu_W + \gamma_W \sigma_W]$, $I_H = [\mu_H - \gamma_H \sigma_H, \mu_H + \gamma_H \sigma_H]$, and $\gamma_t, \gamma_M, \gamma_W, \gamma_H$ are positive scale factors to be chosen suitably large. The computation of R_Q is easily done by using the function `cuhre` that is included in the package `cubature` (Narasimhan et al. 2022) of the statistical software environment R (R Core Team 2022).

Alternatively, one can resort either to Monte Carlo simulation or to the approximation-by-discretization approach, by using one of the methods described in the previous sections.

We will regard the value of reliability recovered by numerical evaluation through the `cuhre` function (with all the γ scaling factors set equal to 9) as the actual (true) value and thus indicate it simply with R ; the value obtained through Monte Carlo simulation is denoted by \hat{R}^{MC} ; the values obtained by the approximation-by-discretization approaches are denoted by \hat{R} , with a superscript identifying the specific discretization: GQ (Gaussian Quadrature), DZ (Drezner and Zerom 2016), CvM (Cramér-von Mises), C (Cramér), AD (Anderson-Darling). In order to measure the degree of accuracy of each technique, one can consider the following synthetic measures, which are: Mean Deviation (MD), Mean Absolute Deviation (MAD), Root Mean Squared Deviation (RMSD), defined as follows,

$$\begin{aligned}
 MD &= \frac{1}{n} \sum_{i=1}^n (\hat{R}_i - R_i) \\
 MAD &= \frac{1}{n} \sum_{i=1}^n |\hat{R}_i - R_i| \\
 RMSD &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{R}_i - R_i)^2}
 \end{aligned}$$

where the subscript i refers to the i -th scenario.

We considered a number $N = 10$ millions of pseudo-random simulations for the Monte Carlo approach, and a number $k = 5$ of approximating points for the discretization approaches.

The results, reported in Table 7, show that Monte Carlo simulation provides overall the best results in terms of MAD and RMSD; the proposed discretization methods, based on the minimization of the Anderson-Darling and Cramér distances, perform better if compared to the Gaussian quadrature method in terms of MD, MAD, and RMDE; they perform worse than the method proposed in Drezner and Zerom (2016), which however requires a much considerable computational effort required by its inner numerical minimization routine and has a narrower applicability, as already remarked in Sect. 2.2. We remark that in this example we focused just on discretization techniques which are somehow comparable: according to their specific criterion, they all compute both the k support points and the corresponding k probabilities simultaneously. Other techniques, mentioned in the Introduction, such as the maximum entropy method, first assign the support points a priori and then calculate the probabilities only.

Table 7 Comparative closeness study for a hollow rectangular tube

Scenario#	μ_X	R	R_{MC}	\hat{R}^{GQ}	\hat{R}^{DZ}	\hat{R}^{CvM}	\hat{R}^{AD}	\hat{R}^C
1	520	0.0102	0.0103	0.0065	0.0098	0.0000	0.0001	0.0000
2	530	0.0134	0.0135	0.0086	0.0111	0.0000	0.0010	0.0005
3	540	0.0173	0.0175	0.0141	0.0132	0.0003	0.0039	0.0025
4	550	0.0223	0.0225	0.0210	0.0180	0.0026	0.0094	0.0077
5	560	0.0284	0.0287	0.0311	0.0276	0.0090	0.0160	0.0152
6	570	0.0358	0.0361	0.0438	0.0378	0.0179	0.0218	0.0214
7	580	0.0448	0.0451	0.0543	0.0489	0.0278	0.0265	0.0274
8	590	0.0554	0.0558	0.0613	0.0594	0.0365	0.0343	0.0349
9	600	0.0680	0.0687	0.0659	0.0693	0.0445	0.0477	0.0462
10	610	0.0827	0.0831	0.0698	0.0835	0.0570	0.0636	0.0626
11	620	0.0996	0.1003	0.0782	0.1028	0.0752	0.0815	0.0812
12	630	0.1190	0.1197	0.0955	0.1236	0.0976	0.1009	0.1006
13	640	0.1410	0.1416	0.1247	0.1484	0.1210	0.1237	0.1219
14	650	0.1654	0.1660	0.1659	0.1754	0.1466	0.1480	0.1476
15	660	0.1921	0.1929	0.2097	0.1967	0.1744	0.1784	0.1759
16	670	0.2218	0.2222	0.2516	0.2275	0.2070	0.2079	0.2081
17	680	0.2538	0.2544	0.2811	0.2591	0.2406	0.2416	0.2424
18	690	0.2880	0.2886	0.2989	0.2945	0.2758	0.2783	0.2782
19	700	0.3242	0.3244	0.3108	0.3289	0.3136	0.3149	0.3155
20	710	0.3619	0.3623	0.3241	0.3667	0.3514	0.3535	0.3533
21	720	0.4012	0.4014	0.3512	0.4027	0.3939	0.3953	0.3939
22	730	0.4417	0.4418	0.3997	0.4390	0.4387	0.4401	0.4393
23	740	0.4826	0.4827	0.4628	0.4835	0.4842	0.4840	0.4836
24	750	0.5235	0.5232	0.5389	0.5273	0.5254	0.5260	0.5253
25	760	0.5642	0.5636	0.6031	0.5644	0.5680	0.5656	0.5674
26	770	0.6043	0.6037	0.6492	0.5990	0.6099	0.6083	0.6082
27	780	0.6433	0.6427	0.6759	0.6369	0.6525	0.6494	0.6513
28	790	0.6808	0.6799	0.6877	0.6734	0.6909	0.6911	0.6913
29	800	0.7166	0.7159	0.6992	0.7118	0.7299	0.7285	0.7268
30	810	0.7503	0.7495	0.7181	0.7469	0.7648	0.7634	0.7620

