*mathematics*

*Article*

# Polynomial Distributions and Transformations

**Yue Yu and Pavel Loskot ***

ZJU-UIUC Institute, Haining 314400, China
* Correspondence: pavelloskot@intl.zju.edu.cn; Tel.: +86-571-8757-2579

**Abstract:** Polynomials are common algebraic structures, which are often used to approximate functions, such as probability distributions. This paper proposes to directly define polynomial distributions in order to describe stochastic properties of systems rather than to assume polynomials for only approximating known or empirically estimated distributions. Polynomial distributions offer great modeling flexibility and mathematical tractability. However, unlike canonical distributions, polynomial functions may have non-negative values in the intervals of support for some parameter values; their parameter numbers are usually much larger than for canonical distributions, and the interval of support must be finite. Hence, polynomial distributions are defined here assuming three forms of a polynomial function. Transformations and approximations of distributions and histograms by polynomial distributions are also considered. The key properties of the polynomial distributions are derived in closed form. A piecewise polynomial distribution construction is devised to ensure that it is non-negative over the support interval. A goodness-of-fit measure is proposed to determine the best order of the approximating polynomial. Numerical examples include the estimation of parameters of the polynomial distributions and generating polynomially distributed samples.

**Keywords:** approximation; distribution; histogram; least-squares; polynomial; probability density

**MSC:** 60E05

check for updates

**Citation:** Yu, Y.; Loskot, P. Polynomial Distributions and Transformations. *Mathematics* **2023**, *11*, 985. https://doi.org/10.3390/math11040985

Academic Editor: Vasile Preda

Received: 9 December 2022
Revised: 8 February 2023
Accepted: 13 February 2023
Published: 15 February 2023



## 1. Introduction

Approximating functions is motivated by reducing the computational complexity and achieving the analytical tractability of mathematical models. This also includes the problems of finding low-complexity and low-dimensional mathematical models for continuous or discrete-time observations, such as time-series data, and empirically determined features, such as histograms. This paper is concerned with the latter problem, i.e., how to effectively model the probability distributions of observation data. In particular, it is proposed to define polynomial probability distributions rather than to assume polynomial approximations of probability distributions. This is a major departure from the reasoning found in the existing literature.

Polynomial distributions provide superior flexibility over other canonical distributions, albeit at a cost of a larger number of parameters, and the support interval is constrained to a finite range of values. The main advantages of polynomial distributions are that they can yield parameterized closed-form expressions and enable the modeling of complex multimodal and time-evolving probability distributions. These distributions are encountered, for example, when describing causal interactions and state transitions in dynamic systems. This may lead to the development of novel probabilistic mathematical frameworks. The disadvantage is that, in the case of a general polynomial function, it may be difficult to ensure that the polynomial is non-negative over the whole intended interval of support. The non-negativity can be guaranteed, for example, by assuming squared polynomials.

The Weierstrass theorem [1] is the fundamental result in the approximation theory of functions. It states that every continuous function can be uniformly approximated with an arbitrary precision over any finite interval by a polynomial of a sufficient order. The uniform approximation can be expressed as a sequence of algebraic polynomials uniformly converging over a given interval to the function of interest. The approximation accuracy can be evaluated by different metrics including $l_p$-norms, minimax norm, and others. The best approximating function from a set or a sequence of functions and its properties can be determined by the Jackson theorem. The Stone–Weierstrass theorem generalizes the function approximation to cases of multivariate functions and functions in multiple dimensions [2].

Runge's phenomenon arises when the approximating polynomials contain a set of predefined points, which can prevent uniform convergence from being possible [3]. The equidistant approximating points can be optimized using the Lebesgue constant as a measure of the approximation accuracy [4,5].

Polynomials can be used to approximate known probability distributions as well as distributions estimated as histograms [6,7]. Reference [8] is one of the earlier works that assume the approximation of probability distributions by a polynomial. The fitting of multivariate polynomials to multivariate cumulative distributions and their partial derivatives is studied in [9], whereas multivariate polynomial interpolation is studied in [10]. The conditions for the coefficients of a polynomial to be a sum of two squared polynomials are determined in [11].

The problem of fitting a polynomial into a finite number of data samples has been investigated in the classic reference [12]. The polynomial curve fitting methods are often available in various software packages [13]. Modeling times series data by piecewise polynomials is considered in [14,15]. The least-squares polynomial approximation of random data samples with standard and induced densities is compared in [16]. A new method for polynomial interpolation of data points within a plane is proposed in [17]. Interestingly, the recent survey [18] on approximating probability distributions does not mention polynomial approximation as one of the available methods.

