PRAMANA — journal of physics

Vol. 85, No. 6
December 2015
pp. 1073–1087

On the discretization of probability density functions and the continuous Rényi entropy

DIÓGENES CAMPOS
Academia Colombiana de Ciencias Exactas, Físicas y Naturales, ACCEFYN, Bogotá, Colombia
E-mail: dcamposr@cable.net.co; dcr257@gmail.com

MS received 27 June 2014; revised 11 September 2014; accepted 27 September 2014
DOI: 10.1007/s12043-015-0955-2; ***e*Publication:** 19 July 2015

Abstract. On the basis of second mean-value theorem (SMVT) for integrals, a discretization method is proposed with the aim of representing the expectation value of a function with respect to a probability density function in terms of the discrete probability theory. This approach is applied to the continuous Rényi entropy, and it is established that a discrete probability distribution can be associated to it in a very natural way. The probability density functions for the linear superposition of two coherent states is used for developing a representative example.

Keywords. Probability density function; Rényi entropy; differential (continuous) entropy; discretization method; mean-value theorem for integrals.

PACS Nos **02.70.−c; 02.50.−r; 03.65.−w**

1. Introduction

Consider a continuous random variable Ξ with a known probability density function $\rho(x) \geq 0$, where the range $a \leq x \leq b$ can be finite, semi-infinite or infinite. The mean value (or expectation value) of an observable property described by the function $f(x)$ is defined by

$$\langle f(x) \rangle := \int_a^b f(x)\rho(x)\mathrm{d}x \,, \quad \int_a^b \rho(x)\mathrm{d}x = \omega, \tag{1}$$

where the constant $0 < \omega \leq 1$ is the degree of completeness assigned to $\rho(x)$.

In addition to the analytical or numerical computation of mean values, eqs (1) are important for most applications or theoretical problems of interest. In statistics, calculation of the moments $\langle x^\nu \rangle$ of a variable x, where $\nu \geq 0$ is an integer, is of interest. In quantum mechanics, the expectation value of an observable quantity described by the function $f(x)$ with respect to a probability density function (PDF) $\rho(x) := |\psi(x)|^2$, where $\psi(x)$ is the wave function in coordinate representation, is often required.

In probability theory, statistics, statistical mechanics, communication theory, and other fields of science, the calculation of Rényi and Tsallis entropies [1–3] for probability density function $\rho(x)$ involves integral $\int_a^b [\rho(x)]^q \, \mathrm{d}x$, where $q \geq 0$ is a parameter.

The aim of this paper is to present a procedure for the discretization of eqs (1) based on the second mean-value theorem (SMVT) for integrals by postulating that: (i) The PDF $\rho(x)$ can be represented by a discrete probability distribution $P := \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$ with $\mathcal{N}$ elements. (ii) The expectation value $\langle f(x) \rangle$ can be described analogously to the case of a discrete probability distribution, i.e., $\langle f(x) \rangle = \sum_{\ell=1}^{\mathcal{N}} f_\ell P_\ell$, where $f := \{f_1, f_2, \ldots, f_{\mathcal{N}}\}$ is a suitable set of values of the function $f(x)$. (iii) The expected value obtained by using the sum must match the value $\langle f(x) \rangle$ given by the integral in eq. (1). Hereafter the notation $\omega_{\mathcal{N}}$ is used instead of ω. Thus, the degree of completeness of the density function $\rho(x)$ is described by the relation $\sum_{m=1}^{\mathcal{N}} P_m = \omega_{\mathcal{N}}$, where P can be complete ($\omega_{\mathcal{N}} = 1$) or incomplete ($0 < \omega_{\mathcal{N}} < 1$).

According to Shannon's fundamental uniqueness theorem [4], the entropy $S(P) := -k_{\mathrm{B}} \sum_{n=1}^{\mathcal{N}} \ln(P_n) P_n$ is uniquely defined for discrete probability distributions $P := \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$ of an exhaustive set of $\mathcal{N}$ mutually exclusive events (k_{B} is a conventional positive constant). The discrete entropy is a measure of the 'amount of uncertainty' in a probability distribution P [5], and $0 \leq S(P) \leq \mathcal{N}$. In the search for the analogous expression of $S(P)$ in the case of a continuous variable x, certain problems arise in the transition $P \to \rho(x)$, i.e., from discrete probability distributions P to continuous probability density functions $\rho(x)$:

(1) To begin, consider the integral $H := -k_{\mathrm{B}} \int_a^b \rho(x) \ln(\rho(x)) \mathrm{d}x$ proposed in 1948 by Shannon [4] as a measure of relative uncertainty for a continuous random variable x. This quantity H, known as continuous (or differential) entropy of $\rho(x)$, has become one of the most important measures in quantum mechanics and other fields of science. When the interval $a \leq x \leq b$ is divided into bins, each one of length Δ and within each bin $[n\Delta, (n+1)\Delta]$ the mean value theorem is used, then the discretized entropy H_Δ is given by $H_\Delta = -k_{\mathrm{B}} \left[\sum_n \rho(\xi_n) \ln(\rho(\xi_n) \Delta - \ln(\Delta) \right]$. The first term approaches the entropy of a discrete probability distribution P, and if the number of elements of P increases more and more, then $\Delta \to 0$, the extra term $-\ln(\Delta)$ goes to ∞, and H differs from H_Δ by an amount $-\infty$ (see [6], §2.2). In conclusion, the Shannon continuous entropy (H) cannot be interpreted as the limit of the Shannon entropy of a discrete probability distribution P with $\mathcal{N}$ elements, when $\mathcal{N} \to \infty$, and both quantities differ by an infinite offset.

(2) In his attempt to change from a discrete distribution P to a continuous probability distribution $\rho(x)$, Jaynes encounters a similar problem (see [5], §4.b). He removes an infinite constant arising from a term $\ln(\mathcal{N})$ to define the finite continuous entropy $H_{\mathrm{c}} := -k_{\mathrm{B}} \lim_{\mathcal{N} \to \infty} [\sum_{n=1}^{\mathcal{N}} \ln(P_n) P_n - \ln(\mathcal{N})] = -k_{\mathrm{B}} \int_a^b \ln(\rho(x)/m(x)) \, \rho(x) \mathrm{d}x$. He assumes that the density of discrete points approaches a reference measure function $m(x)$ that can be interpreted as 'the prior distribution describing complete ignorance of x' (see [7], §VI). H_{c} is invariant under the change of variables but there is an ambiguity in determining $m(x)$ because it is not the result of some definite and explicit limit.

In this work, instead of the transition $P \to \rho(x)$, the transformation $\rho(x) \to P$ is considered. From the numerical point of view, there are different powerful discretization methods for eq. (1), e.g., approximation formulae for the mean and variance or methods based on Gaussian quadrature with N points which can handle $2N$ moments [8]. Instead of trying to improve the existing discretization methods, the focus of this paper is the discretization of eq. (1) by taking advantage of SMVT as a general mathematical approach. Working on this problem is important, e.g., in quantum physics, a theory in which most observable quantities have a continuous spectra or mixed spectra, i.e., one which contains both discrete and continuous components. One should assess the contributions of the two parts of the spectrum of an observable quantity on a common definition, including the case of the entropy. However, as indicated earlier, discrete and continuous Shannon and Jaynes entropies do not share the same definition and when $\mathcal{N} \to \infty$ they differ by an infinite offset.