4 Software implementation

Code in the R programming environment has been developed, which implements the different routines used for finding the optimal discrete approximations. For the Cramér distance, in particular, several parametric distributions are considered. Some functions are also available for plotting discretized distributions, which makes graphical comparison to the original continuous distribution more effective.

For the Cramér-von Mises and Anderson-Darling distances (and potentially, for any other distribution-free distance), the function `DISCR` has to be used, whose arguments

Table 7 continued

Scenario#	μ_X	R	R_{MC}	\hat{R}^{GQ}	\hat{R}^{DZ}	\hat{R}^{CvM}	\hat{R}^{AD}	\hat{R}^C
31	820	0.7818	0.7807	0.7489	0.7772	0.7971	0.7970	0.7962
32	830	0.8107	0.8096	0.7903	0.8013	0.8278	0.8253	0.8277
33	840	0.8372	0.8359	0.8343	0.8287	0.8560	0.8547	0.8540
34	850	0.8610	0.8600	0.8763	0.8519	0.8819	0.8796	0.8804
35	860	0.8824	0.8814	0.9042	0.8774	0.9046	0.9004	0.9021
36	870	0.9014	0.9005	0.9172	0.8962	0.9245	0.9212	0.9198
37	880	0.9180	0.9171	0.9234	0.9150	0.9414	0.9379	0.9378
38	890	0.9324	0.9316	0.9320	0.9286	0.9539	0.9514	0.9527
39	900	0.9447	0.9440	0.9377	0.9405	0.9648	0.9639	0.9639
40	910	0.9552	0.9544	0.9451	0.9513	0.9744	0.9726	0.9722
41	920	0.9641	0.9634	0.9565	0.9634	0.9827	0.9793	0.9804
42	930	0.9715	0.9708	0.9693	0.9714	0.9894	0.9851	0.9866
43	940	0.9776	0.9768	0.9796	0.9792	0.9955	0.9906	0.9914
44	950	0.9825	0.9819	0.9811	0.9843	0.9984	0.9946	0.9955
45	960	0.9865	0.9859	0.9885	0.9878	0.9997	0.9978	0.9984
46	970	0.9897	0.9892	0.9916	0.9905	1.0000	0.9990	0.9995
MD			-0.0002	-0.0016	-0.0004	0.0000	0.0001	0.0000
MAD			0.0006	0.0153	0.0039	0.0155	0.0131	0.0136
RMSD			0.0007	0.0203	0.0047	0.0166	0.0142	0.0146

R value of the reliability parameter obtained numerically through the cubature package, \hat{R} estimate of the reliability parameter obtained from Monte Carlo simulation (MC); or approximation-by-discretization approach: Gaussian Quadrature (GQ); Drezner and Zerom, 2016 (DZ); minimization of the Cramér-von Mises distance (CvM); minimization of the Anderson-Darling distance (AD); minimization of the Cramér distance (C). Note that the values of the mean of X , μ_X , have been actually chosen in order to induce a range of values for R roughly spanning the entire interval between 0.01 to 0.99

are the number of points k and the type of distance (CvM and AD). It returns a list containing the vectors of p_i and q_i (along with the vector of Q_i), from which one can extrapolate the support values x_i through the quantile transformation of the assigned distribution function F .

For the Cramér distance, the main function is called `DiscrF`; its arguments are the number of points k by which we want to approximate the continuous random distribution; the type of continuous distribution (a string identifying it), along with the value of its parameters (a vector, `par`). At the moment, a few families of continuous distributions can be selected, namely, those discussed in Sect. 2.4, for which the first-order condition on the Q_i is available in a closed-form: normal (`norm`), exponential (`exp`), Cauchy (`cauchy`), logistic (`logis`), Lomax (`lomax`), power function (`power`). The output of the function is a list, containing the vector of probabilities p_i and the vector of support values x_i of the discrete approximation, along with the vectors of q_i and Q_i .