The polynomial expansion of chaos for the reliability analysis of systems is proposed in [19]. A polynomial kernel for feature learning from data is considered in [20]. The Stone–Weierstrass theorem is assumed in [21] to design a neural network that can approximate an arbitrary measurable function. The method for function approximation by a polynomial using a neural network is investigated in [22].

Polynomials can be sparse, i.e., only some of their coefficients—including the coefficient determining the order—are non-zero. The polynomials with special properties are named; for example, there are Lagrange, Legendre, Diskson, Chebyshev, and Bernstein polynomials [23,24]. Special polynomials, such as Hermite and Lagrangian polynomials, can form the basis for function decomposition. There is a close link between approximating periodic continuous functions and trigonometric polynomials in the Fourier analysis [25]. A procedure for orthogonal polynomial decomposition of multivariate distributions was devised in [26] in order to compute the output of a multidimensional system with the stochastic input.

Reference [23] is a comprehensive textbook on the theory of polynomials covering fundamental theorems, special polynomials, polynomial algebra, finding and approximating polynomial roots, finding polynomial factors, solving polynomial equations, and defining polynomial inequalities and properties of polynomial approximations. The other textbook [24] includes additional topics, such as critical points of polynomials, the compositions of polynomial functions, theorems and conjectures about polynomials, and defining extremal properties of polynomials. Although the textbook [27] focuses on solving differential equations by polynomial approximations, it also provides a necessary background on polynomials including their definitions and properties. Differential equations are solved by Jacobi polynomial approximation in [28].

The properties of minima and maxima of polynomials were studied in [29]. An algorithm for finding the global minimum of a general multivariate polynomial was developed in [30]. The number of local minima of a multivariate polynomial is bounded in [31]. Sturm series are assumed in [32] to find the maxima of a polynomial.

In this paper, polynomial distributions are introduced in Section 2 including transformations of the polynomial distributions, fitting a histogram with a polynomial distribution, constructing a piecewise polynomial distribution, and defining the basic properties of a random polynomial function. In Section 3, selected properties of the polynomial distributions are derived. The estimation problems involving polynomial distributions and determining the polynomial order are considered in Section 4. Numerical examples are presented in Section 5 including constructing a piecewise polynomial, generating polynomially distributed random samples, estimating parameters of polynomial distributions by the method of moments and by fitting the observations, and approximating distributions by Lagrange interpolation. Section 5 ends with a summary of key findings. The paper is concluded in Section 6. In addition, the key expressions for polynomial functions in Form I, II, and III are summarized in Appendices A–C, respectively.

The following notations are used in the paper: $X$ denotes a random variable whereas $x$ denotes a specific value of this random variable; $(\cdot)^T$ is matrix transpose; $(\cdot)^{-1}$ is matrix inverse; the operators, $\mathrm{E}[\cdot]$ and $\mathrm{var}[\cdot]$, denote expectation and variance, respectively; $(\cdot)!$ denotes factorial, $\stackrel{!}{=}$ is used to find a value satisfying the indicated equality, and $\langle\cdot,\cdot\rangle$ denotes the dot-product of two vectors.

## 2. Defining Polynomial Distributions

Given a continuous and finite interval, $(l,u)$, $-\infty < l < u < +\infty$, the probability density function (PDF), $p(x)$, of a random variable, $X$, with the support, $(l,u)$, must satisfy the following two conditions,

$$\begin{aligned} p(x) &\geq 0, \quad \forall x \in (l,u) \\ \int_l^u p(x)\,\mathrm{d}x &= 1. \end{aligned} \tag{1}$$

Assume that the PDF, $p(x)$, can be linearly expanded as

$$p(x) = a_0 + \sum_{i=1}^{n} a_i b_i(x) \tag{2}$$

into the $n$-dimensional basis of generally non-linear functions, $b_i(x)$. These functions can also be parameterized as $b_i(x) \equiv b(x;\theta_i)$. Provided that the functions, $b_i(x)$, are themselves PDFs, i.e., they satisfy the conditions (1), then the PDF (2) is referred to as mixture distribution, and $\sum_{i=0}^{n} a_i = 1$.