In this study, considerations are restricted to the Rényi entropy of order q ($q \geq 0$, $q \neq 1$) for the probability density function $\rho(x)$ [1,2] (see the definition of $S_{\mathrm{R}}(q,h)$ in eq. (11)). The Rényi entropy is a one-parameter axiomatic generalization of the Shannon entropy that has been successfully applied in many fields such as physics, engineering, economics and biology. The Shannon continuous entropy H results as a special case of $S_{\mathrm{R}}(q,h)$ in the limit $q \to 1$.

The proposed method is applied to a quantum mechanical particle in a state $\Phi(x)$ described by the superposition of coherent states, which implies a probability density function $\rho(x) = |\Phi(x)|^2$. This system supports a probabilistic description and exhibits quantum mechanical interference effects, a distinguishing feature of a quantum system. The idea underlying this work is that one can differentiate and characterize quantum and classical behaviours by using the Rényi entropy [1,2] associated with the finite discrete probability distribution P that emerges by using the SMVT discretization procedure.

This paper is organized as follows. Section 2 describes the basis of discretization method of the probability density function $\rho(x)$. Section 3 deals with the Shannon and Rényi entropies [1,4] of a continuous distribution with PDF $\rho(x)$. Section 4 considers an application to a quantum system described by a linear superposition of two coherent states. Finally, §5 contains some additional comments.

2. Discretization method

2.1 *Basis of the method*

The starting point is the SMVT for integrals [9–13], i.e., if $f(x)$ is strictly monotonic in the interval (p,q), where $p < q$, and if $\rho(x) \geq 0$ is integrable over that interval, then there exists at least one point ξ (with $p \leq \xi \leq q$) such that

$$\int_p^q f(x)\rho(x)\mathrm{d}x = f(p)\int_p^{\xi} \rho(x)\mathrm{d}x + f(q)\int_{\xi}^q \rho(x)\mathrm{d}x. \tag{2}$$

This theorem applies to a strictly monotonic function, i.e., defining the increment of $f(x)$ as $\Delta f(x) := f(x') - f(x)$, for $x' - x > 0$, then $f(x)$ is either strictly increasing ($\Delta f(x) > 0$) or strictly decreasing ($\Delta f(x) < 0$) in the interval $p \leq x \leq q$. When $f(x)$ is a constant (say, c), choose $\xi = (q+p)/2$ as the midpoint of the interval $p \leq x \leq q$ and $f(q) = f(p) = c$. By convention, other statements in [9] are formally included in eq. (2) with suitable reinterpretation.

Now consider the interval $[p, q] := p \leq x \leq q$ and associate to it the quantity

$$\mathcal{P}([p, q]) := \int_p^q \rho(x)\mathrm{d}x = \mathcal{P}([a, q]) - \mathcal{P}([a, p]). \tag{3}$$

As $\rho(x)$ is a PDF, then $\rho(x) \geq 0$ and $\mathcal{P}([p, q])$ is the probability of choosing one value of x in the range $p \leq x \leq q$. The quantity $\mathrm{CDF}(u) := \mathcal{P}([a, u])$, defined for $a \leq u \leq b$, is known as the cumulative distribution function (CDF) of the continuous random variable Ξ. The PDF and the CDF contain all the probabilistic information about Ξ; i.e., the probability distribution of Ξ can be described by either of them.

2.2 *Natural discretization of $\rho(x)$ induced by $f(x)$*

Consider a piecewise monotonic function $f(x)$ on the interval $[a, b]$, with $a < b$, and assume that: (1) The function $f(x)$ has a minimal number of different regions or segments ($N \geq 1$) defined by a set of points $a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$ covering the whole interval $[a, b]$, (2) on each segment $[x_{n-1}, x_n]$, $f(x)$ is strictly increasing, strictly decreasing or a constant. By convention, the bending points of $f(x)$ are the points $x_0, x_1, x_2, \ldots, x_N$ between the segments (including x_0 and x_N), i.e., the points in which the curve $f(x)$ changes its course (figure 1). The N subintervals $[x_{n-1}, x_n]$ are named the 'regions' or 'segments' of $f(x)$, with $n = 1, 2, \ldots, N$.

The mean value $\langle f(x)\rangle$ given by eq. (1) can be written as $\langle f(x)\rangle = \sum_{n=1}^{N} I_n$, where $I_n := \int_{x_{n-1}}^{x_n} f(x)\rho(x)\mathrm{d}x$. Under the conditions of SMVT, for the nth segment there is a point ξ_n ($x_{n-1} < \xi_n < x_n$) such that $I_n = f(x_{n-1})\,\mathcal{P}([x_{n-1}, \xi_n]) + f(x_n)\mathcal{P}([\xi_n, x_n])$. Then, eq. (1) can be written as

$$\langle f(x)\rangle = \sum_{n=1}^{N} \left[f(x_{n-1})\,\mathcal{P}([x_{n-1}, \xi_n]) + f(x_n)\mathcal{P}([\xi_n, x_n]) \right]. \tag{4}$$

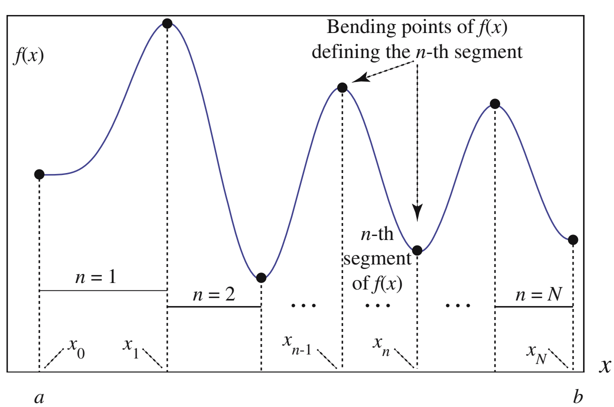

Figure 1. A piecewise monotonic function $f(x)$ on the interval $[a, b]$ is partitioned into N segments $a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$, in which $f(x)$ is either monotonic increasing, monotonic decreasing or a constant. The interval $x_{n-1} \leq x \leq x_n$ is named the nth segment or region.

This result is called the 'natural discretization' of $\rho(x)$ induced by the function $f(x)$. By 'natural', one means that the only assumptions involved in eq. (4) are the ones established by the SMVT, and that this theorem is applied to each of the complete segments of $f(x)$. It is to be noted that the nth segment contributes with the weights $\mathcal{P}([x_{n-1}, \xi_n])$ and $\mathcal{P}([\xi_n, x_n])$, and function $f(x)$ with the values $f(x_{n-1})$ and $f(x_n)$. As annotated in §5, the segments of $f(x)$ may be further subdivided into subintervals to have an 'artificial discretization' of $\rho(x)$, which may be necessary in some circumstances to improve the numerical results.