The companion function `moments` receives as input a vector of support points and a vector of corresponding probabilities possibly obtained as a result from `DiscrF` or

`Discr`, and computes the expectation, variance, skewness and kurtosis of the related discrete rv. This is useful if one wants to keep under control the effects of discretization over the first (normalized) moments, since we know that the techniques we introduced do not offer any guarantee, in general, that the moments of the continuous rv are preserved.

The graphical function `plotdist` receives as its first argument the result of `Discr` or `DiscrF` and plots the corresponding discrete approximation (its pmf or its cdf, to be selected through the argument `plot`, to be set equal to "pmf" or "cdf"). This can be plotted over the unit square (by setting the argument `xaxis` equal to "q": on the x axis the q_i , on the y axis the p_i or the Q_i , see Fig. 1); or on the usual $\mathbb{R} \times [0, 1]$ space (by setting the argument `xaxis` equal to "x": on the x axis the x_i , which should be supplied if the first argument comes from `Discr`; on the y axis the p_i or the Q_i).

In the supplementary material, available at <https://tinyurl.com/STPA-D-21-00382>, the relevant R code along with some examples is supplied.

5 Conclusion

In this work, we discussed a class of discretization techniques that calculate a k -point discrete approximation of a continuous rv by minimizing a distance between the two cdf. For one distance in this class, the optimal discrete approximation (i.e., both the points x_i and their probabilities p_i , $i = 1, \dots, k$) turns out to have an analytical expression for any k . For other distances, a closed-form solution is not available, but the x_i and the p_i can be iteratively computed by alternately solving two sets of equations till convergence, in a similar fashion to optimal quantization. It may happen that the solution is "distribution-free", meaning that the probability transforms $F(x_i)$ and p_i do not depend on the particular distribution function F selected; or, on the contrary, that the solution directly depends on F : in the latter case, we derived the sets of equations to be satisfied by the solution for a wide array of continuous random distributions.

We underline that this class of discretization techniques represents a valid alternative to the other extant procedures, among them, the consolidated moment-matching technique, whose applicability is however hindered by the often unattainable hypothesis of finiteness of the first integer moments. This class is also competitive if compared to optimal quantization and modifications thereof, since for the latter the algorithm leading to the best discrete approximation is very similar, being based on the minimization of the mean squared error between the two rvs. What we cannot expect from the suggested class of discrete approximations is the preservation of the first (and second) moment; this can be seen as a shortcoming, but only if one is used to deal with or synthesize random distributions in terms of moments or if the transformation of rvs one needs to approximate is a smooth function. In the financial or insurance fields, one is more familiar with quantiles (the most popular risk measure for market risk is the Value at Risk (VaR), which is nothing else than a quantile of a loss distribution over a fixed time horizon; see, e.g., McNeil, Frey and Embrechts 2005) and then adopting a discrete approximation which minimizes a distance between cdfs may be intuitively

more appropriate and convenient. Moreover, many continuous distributions do not even possess the second or the first moment (just think of the Cauchy distribution) and then moment-matching and quantization techniques would fail in providing a discrete approximation, which is instead guaranteed by the proposed class (with rare exceptions arising if the Cramér distance is considered). The practical problem presented in the third section, concerning a complex function of several random variables, illustrates how the proposed procedures can overperform the standard moment matching approach.

Future research will examine other statistical distances between cdf, in particular asymmetrical ones, and derive the corresponding optimal k -point discrete approximation from their minimization. We will also examine possible extensions of this class of discretization techniques to the bivariate and more generally d -variate case. Although it can be naïvely judged to be straightforward, high dimensionality leads to non-negligible theoretical and computational issues: apart from the choice of the statistical distance to be minimized (an analogous quadratic distance between joint cdfs or the energy distance, see Rizzo and Székely 2016, can be considered) the most challenging matter is represented by the selection of all feasible d -variate supports for the discrete approximation (they are not necessarily a d -variate Cartesian product of univariate supports, and this complicates the evaluation of the statistical distance) and by the possible statistical dependence between the univariate components, which unavoidably calls for the use of copulas. Similar kinds of problems arise in the computation of principal points for multivariate random distributions (Flury 1990).