In this paper, the functions, $b_i(x) = x^i$, are assumed, so that, the expression (2) represents an ordinary univariate polynomial of degree, $n$. The coefficients, $a_i$, can be a function of another common variable, e.g., $a_i(y)$, $i = 0,1,\ldots,n$; this multivariate polynomial is referred to as algebraic function. The multivariate polynomial having the same degree of a non-zero term is referred to as being homogeneous (formerly a quantic polynomial).

The following three representations of real-valued polynomial functions are considered in this paper.

**Definition 1.**

$$\text{Form I}: \qquad p_n(x) = \sum_{i=0}^{n} a_i\, x^i, \qquad a_i \in \mathcal{R},\ a_n \neq 0 \tag{3a}$$

$$\text{Form II}: \quad p_n(x) = a_n \prod_{i=1}^{n} (x - r_i), \quad r_i \in,\ a_n \neq 0 \tag{3b}$$

$$\text{Form III}: \qquad p_n(x) = \sum_{i=1}^{n} \frac{a_i}{x - r_i}, \qquad a_i \neq 0,\ r_i \neq r_j\ \forall i \neq j \tag{3c}$$

Form I is a canonical polynomial function. Form II indicates that every $n$-degree polynomial has exactly $n$, generally complex-valued, roots $r_i$ [33]. The number of real-valued roots can be determined by Sturm's theorem. Form III is a rational polynomial function. The basic properties of the polynomial Forms I, II, and III are summarized in Appendices A–C, respectively, including roots, indefinite and definite integrals, derivatives, general statistical moments, and characteristic or moment-generating functions. Note that every polynomial function, $p_n(x)$, of any order, $n$, diverges when its argument, $x$, becomes unbounded. In addition, Forms I and II are equivalent as shown in Appendix B, and, for complex-conjugate roots, $(x - r_i)(x - r_i^*) = (x - \mathrm{Re}(r_i))^2 + \mathrm{Im}(r_i)^2 > 0$. Form I defined by (3a) can also be computed recursively as

$$\begin{aligned} p_n(x) &= (\dots(((a_n x + a_{n-1})x + a_{n-2})x + a_{n-3})\dots)x + a_0 \\ &= x p_{n-1}(x) + a. \end{aligned} \tag{4}$$

A Form I or II polynomial, $p_n(x)$, of degree $n$ and all of its derivatives, $p_n^{(k)}(x) = \frac{\mathrm{d}^k}{\mathrm{d}x^k} p_n(x)$, $k \leq n$, is continuous and strictly bounded over a finite interval, $x \in (l, u)$. However, the polynomial forms in Definition 1 represent a PDF, *if and only if*, they satisfy both conditions (1). This can be achieved by using linear and non-linear transformations, which are defined in the following lemma.

**Lemma 1.** *A polynomial, $p_n(x)$, can become a PDF by using either of the following transformations.*

(a) *There exist finite real constants, $A$ and $B$, such that the linearly transformed polynomial, $Ap_n(x) + B$, satisfies PDF conditions (1).*
(b) *There exists a real positive constant, $A > 0$, such that the polynomial, $A|p_n(x)|$, or, $A(p_n(x))_+$, satisfies PDF conditions (1) where $|\cdot|$ denotes absolute value, and $(\cdot)_+$ changes the negative values of its argument to zero.*
(c) *There exists a low-degree polynomial, $q_k(x)$, such that the polynomial, $q_k(p_n(x))$, satisfies PDF conditions (1); for instance, $q_1(x) = Ax + B$ [cf. (a)], or, $q_2(x) = Ax^2$, $A > 0$.*

**Proof.**

(a) Let, $b = \min_{x \in (l,u)} p_n(x)$, so that, $p_n(x) - b \geq 0$. Then, $A^{-1} = \int_l^u p_n(x) - b \,\mathrm{d}x$, and, $B = -Ab$.
(b) For any, $x$, the functions, $|p_n(x)| \geq 0$, and, $(p_n(x))_+ \geq 0$. Then, $A^{-1} = \int_l^u |p_n(x)| \,\mathrm{d}x$, or, $A^{-1} = \int_l^u (p_n(x))_+ \,\mathrm{d}x$, respectively.
(c) The linear transformation, $q_1(x)$, was considered in (a). The transformation, $p_n^2(x) \geq 0$, and, $A^{-1} = \int_l^u p_n^2(x) \,\mathrm{d}x$.