2.3 *Discrete probability distribution assigned to the segments of $f(x)$*

To gain further insight into the meaning of the results obtained, it is convenient to consider the nth segment of the function $f(x)$, i.e., the interval $[x_{n-1}, x_n]$. The relations $f(x_{n-1})\,\mathcal{P}([x_{n-1}, \xi_n]) + f(x_n)\,\mathcal{P}([\xi_n, x_n]) = I_n$ and $\mathcal{P}([x_{n-1}, \xi_n]) + \mathcal{P}([\xi_n, x_n]) = J_n$ form a system of equations, for which one assumes that $f(x_{n-1})$ and $f(x_n)$ are bounded, and that the following integrals can be evaluated analytically or numerically:

$$I_n := \int_{x_{n-1}}^{x_n} f(x)\rho(x)\mathrm{d}x, \quad J_n := \int_{x_{n-1}}^{x_n} \rho(x)\mathrm{d}x. \tag{5}$$

As the nth segment of $f(x)$ is strictly monotonic and $f(x_n) - f(x_{n-1}) \neq 0$ is bounded, the solution of the system of equations is given by

$$\mathcal{P}([x_{n-1}, \xi_n]) = +\frac{f(x_n)\,J_n - I_n}{f(x_n) - f(x_{n-1})},$$
$$\mathcal{P}([\xi_n, x_n]) = -\frac{f(x_{n-1})\,J_n - I_n}{f(x_n) - f(x_{n-1})}. \tag{6}$$

Otherwise, if the nth segment of $f(x)$ is a constant ($f(x) = c$) in the whole interval $x_{n-1} \le x \le x_n$, then by taking $\xi_n = (x_{n-1} + x_n)/2$ as the midpoint, one gets $\mathcal{P}([x_{n-1}, \xi_n]) = \mathcal{P}([\xi_n, x_n]) = J_n/2$.

In general, the existence of an appropriate value ξ_n, between x_{n-1} and x_n, guarantees that the nth segment of the function $f(x)$ is divided into two parts, $x_{n-1} \le x \le \xi_n$ and $\xi_n \le x \le x_n$, and that the weights $\mathcal{P}_n([x_{n-1}, \xi_n])$ and $\mathcal{P}_n([\xi_n, x_n])$ can be assigned to them. Then, if the function $f(x)$ has N segments, a discrete probability distribution $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$ and discrete variable $\{f_1, f_2, \ldots, f_{\mathcal{N}}\}$ are introduced each with $\mathcal{N} = 2N$ elements, where the nth segment of $f(x)$ contributes with one odd and one even element, namely

$$P_{2n-1} := \mathcal{P}([x_{n-1}, \xi_n]), \quad P_{2n} := \mathcal{P}([\xi_n, x_n]),$$
$$f_{2n-1} := f(x_{n-1}), \quad f_{2n} := f(x_n), \quad \text{for } n = 1, 2, \ldots, N. \tag{7}$$

By using this change of notation, in eq. (1) the mean value $\langle f(x)\rangle$ can be expressed in an analogous manner to the case of a discrete probability distribution, i.e.,

$$\langle f(x)\rangle = \sum_{n=1}^{N} \left[f_{2n-1} P_{2n-1} + f_{2n} P_{2n} \right] = \sum_{\ell=1}^{\mathcal{N}} f_\ell\, P_\ell\,, \tag{8}$$

where the odd and even contributions have been unified into a single sum with summation index ℓ. The normalization condition of P becomes $\sum_{n=1}^{N}(P_{2n-1}+P_{2n}) = \sum_{\ell=1}^{\mathcal{N}} P_\ell = \omega_{\mathcal{N}}$ with the degree of completeness $\omega_{\mathcal{N}} := \omega$.

The quantity $S(P|\rho, f) = -(k_{\mathrm{B}}/\omega_{\mathcal{N}})\sum_{n=1}^{N}\left[\ln(P_{2n-1})P_{2n-1} + \ln(P_{2n})P_{2n}\right]$ is the Shannon entropy of the discrete probability distribution $P := \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$, where the positive constant k_{B} fixes the unit of measurement of entropy. Thus, one gets

$$S(P|\rho, f) = -\,(k_{\mathrm{B}}/\omega_{\mathcal{N}})\sum_{\ell=1}^{\mathcal{N}} \ln(P_\ell)P_\ell\,, \tag{9}$$

where the notation $S(P|\rho, f)$ means that for a given $\rho(x)$, the piecewise monotonic function $f(x)$ generates a discrete probability set P to which corresponds a Shannon entropy $S(P|\rho, f)$.

2.4 *Moments of $f(x)$*

Under the assumptions of eq. (2), to calculate the moments of $f(x)$ one may be interested in applying the SMVT to the function $g(x) := \left[f(x)\right]^\nu$, where ν is an integer, $\nu \geq 1$. It is to be noted that not all the bending points of $f(x)$ belong to $g(x)$ but depending on the function $f(x)$ and the exponent n, $g(x)$ may have some additional bending points. For e.g., (a) $x = 0$ is not a bending point of $f(x) = x$ but it is a bending point of $g(x) = x^2$, if $a < x < b$ and $a < 0$ and $b > 0$. (b) The exponent $n = 2$ transforms the odd function $f(x) = \sin(x)$ into the even function $g(x) = \sin^2(x)$, and in the interval $-2\pi \leq x \leq 2\pi$ they exhibit six and nine bending points, respectively. Also it is to be noted that even for common intervals eqs (6) generate different weights for $f(x)$ and $g(x)$. Thus, the calculation of variance and moments of $f(x)$ leads to different weights $\mathcal{P}_n([x_{n-1}, \xi_n(\nu)])$ and $\mathcal{P}_n([\xi_n(\nu), x_n])$, because the points ξ_n involved in the SMVT depend not only on the PDF $\rho(x)$ but also on the function in consideration, $f(x)$ or $g(x) = \left[f(x)\right]^\nu$.

In general, (i) the discretization ($\rho(x) \xrightarrow{f(x)} P$) of a given PDF $\rho(x)$ into a discrete probability distribution $P = \{P_1, P_2, \ldots, P_{\mathcal{N}}\}$ depends on the function $f(x)$ and its N segments, where each segment contributes with two elements and $\mathcal{N} = 2N$. (ii) Equation (8) can be interpreted as the $\mathcal{N}$-dimensional discrete representation of the expectation value $\langle f(x)\rangle$ given by eq. (1), which involves not only $\rho(x)$ but also $f(x)$. (iii) The given two different functions $f_\alpha(x)$ and $f_\beta(x)$ lead to distinct sets of discrete probabilities (P_α and P_β) and their dimensions ($\mathcal{N}_\alpha \neq \mathcal{N}_\beta$) are generally different, except if the functions have the same number of segments. (iv) Under the conditions (iii), as $\langle f_\alpha(x)\rangle \neq \langle f_\beta(x)\rangle$ it is natural that the set of discrete probabilities associated with $\rho(x)$ will not be unique. Note also that, according to eqs (7), the discretization applies not only to the PDF $\rho(x)$ but also to the function $f(x)$.