Another aspect that can be investigated is related to the construction of discrete approximations with “a priori” assigned support points x_i . This can be useful when one wants to construct a discrete counterpart or analog to a continuous probability distribution, and not just a discrete approximation. Typically, the discrete counterpart to a continuous distribution supported on \mathbb{R} (or \mathbb{R}^+) is constructed as the discrete distribution supported on \mathbb{Z} (or \mathbb{N}) preserving the expression of the continuous cdf at the integer support points (Chakraborty 2015). Alternatively, a discrete counterpart can be constructed by considering the same integer support as above and minimizing one of the statistical distances examined in this work with respect to the probabilities p_i , which now constitute a countable set. By considering any parametric random distribution with infinite support, we can thus generate several new count distributions that can be actually regarded as its discrete counterparts and can be employed for modeling count data.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00362-022-01356-2>.

Acknowledgements We wish to thank the Editor and two anonymous reviewers for their careful reading of our manuscript and their insightful comments and suggestions.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence,

and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barbiero A, Hitaj A (2022) Approximation of continuous rvs for the evaluation of the reliability parameter of complex stress–strength models. *Ann Oper Res* 315(2):1–26
- Baringhaus L, Henze N (2017) Cramér–von Mises distance: probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *J Nonparametr Stat* 29(2):167–188
- Chakraborty S (2015) Generating discrete analogues of continuous probability distributions: a survey of methods and constructions. *J Stat Distrib Appl* 2(1):1–30
- Cramér H (1928) On the composition of elementary errors: first paper: mathematical deductions. *Scand Actuar J* 1928(1):13–74
- Drezner Z, Zerom D (2016) A simple and effective discretization of a continuous random variable. *Commun Stat Simul Comput* 45(10):3798–3810
- English JR, Sargent T, Landers TL (1996) A discretizing approach for stress/strength analysis. *IEEE Trans Reliability* 45(1):84–89
- Farmer LE, Toda AA (2017) Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments. *Quant Econom* 8(2):651–683
- Flury BA (1990) Principal points. *Biometrika* 77(1):33–41
- Golub GH, Welsch JH (1969) Calculation of Gauss quadrature rules. *Math Comput* 23(106):221–230
- Gray RM, Neuhoff DL (1998) Quantization. *IEEE Trans Inform Theory* 44(6):2325–2383
- Hanebeck UD, Klumpp V (2008) Localized cumulative distributions and a multivariate generalization of the Cramér–von Mises distance. In: 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp 33–39
- Jamshidian F, Zhu Y (1996) Scenario simulation: theory and methodology. *Finance Stoch* 1(1):43–67
- Johnson RA (1988) Stress–strength models for reliability. In: Krishnaiah PR, Rao CR (eds) *Handbook of statistics*, vol 7. Elsevier, Amsterdam, pp 27–54
- Kennan J (2006) A note on discrete approximations of continuous distributions. University of Wisconsin, Madison
- Lloyd S (1982) Least squares quantization in PCM. *IEEE Trans Inform Theory* 28(2):129–137
- Luceno A (1999) Discrete approximations to continuous univariate distributions: an alternative to simulation. *J R Stat Soc Ser B* 61(2):345–352
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative risk management: concepts, techniques and tools*, 2nd edn. Princeton University Press, Princeton
- Miller AC, Rice TR (1983) Discrete approximations of probability distributions. *Manag Sci* 29(3):352–362
- Narasimhan B, Johnson SG, Hahn T, Bouvier A, Kiêu K (2022). *cutature: Adaptive Multivariate Integration over Hypercubes*. R package version 2.0.4.4. <https://CRAN.R-project.org/package=cutature>
- Owen DB (1980) A table of normal integrals. *Commun Stat Simul Comput* 9(4):389–419
- Panjer HH (1981) Recursive evaluation of a family of compound distributions. *ASTIN Bull* 12(1):22–26
- Pavlikov K, Uryasev S (2018) CVaR distance between univariate probability distributions and approximation problems. *Ann Oper Res* 262(1):67–88
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Rizzo ML, Székely GJ (2016) Energy distance. *Wiley Interdiscip Rev* 8(1):27–38
- Roy D, Dasgupta T (2001) A discretizing approach for evaluating reliability of complex systems under stress–strength model. *IEEE Trans Reliab* 50(2):145–150
- Schrempf OC, Brunn D, Hanebeck UD (2006) Dirac mixture density approximation based on minimization of the weighted Cramér–von Mises distance. In: 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp 512–517
- Stein EM, Shakarchi R (2011) *Functional analysis: introduction to further topics in analysis*. Princeton University Press, Princeton

- Tanaka KI, Toda AA (2013) Discrete approximations of continuous distributions by maximum entropy. *Econom Lett* 118(3):445–450
- Tanaka KI, Toda AA (2015) Discretizing distributions with exact moments: error estimate and convergence analysis. *SIAM J Numer Anal* 53(5):2158–2177
- von Mises R (1931) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Mary S, Rosenberg

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.