□

The polynomial PDFs defined in Lemma 1 can be further constrained by the required number of local minima, maxima, and roots within the interval of support, $(l, u)$. There can also be additional constraints on smoothness expressed in terms of the minimum required polynomial order.

By Bolzano's theorem, a continuous function having opposite sign values in an interval also has a root between these values. Consequently, a polynomial, $p_n(x)$, of order $n$ have at least one maximum or minimum between every two adjacent roots, and there can be a maximum or minimum located at the roots themselves [29]. Moreover, provided that the polynomial is considered over a finite interval, the boundary points of the support interval should be treated as additional roots, i.e., the boundary points can create local maximum or minimum as well as allow additional extrema to exist before the first nearest root. In the case of Form II polynomials, the condition of the first derivative to be zero can be equivalently expressed as

$$\frac{\mathrm{d}}{\mathrm{d}x} p_n(x) \stackrel{!}{=} 0 \quad \Leftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}x} \log p_n(x) = \frac{\dot{p}_n(x)}{p_n(x)} = \sum_{i=1}^{n} \frac{1}{x - r_i} \stackrel{!}{=} 0. \tag{5}$$

However, this approach still requires finding the roots of (5) for every sub-interval, $(r_i, r_{i+1})$, $i = 0, 1, \ldots, n$, where $r_0 \equiv l$ and $r_{n+1} \equiv u$. It may be much easier to find the local extrema by considering the recursion,

$$p_n(x) = \int p_{n-1}(x)\,\mathrm{d}x = \sum_{i=0}^{n-1} \frac{a_i}{i+1} x^{i+1} + c \tag{6}$$

provided that the roots of the polynomial, $p_{n-1}(x) = a_{n-1} \prod_{i=1}^{n-1}(x - r_i)$, are known, and, $c = a_0$, denotes the constant of integration. These roots can be known by design, i.e., the locations of minima and maxima are selected a priori in a given interval of support. More importantly, in the case of a polynomial PDF, the local maxima represent the modes of such a distribution.

The Form I polynomial PDF can be generalized as

$$p_n(x) = \sum_{i=0}^{n} a_i g^i(x), \quad x \in (l, u) \tag{7}$$

where $g(x)$ is a mathematical expression (i.e., not a transformation). For instance, it is possible to assume polynomials with fractional rather than integer powers of the independent variable [34].

For $g(x) = \mathrm{e}^{\mathrm{j}\omega_0 x}$, $\mathrm{j} = \sqrt{-1}$, $\omega_0 = 2\pi/(u-l) > 0$, the PDF (7) becomes the truncated exponential Fourier series, i.e.,

$$p_n(x) = \sum_{i=0}^{n} a_i\, \mathrm{e}^{\mathrm{j}\omega_0 i x}, \quad a_i = \frac{1}{u-l} \int_l^u p_n(x)\, \mathrm{e}^{-\mathrm{j}\omega_0 i x}\,\mathrm{d}x. \tag{8}$$

The corresponding $k$-th general moments are then computed as

$$\int_l^u x^k\, p_n(x)\,\mathrm{d}x = \sum_{i=0}^{n} a_i \int_l^u x^k\, \mathrm{e}^{\mathrm{j}\omega_0 i x}\,\mathrm{d}x = \sum_{i=0}^{n} a_i (-1)^k W^{(k)}(\mathrm{j}\omega_0 i) \tag{9}$$

where $W(\mathrm{j}\omega) = \int_l^u \mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}x$ is the Fourier transform of a rectangular window located over the interval, $(l, u)$.

For $g(x) = \mathrm{e}^x$, PDF (7) becomes,

$$p_n(x) = \sum_{i=0}^{n} a_i\, \mathrm{e}^{ix}, \quad x \in (l, u) \tag{10}$$

which can be readily integrated, although general statistical moments can only be expressed as special functions.

Consider now the general case of PDF (7), for $n = 2$. Thus, given $p_2(x)$ and $g(x)$, and positive integers $i_1$ and $i_2$, the task is to find the coefficients $a_0$, $a_1$ and $a_2$ of the PDF decomposition,

$$p_2(x) = a_2 g^{i_2}(x) + a_1 g^{i_1}(x) + a_0, \quad x \in (l, u). \tag{11}$$

Multiplying both sides of (11) by $g^{-i}(x)\dot{g}(x)$ and integrating, we obtain,

$$\int_l^u p_2(x) g^{-i}(x)\dot{g}(x)\,\mathrm{d}x = a_2 \int_l^u g^{i_2}(x) g^{-i}(x)\dot{g}(x)\,\mathrm{d}x + a_1 \int_l^u g^{i_1}(x) g^{-i}(x)\dot{g}(x)\,\mathrm{d}x + a_0 \int_l^{g(u)} g^{-i}(x)\dot{g}(x)\,\mathrm{d}x. \tag{12}$$