3. On the continuous Shannon and Rényi entropies

Consider the function $\mathcal{H}(x, h) := -k_{\mathrm{B}}\ln(h\rho(x))$ and define the differential entropy (or continuous entropy) for a PDF $\rho(x)$ as [4,6,14–16]

$$H\,(\rho, h) := -(k_{\mathrm{B}}/\omega)\int_a^b \ln(h\rho(x))\rho(x)\mathrm{d}x. \tag{10}$$

For $h = 1$, eq. (10) reduces to the standard definition of differential entropy and for $h > 0$, to the modified differential entropy proposed in [15], in which the range $a \le x \le b$ is divided into bins of width h. The constant $k_B > 0$ fixes the unit of measurement of $H(\rho, h)$, $\ln(\cdots)$ is the natural logarithm giving the information in nats, and by convention $0 \ln 0 = 0$ (for visual representation, plot $x \ln(x)$).

For a segment of $\mathcal{H}(x, h)$ defined by an interval $[x_{n-1}, x_n]$ in which either $\rho(x_{n-1}) \to 0$ or $\rho(x_n) \to 0$, then $\mathcal{H}(x_{n-1}, h) \to \infty$ or $\mathcal{H}(x_n, h) \to \infty$ but, according to the convention, $\mathcal{H}(x_{n-1}, h) \times 0 \to 0$ or $\mathcal{H}(x_n, h) \times 0 \to 0$. The system of equations in §2.3 cannot be formulated and, therefore, eqs (6) are not directly applicable to that segment.

3.1 *Discretization of the continuous Rényi entropy*

A method to deal with the case $\mathcal{H} \to \infty$ is to take advantage of the continuous Rényi entropy of order q ($0 \le q < \infty, q \ne 1$) defined as

$$S_R(q, h) := \frac{k_B}{1-q} \ln\left(\frac{h^{q-1}}{\omega} \int_a^b [\rho(x)]^q \, dx\right) = S_R(q, 1) - k_B \ln(h), \tag{11}$$

where $S_R(q) := S_R(q, 1)$ is the standard Rényi entropy for a PDF of degree of completeness ω [1,2]. In the limit $q \to 1$, the entropy $S_R(q, 1)$ becomes a 0/0 indeterminate form, and by using the L'Hôpital rule and the relation $\partial(\rho(x))^q / \partial q = \ln(\rho(x)) (\rho(x))^q$, one gets $S_R(1, h) := \lim_{q \to 1} S_R(q, h) = H(\rho, h) = H(\rho, 1) - k_B \ln(h)$.

The function $f(x, q) := [\rho(x)]^q$, the identity $\int_a^b [\rho(x)]^q \, dx = \int_a^b f(x, q-1)\rho(x) dx$, the notation $\omega_{\mathcal{N}} = \omega$, and eqs (8) and (7) give

$$S_R(q,h) = \frac{k_B}{1-q} \ln\left(\frac{h^{q-1}}{\omega_{\mathcal{N}}} \sum_{n=1}^{N} \left[f(x_{n-1}, q-1) P_{2n-1}(q) + f(x_n, q-1) P_{2n}(q)\right]\right). \tag{12}$$

In this equation, $I_n(q) := \int_a^b [\rho(x)]^q \, dx$, the entities $P_\ell(q)$ and $f_\ell(q)$ are determined by applying eqs (6) and (7) to the function $f(x, q-1) := [\rho(x)]^{q-1}$, and $q \ge 0$ is a parameter sometimes called the entropic index.

To gain an additional perspective, in eq. (12) consider the limit $q \to 1$ by introducing the notation $u'(q) := \partial u(q)/\partial q$ and using the L'Hôpital rule. According to §2.3, the nth summand in eq. (12) can be written as $A_n := \left[\rho(x_{n-1})\right]^{q-1} P_{2n-1}(q) + [\rho(x_n)]^{q-1} P_{2n}(q) = I_n(q)$. Differentiation of this relation with respect to q implies $A'_n(q) = I'_n(q) = \int_a^b \ln(\rho(x)) [\rho(x)]^q \, dx$ and, after completing the calculation of the limit, one gets $S_R(1, h) = \sum_{n=1}^{N} \int_{x_{n-1}}^{x_n} \mathcal{H}(x, h)\rho(x) dx = H(\rho, h)$, as expected.

By applying eqs (6) and (7) to the function $f(x, q-1)$ one gets the weights $P_{2n-1}(q)$ and $P_{2n}(q)$, for $n = 1, 2, \ldots, N$. In particular, for $q = 0$, as $\rho(x_{n-1}) = 0$ implies $P_{2n-1}(0) = 0$ and $\rho(x_n) = 0$ results in $P_{2n}(0) = 0$, then $S_R(0, h)$ is the logarithm of the nonzero summands in the sum of eq. (12). For the limit value $q \to 1$, the Hôpital rule with the auxiliary quantity $H_n := -\int_{x_{n-1}}^{x_n} \ln(\rho(x)) \rho(x) dx$, gives

$$P_{2n-1}(1) = +\frac{\ln(\rho(x_n)) J_n + H_n}{\ln(\rho(x_n)) - \ln(\rho(x_{n-1}))},$$

$$P_{2n}(1) = -\frac{\ln(\rho(x_{n-1}))J_n + H_n}{\ln(\rho(x_n)) - \ln(\rho(x_{n-1}))}. \tag{13}$$

In conclusion, the continuous Rényi entropy $S_R(q, h)$ can be represented by the discrete expression given by eq. (12) and, in the limit $q \to 1$, it also retrieves the continuous entropy $H(\rho, h) = H(\rho, 1) - k_B \ln(h)$. As the parameter $h > 0$ is arbitrary, its role is to set a 'reference value' for the measurement of $S_R(q, h)$ and $H(\rho, h)$ entropies.

4. An illustrative example with a quantum system

4.1 *On the normal (or Gaussian) distribution*

Consider a coherent state of the harmonic oscillator described by the wave function

$$\phi(x; \mu, \Delta x, \bar{p}) = \frac{1}{\sqrt{\sqrt{2\pi}\,\Delta x}} \exp\left(-\frac{1}{4}\left(\frac{x-\mu}{\Delta x}\right)^2\right) \exp\left(\frac{i}{\hbar} x \bar{p}\right), \tag{14}$$

defined for $-\infty < x < \infty$ with mean position $\bar{x} = \mu$, uncertainty $\Delta x := \sigma$ (i.e., variance σ^2) and mean momentum $\bar{p}$. In terms of the mass m and frequency Ω of the oscillator the quantities $x_0 = \sqrt{\hbar/(m\Omega)} = \sqrt{2}\,\Delta x$ and $p_0 = \sqrt{\hbar m \Omega} = \sqrt{2}\,\Delta p$ satisfy the minimum uncertainty relation $x_0 p_0 = \hbar$ or $\Delta x \Delta p = \hbar/2$.