Assuming a substitution, $y = g(x)$, Equation (12) can be rewritten as

$$\int_{g(l)}^{g(u)} p_2(g^{-1}(y)) y^{-i}\,\mathrm{d}y = a_2 \int_{g(l)}^{g(u)} y^{i_2 - i}\,\mathrm{d}y + a_1 \int_{g(l)}^{g(u)} y^{i_1 - i}\,\mathrm{d}y + a_0 \int_{g(l)}^{g(u)} y^{-i}\,\mathrm{d}y. \tag{13}$$

Provided that, $g(u) = -g(l) = v$, and $i_1 > 0$ is an odd-integer, and $i_2 > 0$ is an even-integer, then, for $i = i_1$ and $i = i_2$, respectively,

$$\begin{aligned} \int_{-v}^{v} p_2(g^{-1}(y))y^{-i_1}\,\mathrm{d}y &= a_2 \underbrace{\int_{-v}^{v} y^{i_2-i_1}\,\mathrm{d}y}_{=0} + 2a_1 v + a_0 \underbrace{\int_{-v}^{v} y^{-i_1}\,\mathrm{d}y}_{=0} \\ \int_{-v}^{v} p_2(g^{-1}(y))y^{-i_2}\,\mathrm{d}y &= 2a_2 v + a_1 \underbrace{\int_{-v}^{v} y^{i_1-i_2}\,\mathrm{d}y}_{=0} + a_0 \int_{-v}^{v} y^{-i_2}\,\mathrm{d}y \end{aligned} \tag{14}$$

and, therefore,

$$\begin{aligned} a_1 &= \frac{1}{2v} \int_{-v}^{v} p_2(g^{-1}(y))y^{-i_1}\,\mathrm{d}y \\ a_2 &= \frac{1}{2v} \int_{-v}^{v} \Big(p_2(g^{-1}(y)) - a_0\Big)y^{-i_2}\,\mathrm{d}y. \end{aligned} \tag{15}$$

The offset, $a_0$, must be computed from some other constraint, for example, as the minimum value to guarantee a non-negativity of $p_2(x)$. Note that the function, $g(x)$, in (11) must be chosen, so the integrals (15) converge.

### 2.1. Probability Density Transformations

In general, if $g(x)$ is an invertible transformation of a random variable, $X$, having the PDF, $p(x)$, the PDF, $q(x)$, of random output variable, $g(x)$, is, [35]

$$q(x) = \frac{p(g^{-1}(x))}{|\dot{g}(g^{-1}(x))}) = p(g^{-1}(x))\Big|\dot{g}^{-1}(x)\Big| \tag{16}$$

where $\dot{g}(x) = \frac{\mathrm{d}}{\mathrm{d}x}g(x)$. Assuming $p(x)$ is a Form I polynomial PDF, $p_n(x)$, the transformed PDF is also a Form I polynomial, i.e.,

$$q_n(x) = |\dot{g}^{-1}(x)| \sum_{i=1}^{n} a_i g^{-i}(x) = \sum_{i=1}^{n} a_i \left(\frac{\sqrt[i]{|\dot{g}^{-1}(x)|}}{g(x)}\right)^i \tag{17}$$

in the variable, $\sqrt[i]{|\dot{g}^{-1}(x)|}g^{-1}(x)$.

Assuming a linear transformation, $g_1(x) = b_1 x + b_0$, the PDF (17) is also a polynomial PDF of the same order, i.e.,

$$q_n(x) = \sum_{i=1}^{n} \frac{a_i}{|b_1|b_1^i}(x - b_0)^i. \tag{18}$$

However, the linear transformation changes the support, $(l, u)$ of $p_n(x)$, to $(b_1 l + b_0, b_1 u + b_0)$, if $b_1 > 0$, and $(b_1 u + b_0, b_1 l + b_0)$, if $b_1 < 0$.

Another example of a non-linear transformation with memory that preserves polynomial form of the resulting distribution is an integrator. In particular, let, $g^{-1}(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \equiv F(x)$, i.e., $g(x) = F^{-1}(x)$, such that, $f(u) \geq 0$, for $\forall u$. Then, substituting into (16), the transformed PDF can be written as

$$q_m(x) = b_1 f(x) p_n(b_1 F(x) + b_0) \tag{19}$$

where $b_1 \neq 0$ and $b_0$ are arbitrary real constants. Provided that $f(x)$ is a polynomial of order, $k$, $F(x)$ is a polynomial of order, $(k+1)$ (cf. Appendix A) and, thus, $m = n(k+1)k$. The family of PDFs with a form similar to (19) are considered in [36], which could be investigated also for our case of the polynomial distributions.

Consider now a general case of a polynomial nonlinear transformation, $g_k(x)$, and denote as $x_i$, $i = 1, 2, \ldots, N(y)$, all the roots of, $g_k(x) = y$. Then, the PDF (16) is rewritten as

$$q_m(y) = \sum_{j=1}^{N(y)} \frac{p_n(x_j(y))}{|\dot{g}_k(x_j(y))|} \tag{20}$$

i.e., it is a sum of ratios of polynomials, i.e., $q_m(y)$ is a polynomial of a certain order, $m$.

Linear and non-linear transformations of a random variable can be used to change the interval of support of its probability distribution. The following Lemma 2 assumes linear transformations to convert the interval of support, $(l, u)$, into $(-1, +1)$ and vice versa. Lemma 3 proposes two transformations on how to convert the interval of support, $(-1, +1)$, to semi-finite or infinite intervals of support, respectively.

**Lemma 2.** *The interval of support, $(l, u)$, of a PDF, $p(x)$, is changed to the interval, $(-1, +1)$, by a linear transformation, $\frac{2}{u-l}x - \frac{u+l}{u-l}$, which transforms the PDF, $p(x)$, to the PDF, $\frac{u-l}{2}p\left(\frac{(u-l)x-(u+l)}{2}\right)$. Furthermore, the linear transformation, $\frac{(u-l)}{2}x + \frac{u+l}{2}$, transforms the PDF, $p(x)$, with support, $(-1, +1)$, into the PDF, $\frac{2}{u-l}p\left(\frac{2x-(u+l)}{u-l}\right)$, with the interval of support, $(l, u)$.*

**Proof.** The linear transformation, $Ax + B$, transforms any PDF, $p(x)$, into the PDF, $|A|^{-1}p\left(\frac{x-B}{A}\right)$, [35]. □

**Lemma 3.** *The PDF, $p(x)$, defined over the finite interval of support, $(-1, +1)$, can be transformed into the PDF, $(x^2 + x + 1/4)^{-1}p\left(\frac{2x-1}{2x+1}\right)$, with semi-infinite support, $(0, +\infty)$, using the non-linear transformation, $\frac{1}{2}\left(\frac{1+x}{1-x}\right)$. Similarly, the non-linear transformation, $\operatorname{atanh}(x)$, can be assumed to extend the support to all real numbers, for a PDF, $p(x)$, defined over the support interval, $(-1, +1)$. The transformed PDF becomes $\cosh^{-2}(x)p(\tanh(x))$.*

**Proof.** The functions, $g(x) = \frac{1}{2}\left(\frac{1+x}{1-x}\right)$, and, $g(x) = \operatorname{atanh}(x)$, are increasing, i.e., invertible in the interval, $(-1, +1)$. Then, in (20), $N(y) = 1$, the inverse transforms, $\frac{2x-1}{2x+1}$, and, $\tanh(x)$, and their derivatives, $(x^2 + x + 1/4)^{-1}$, and, $\cosh^{-2}(x)$, respectively. □

Furthermore, there are scenarios when deterministic values, $x \in (l, u)$, are transformed by a polynomial function, $p_n(x)$, having the random coefficients, $a_i$. The output value of such a transformation,

$$y = p_n(x) = \sum_{i=0}^{n} a_i x^i \tag{21}$$

is a random variable. Provided that $a_i$ are independent and distributed as $f_{a_i}(a_i)$, and conditioned on $x$, the random variable $Y$ is a sum of independent random variables, so its PDF is given by a multi-fold convolution,

$$f_{y|x}(y|x) = \left(\prod_{i=0}^{n} |x^{-i}|\right) f_{a_0}(y) \circledast f_{a_1}(y/x) \circledast \cdots \circledast f_{a_n}(y/x^n) \tag{22}$$

since, $a_i x^i$, are distributed as $|x^{-i}| f_{a_i}(y/x^i)$. Furthermore, the conditional mean and variance of $Y$, respectively, are,

$$\mathrm{E}[Y|X] = \sum_{i=0}^{n} \mathrm{E}[a_i]x^i, \quad \text{and,} \quad \mathrm{var}[Y|X] = \sum_{i=0}^{n} \mathrm{var}[a_i]x^{2i}. \tag{23}$$