The PDF associated with $\phi(x; \mu, \Delta x, \bar{p})$ is given by

$$\rho(x; \mu, \Delta x) = \frac{1}{\sqrt{2\pi}\,\Delta x} \exp\left(-\frac{(x-\mu)^2}{2(\Delta x)^2}\right) \tag{15}$$

and, therefore $f(x, q) := [\rho(x; \mu, \sigma)]^q = \rho\left(x; \mu, \sigma/\sqrt{q}\right)/(\sqrt{2\pi}\,\sigma)^{q-1}$. Consequently, $f(x, q)$ has two segments defined by the points $\{x_0, x_1, x_2\}$, where $x_0 \to -\infty$, $x_1 = \mu$ and $x_2 \to \infty$.

Consider the integral of $f(x, q)$ from $-\infty$ to μ (segment 1) and from μ to ∞ (segment 2). Integration of $\rho(x; \mu, \sigma) = f(x, 1)$ gives $J_1 = J_2 = 1/2$, and integration of $f(x, q)$ leads to $I_1(q) = I_2(q) = \left[\pi^{1-q}/(2^{1+q}q)\right]^{1/2} \sigma^{1-q}$. Recalling the details in §3.1 (after eq. (12)) evaluation at the points $\{x_0, x_1, x_2\}$ gives $f(\pm\infty, q-1) = 0$ and $f(\mu, q-1) = (2\pi\sigma^2)^{(1-q)/2}$. Thus, eqs (5) and (6) imply the existence of a discrete probability set $P(q) := \{P_1(q), P_2(q), P_3(q), P_4(q)\}$ with elements $P_2(q) = P_3(q) = 1/(2\sqrt{q})$ and $P_1(q) = P_4(q) = 1/2 - 1/(2\sqrt{q})$. As the weights $P_1(q)$ and $P_4(q)$ do not contribute to the sum in eq. (12), the Rényi entropy is given by

$$S_R(q, h) = \frac{k_B}{1-q} \ln\left((2\pi\,\sigma^2)^{(1-q)/2} h^q/\sqrt{q}\right), \quad q > 0, \quad q \neq 1. \tag{16}$$

On other hand, by using eq. (9) one gets the discrete Shannon entropy associated with the probability distribution $P(q)$, defined for $q > 0$ as

$$S(P|\rho, f) = -k_B\left[\left(1 - \frac{1}{\sqrt{q}}\right)\ln\left(1 - \frac{1}{\sqrt{q}}\right) + \frac{1}{\sqrt{q}}\ln\left(\frac{1}{\sqrt{q}}\right) - \ln(2)\right].$$

When q increases, Rényi entropy decreases monotonously and for $q \to \infty$ it reaches the value $(k_B/2)\ln\left(2\pi(\sigma/h)^2\right)$, whereas $S(P|\rho, f)$ takes the value $k_B \ln 2$ when $q \to 1$, reaches a maximum at $q = 4$ and the value $k_{\rm B}\ln(2)$ when $q \to \infty$; both entropies are nonnegative. Finally, by taking the limit $q \to 1$ and using the relation $\lim_{q\to 1} S_{\rm R}(q, h) = H(\rho, h)$, one gets the Shannon entropy $H(\rho, h) = (k_{\rm B}/2)\ln(2\pi e(\Delta x/h)^2)$ nats, which is negative for an effective uncertainty $\sigma_{\rm eff} := \Delta x/h$ such that $2\pi e(\Delta x/h)^2 < 1$, i.e., $\sigma_{\rm eff} < 1/\sqrt{2\pi e} \approx 0.241971$.

At this point it is appropriate to emphasize, as mentioned in §1, that the Shannon-continuous entropy $H(\rho, h)$ is not the limit of the Shannon entropy $S(P)$ of a discrete probability distribution P with $\mathcal{N}$ elements, when $\mathcal{N} \to \infty$. As a result of the definition given by eq. (10), $H(\rho, h)$ can be positive or negative, contrary to $S(P)$ that remains nonnegative. Negative values of H are not associated with the violation of Heisenberg uncertainty principle $\Delta x \Delta p \geq \hbar/2$, because the Gaussian wave function $\phi(x; \mu, \Delta x, \bar{p})$ defined by eq. (14) describes a state of minimum uncertainty: $\Delta x = \sigma/\sqrt{2}$ and $\Delta p = \hbar/(\sigma\sqrt{2})$. Similarly, the continuous Shannon entropies (H and $\tilde{H}$) satisfy the uncertainty relation [17] $H(\rho, 1) + \tilde{H}(\tilde{\rho}, 1) = k_{\rm B}\ln(\pi e)$, where ρ and $\tilde{\rho}$ are the probability densities in position-space and wave-vector space (i.e., $|k\rangle := \sqrt{\hbar}|p\rangle|_{p=\hbar k}$, where $|p\rangle$ is a momentum eigenstate).

4.2 *Comparison between quantum and classical behaviours*

In this section, consider the function $\Phi(x) = N(\alpha_1\phi_1 + \alpha_2\phi_2)$ that describes a quantum superposition of two coherent states, $\phi_1 := \phi(x; \mu_1, \sigma_1, \bar{p}_1)$ and $\phi_2 = \phi(x; \mu_2, \sigma_2, \bar{p}_2)$, where α_1 and α_2 in general, are complex numbers and N is the normalization constant. The PDF is given by

$$\rho(x) = |\Phi(x)|^2 = |N|^2\left[|\alpha_1|^2|\phi_1|^2 + |\alpha_2|^2|\phi_2|^2 + \mathcal{I}_{12}\right], \tag{17}$$

where $\mathcal{I}_{12}(x) = \alpha_1^\star\alpha_2\phi_1^\star\phi_2 + \alpha_1\alpha_2^\star\phi_1\phi_2^\star$ is the interference term, and

$$\phi_1^\star\phi_2 = \phi\left(0; \mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}, 0\right)\phi(x; \mu, \sigma, \bar{p}_2 - \bar{p}_1), \tag{18}$$

with

$$\sigma = \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}, \quad \mu = \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right)\sigma^2 = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}. \tag{19}$$

In eq. (18), one uses the relation $\phi(x; \mu, \sigma, \bar{p}_2 - \bar{p}_1) = \phi(x; \mu, \sigma, 0)\exp(-ix\bar{p}/\hbar)$ with $\bar{p} := \bar{p}_1 - \bar{p}_2$, and the Fourier transform gives

$$\begin{aligned}\tilde{\phi}(\bar{p}, \Delta p, \mu) &:= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}\phi(x; \mu, \sigma, 0)\exp(-ix\bar{p}/\hbar)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{\sqrt{2\pi}\,\Delta p}}\exp\left(-\frac{1}{4}\left(\frac{\bar{p}}{\Delta p}\right)^2\right)\exp\left(-\frac{i}{\hbar}\mu\,\bar{p}\right),\end{aligned} \tag{20}$$

where $\Delta p = \hbar/(2\sigma)$. Then the normalization constant can be written as

$$N = \left[|\alpha_1|^2 + |\alpha_2|^2 + \phi\left(0; \mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}, 0\right)N_{12}\right]^{-1/2}, \tag{21}$$

Table 1. Parameters adopted for describing the state $\Phi(x)$ and PDFs $\rho(x) = |\Phi(x)|^2$ and $\rho_{\rm cl}(x)$.

μ_1	σ_1	p_1	μ_2	σ_2	p_2	α_1	α_2
6.5	0.8	−3.0	9.0	0.7	3.0	0.5	0.5

with the auxiliary quantity

$$N_{12} := 2\sqrt{2\pi\hbar}\,\tilde{\phi}\,(\bar{p}_1 - \bar{p}_2, \Delta p, 0)\,\mathrm{Re}\left(\alpha_1^\star \alpha_2 \exp\left(-\frac{i}{\hbar}\mu(\bar{p}_1 - \bar{p}_2)\right)\right).$$

By omitting in eqs (17) and (21) the contributions due to the interference term, one gets the classical PDF as

$$\rho_{\rm cl}(x) = \frac{1}{|\alpha_1|^2 + |\alpha_2|^2}\left[|\alpha_1|^2|\phi_1|^2 + |\alpha_2|^2|\phi_2|^2\right]. \tag{22}$$

For illustrational purposes, table 1 specifies the values assigned to the parameters, and in figure 2 $\rho(x)$ and $\rho_{\rm cl}(x)$ are plotted as a function of x. This graph clearly shows that the interference term $\mathcal{I}_{12}(x)$ generates four additional bending points in $\rho(x)$ in comparison with the number of bending points that $\rho_{\rm cl}(x)$ exhibits. By following the method described in §3, one sets the functions $f(x,q) := [\rho(x)]^q$, $\rho(x) = f(x,1)$, and similar relationships for $\rho_{\rm cl}(x)$.

First, one looks for the values $\{x_1, x_2, \ldots, x_N\}$ in which the function $f(x, q-1)$ exhibits maxima and minima. The results are shown in table 2 for Cases (a) and (b):

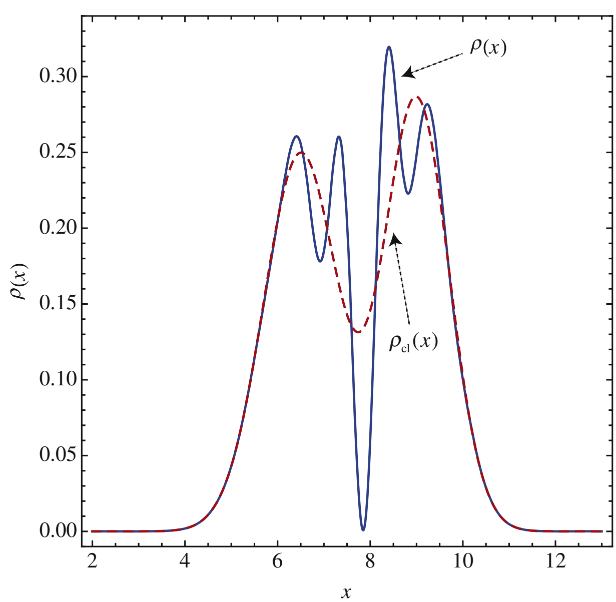

Figure 2. PDFs $\rho(x) = |\Phi(x)|^2$ and $\rho_{\rm cl}(x)$ given by eqs (17) and (22), respectively, calculated with the parameters shown in table 1.

Table 2. Set of points $\{x_0, x_1, \ldots, x_N\}$ defining the segments of $\rho(x)$ and $\rho_{\text{cl}}(x)$ for the parameters in table 1. The points $x_0 \to -\infty$ and $x_N \to \infty$ are approximated by finite values because at these points the probability density attains zero value.

Case	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
(a)	2.0	6.409	6.919	7.322	7.848	8.404	8.820	9.230	13.0
(b)	2.0	6.507	7.741	8.987	13	–	–	–	–

$\rho(x)$ and $\rho_{\text{cl}}(x)$ exhibit $N = 8$ and 4 segments, respectively. Second, by applying formulae (6) and (7) one gets the weights $P(q) := \{P_1(q), P_2(q), \ldots, P_{2N}(q)\}$ assigned to N segments of the function $f(x, q-1)$ and for Cases (a) and (b) of table 2, the results are displayed in figure 3. Third, after the evaluation of the coefficients $f_{2n-1}(q) = f(x_{n-1}, q-1)$ and $f_{2n}(q) = f(x_n, q-1)$, the discretization of the Rényi entropy proceeds according to eq. (12).

Figure 3 shows that the weights satisfy the normalization condition, $\omega_{\mathcal{N}} = \sum_{\ell=1}^{\mathcal{N}} P_\ell(q) = 1.0$ for all the values of q. For $q \to 0$ the importance of some weights increases, whereas the significance of other weights decreases to zero. As a general rule, if $\rho(x_{n-1}) = 0$, then $P_{2n-1}(0) = 0$, and similarly, if $\rho(x_n) = 0$, then $P_{2n}(0) = 0$. As shown in figure 2 and table 2, one has: for Case (a): $\rho(x) = 0$ for $x = x_0$, x_4 and x_8, and for Case (b): $\rho_{\text{cl}}(x) = 0$ for $x = x_0$ and x_4. Thus, $P_1(0) = P_8(0) = P_9(0) = P_{16}(0) = 0$ for Case (a), whereas $P_1(0) = P_8(0) = 0$ for Case (b). On the other hand, as q increases some weights are gradually eliminated, while some other weights still survive. For $q \geq q_\infty$, where q_∞ has a large value, in eq. (6) $[\rho(x_n)]^{q-1} - \left[\rho(x_{n-1})\right]^{q-1} \to 0$ and, in this case, such equations are no longer valid.

Figure 4 shows the Rényi entropy $S_{\text{R}}(q, 1)$ as function of q, calculated for Case (a) of table 2, with the parameters of table 1. $S_{\text{R}}(q, 1)$ decreases monotonously when q

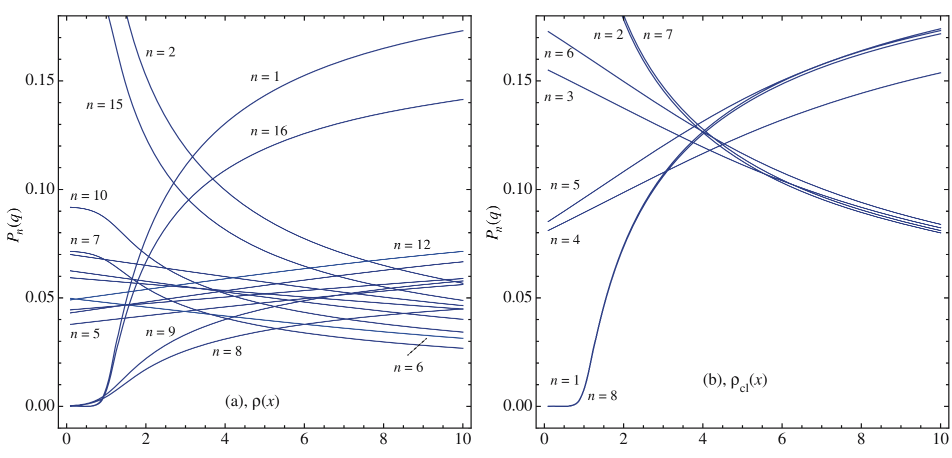

Figure 3. Behaviour of the weights $P_n(q)$ as a function of q, evaluated for the parameters of table 1 and Cases (a) and (b) of table 2.

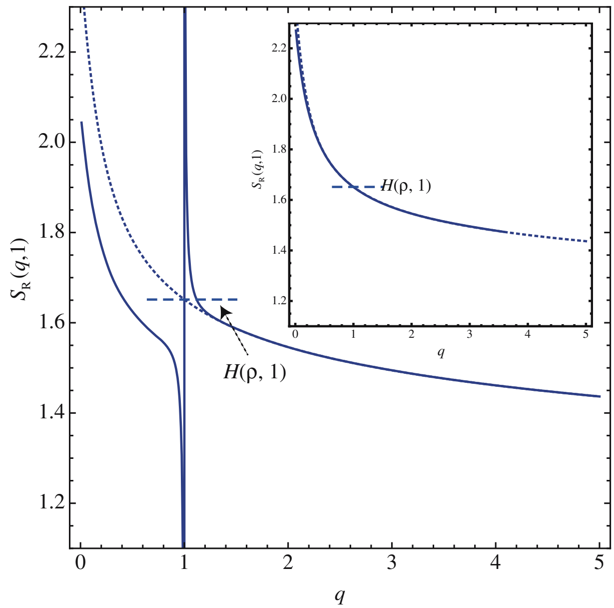

Figure 4. Rényi entropy $S_R(q, 1)$ for Case (a) in table 2. The solid line represents the natural discretization given by eq. (12), whereas the tiny dotted line exhibits the results from eq. (11). For improving the behaviour in the neighbourhood of $q = 1$ an artificial discretization is required, and the results are shown in the inset. The differential entropy takes the value $H(\rho, 1) \approx 1.6514$ nats. For Case (b) in table 2 a similar general behaviour is observed and $H(\rho_{cl}, 1) \approx 1.7001$ nats.

increases, as $S'_R(q, 1)$ is a negative quantity for all the values of q. However, in the vicinity of $q = 1$ and specially for $q < 1$, eq. (12) gives results which differ significantly from those given by eq. (11). Then, for dealing with the singularity of eq. (12) in the vicinity of $q = 1$, one uses the relation $S_R(1, h) = H(\rho, 1)$, chooses a cut-off point $q_\star$ (e.g., $q_\star = 1.2$) and, for $q \leq q_\star$, introduces an artificial discretization of the segments of $f(x, q-1)$: the nth segment is divided into M_n subintervals, each of length $\Delta_n := (x_n - x_{n-1})/M_n$, and $x_{n,m} = x_{n-1} + m\Delta_n$, for $m = 0, 1, \ldots, M_n$. The partition $M = \{3, 2, 2, 3, 3, 2, 2, 3\}$ significantly improves the results, which can be observed in the inset of figure 4, for Case (a). For Case (b) of table 2 one proceeds similarly by choosing a partition $M = \{4, 4, 4, 4\}$.

The entropies are shown in figure 5 as a function of q, for Cases (a) and (b) of table 2. Rényi entropy $S_R(q, 1)$ is calculated by using eq. (12), with the natural discretization for $q > q_\star$ and artificial discretization for $q \leq q_\star = 1.2$. One observes that, in general, the Rényi entropy for Case (b) is greater than that for Case (a), i.e., the interference term $\mathcal{I}_{12}$ in eq. (17) reduces the Rényi entropy in comparison with that for the classical probability density $\rho_{cl}(x)$ given by eq. (22): $S_R^{(b)}(q, 1) > S_R^{(a)}(q, 1)$. Rényi has shown that $S_R(q, 1)$ represents the disclosed information (or removed ignorance) after observing the result of an experiment depending on chance [1,16,18]. Also, it is observed that $H(\rho_{cl}, 1) > H(\rho, 1)$.

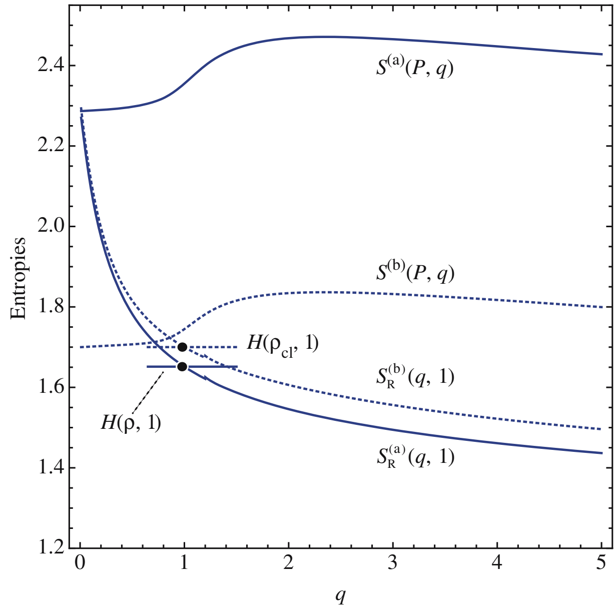

Figure 5. Entropies for the PDFs $\rho(x)$ (——) and $\rho_{\mathrm{cl}}(x)$ $(\cdot\cdot\cdot\cdot\cdot)$ given by eqs (17) and (22), respectively, calculated with the parameters shown in table 1. The differential entropy takes the value $H(\rho, 1) \approx 1.6514$ nats for Case (a) and $H(\rho_{\mathrm{cl}}, 1) \approx 1.7001$ nats for Case (b).

By using eq. (9) one obtains the Shannon entropy $S(P, q) := S\big(P|\rho, \rho^{q-1}\big)$, i.e., the averaged information associated with the PDF $\rho(x)$ and, therefore with the weights $P = \{P_{2n-1}, P_{2n}|n = 1, 2, \ldots, N\}$ deduced for the natural discretization of $\rho(x)$. Figure 5 shows that $S^{(a)}(P, q) > S^{(b)}(P, q)$, which indicates that the 'amount of uncertainty' [19] represented by the quantum probability density function $\rho(x)$ is greater than that represented by $\rho_{\mathrm{cl}}(x)$. This characteristic can be understood because the interference term $\mathcal{I}_{12}$ in eq. (17) produces four additional bending points in $\rho(x)$ compared to the four of $\rho_{\mathrm{cl}}(x)$ (figure 2 and table 2).

5. Additional comments and conclusions

As mentioned in §2.2, the segments of $f(x)$ can be subdivided into subintervals to have an artificial discretization of $\rho(x)$. Explicitly, the nth segment of $f(x)$, $x_{n-1} \le x \le x_n$, is divided into M_n bins, i.e., subintervals of widths defined by an arbitrary set of $(M_n + 1)$ points $X_n := \{x_{n,0}, x_{n,1}, x_{n,2}, \ldots, x_{n,M_n}\}$ such that $x_{n,0} := x_{n-1}$, $x_{n,M_n} := x_n$, and $x_{n,m-1} < x_{n,m}$, for $m = 0, 1, 2, \ldots, M_n$ with $M_n \ge 1$. A common way is to choose M_n subintervals each of length $\Delta_n := (x_n - x_{n-1})/M_n$, so that $x_{n,m} = x_{n-1} + m\Delta_n$. The mth subinterval of the nth segment is named the (n, m)th bin, i.e., $x_{n,m-1} \le x \le x_{n,m}$.

As the choice of the set X_n is arbitrary, this process is an artificial discretization of the PDF $\rho(x)$, which contrasts with the natural discretization described in §2.2. This procedure allows the application of the SMVT to each bin for improving the results of numerical calculations, as shown in §4.2 when the Rényi entropy $S_{\mathrm{R}}(q, 1)$ was calculated for $q < q_\star$.

By using an artificial discretization procedure, instead of eq. (4) one has

$$\langle f(x)\rangle = \sum_{n=1}^{N} \underbrace{\sum_{m=1}^{M_n} \left[f\big(x_{n,m-1}\big)\mathcal{P}^{\leftarrow}_{n,m} + f\big(x_{n,m}\big)\,\mathcal{P}^{\rightarrow}_{n,m}\right]}_{=I_n}. \tag{23}$$

The SMVT ensures the existence of a point $\xi_{n,m}$ which decomposes the (n, m)th bin into two subintervals, one at the left ($\leftarrow$) and other at the right ($\rightarrow$) of $\xi_{n,m}$. Thus, in analogy with eqs (6), the (n, m)th bin contributes with two weights,

$$\begin{aligned} \mathcal{P}^{\leftarrow}_{n,m} &:= \mathcal{P}\big([x_{n,m-1}, \xi_{n,m}]\big) = +\frac{f\big(x_{n,m}\big)J_{n,m} - I_{n,m}}{f\big(x_{n,m}\big) - f\big(x_{n,m-1}\big)} \geq 0, \\ \mathcal{P}^{\rightarrow}_{n,m} &:= \mathcal{P}\big([\xi_{n,m}, x_{n,m}]\big) = -\frac{f\big(x_{n,m-1}\big)J_{n,m} - I_{n,m}}{f\big(x_{n,m}\big) - f\big(x_{n,m-1}\big)} \geq 0, \end{aligned} \tag{24}$$

with the auxiliary quantities

$$I_{n,m} := \int_{x_{n,m-1}}^{x_{n,m}} f(x)\rho(x)\mathrm{d}x, \quad J_{n,m} := \mathcal{P}\big(x_{n,m-1} \leq x \leq x_{n,m}\big). \tag{25}$$

The weights assigned to the nth segment of $f(x)$ satisfy $J_n := \mathcal{P}([x_{n-1}, x_n]) = \sum_{m=1}^{M_n} \left[\mathcal{P}^{\leftarrow}_{n,m} + \mathcal{P}^{\rightarrow}_{n,m}\right]$ with the normalization condition $\sum_{n=1}^{N} \mathcal{P}([x_{n-1}, x_n]) = \omega_{\mathcal{N}}$.

To conclude, in this paper a discretization method of PDFs has been presented such that: (i) it is based on a well-defined mathematical property, namely, the SMVT; (ii) it allows one to apply the discrete probability theory to describe continuous probability distributions and (iii) it has been successfully applied to the discretization of the continuous Rényi entropy for a $\rho(x)$ associated with the superposition of two coherent states. This work also gives an answer to a remark in [11] concerning the third mean value theorem for integrals (in this study SMVT), which quotes as: "To be quite frank, I haven't the slightest idea what is good for" the SMVT. This paper is indeed a useful application of this theorem.

Acknowledgements

Mathematica was used for the calculations and graphics [20]. The author thanks the anonymous reviewer for important remarks and valuable comments. Thanks are also due to Professor Frank G Stiles for a revision of the text.

References

[1] A Rényi, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, California, 1961), Volume 1: Contributions to the Theory of Statistics, pp. 547–561

[2] A Rényi, *Rev. Int. Stat. Inst.* **33(1)**, 1 (1965)

[3] C Tsallis, *J. Stat. Phys.* **52**, 479 (1988)

[4] C E Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948)
[5] E T Jaynes, in: *Information theory and statistical mechanics* edited by K Ford, Statistical Physics, Vol. 3 (Benjamin, 1962), Brandeis Summer Institute in Theoretical Physics, 1963, ch. 3, pp. 182–218)
[6] J M Nichols, J V Michalowicz and F Bucholtz, *Handbook of differential entropy* (Taylor and Francis Group, LLC, 2014)
[7] E T Jaynes, *IEEE Trans. Syst. Sci. Cyber.* **4**, **3**, 227 (1968), Reprinted in E T Jaynes, *Papers on probability, statistics and statistical physics* edited by R D Rosencrantz (D. Reidel Publishing Company, Holland) pp. 114–130
[8] R K Hammond and J E Bickel, *Decision Analysis* **10(1)**, 6 (2013)
[9] L S Gradshteyn and I M Ryzhik, *Table of integrals, series, and products* (Academic Press, San Diego, 2000)
[10] T M Apostol, *Calculus* (Blaisdell Publishing Company, New York, 1964) Vol. 1
[11] M Spivak, *Calculus* (W A Benjamin Inc., Menlo Park, California, 1967)
[12] E L Stark, *Am. Math. Mon.* **85**, 481 (1978)
[13] E Hetmaniok, R Witula and D Slota, *Comp. Math. Appl.* **64**, 1612 (2012)
[14] T M Cover and J A Thomas, *Elements of information theory* (Wiley-Interscience, 1991)
[15] C G Chakrabarti and I Chakrabarty, *Int. J. Math. Math. Sci.* **17**, 2847 (2005)
[16] D Xu and D Erdogmuns, *Information theoretic learning* (Springer, 2010), chap. Renyi's Entropy, Divergence and Their Nonparametric Estimators, pp. 47–102, Information Science and Statistics
[17] I Bialynicki-Birula and J Mycielski, *Commun. Math. Phys.* **44(2)**, 129 (1975)
[18] P Jizba, *Decoherence and entropy in complex systems* (Springer, Berlin, 2004) vol. 633 of *Lecture Notes in Physics*, chap. Information Theory and Generalized Statistics, pp. 362–376
[19] E T Jaynes, *Phys. Rev.* **106**, 620 (1957)
[20] *Mathematica*, version 9.0 (Wolfram Research, Inc., Champaign, IL, 2013)