Provided that $a_i$ have equal means, then for any $n$ and $x \in (-1, +1)$, $\mathrm{E}[Y|X] > 0$, if $\mathrm{E}[a_i] > 0$, and $\mathrm{E}[Y|X] < 0$, if $\mathrm{E}[a_i] < 0$. Note also that, if $a_i$ have equal variances, then the variance of $Y$ is minimized for $x = 0$, and, it is equal to, $\mathrm{var}[Y|X = 0] = \mathrm{var}[a_0]$.

The bounds for the number of real roots of random but sparse polynomials were provided in [37]. A numerical method for efficiently finding the zeros of complex-valued polynomials of very large orders was developed in [38]. Another method for a rapid root finding of polynomials is presented in [39].

## 2.2. Polynomial PDF Fit of a Histogram

Approximating a continuous function by a polynomial over a finite interval is formalized by the well-known Weierstrass theorem [1]. Runge's phenomenon occurs when the approximating function must contain predefined points [3]. The polynomial approximation represents the problems of existence as well as the uniqueness of such a polynomial, and also the problem of how to find it. These problems are crucially dependent on the choice of metric for the goodness of approximation.

Hence, consider the problem of approximating a PDF having the finite support by a polynomial PDF. For instance, the empirical histogram can be fitted by a polynomial function, or a known PDF can be approximated by a polynomial in order to achieve mathematical tractability. However, in neither of these cases, the resulting polynomial is guaranteed to satisfy conditions (1), since the polynomial coefficients are normally chosen to obtain the best fit.

The polynomial PDF can be obtained by assuming a polynomial function, which is non-negative over a given interval for any values of its coefficients. One example of such a polynomial is, $p_n^2(x)$, which has degree, $2n$. The true PDF, $q(x)$, can be then approximated as

$$q(x) \approx p_n^2(x), \quad \text{or,} \quad \sqrt{q(x)} \approx p_n(x). \tag{24}$$

The latter strategy by first transforming $q(x)$ with a square root is numerically more stable. Other invertible transformations of $q(x)$ can also be assumed, provided that they yield the non-negative polynomial, $p_n(x)$, since scaling $p_n(x)$ to have a unit area usually does not affect the approximation error significantly.

For instance, the data points, $(x_i, \sqrt{y_i})$, $l \le x_i \le u$, $i = 1, 2, \ldots, M$, can be interpolated by Lagrange polynomials [5],

$$L_i(x) = \prod_{\substack{j=1 \\ i \neq j}}^{M} \frac{x - x_i}{x_i - x_j}. \tag{25}$$

Then, the true PDF, $p(x)$, is approximated as

$$p(x) \approx \left( \sum_{i=1}^{M} \sqrt{y_i}\, L_i(x) \right)^2 \equiv q_{2(M-1)}(x) \tag{26}$$

which is a polynomial of order, $2(M-1)$. In order to normalize the approximation (26), let,

$$c_{ij} = \prod_{\substack{j_1=1 \\ j_1 \neq i}}^{M} \prod_{\substack{j_2=1 \\ j_2 \neq j}}^{M} (x_i - x_{j_1})(x_j - x_{j_2}) \tag{27}$$

and,

$$\begin{aligned} s_{ij} = \int_l^u L_i(x) L_j(x)\,\mathrm{d}x &= c_{ij}^{-1} \int_l^u \prod_{\substack{j_1=1 \\ j_1 \neq i}}^{M} \prod_{\substack{j_2=1 \\ j_2 \neq j}}^{M} (x - x_{j_1})(x - x_{j_2})\,\mathrm{d}x \\ &= c_{ij}^{-1} \sum_{k=0}^{2(M-1)} a_k \int_l^u x^k\,\mathrm{d}x = c_{ij}^{-1} \sum_{k=0}^{2(M-1)} \frac{a_k}{k+1} (u^{k+1} - l^{k+1}). \end{aligned} \tag{28}$$

Then, the area is calculated as

$$\int_l^u q_{2(M-1)}(x)\,\mathrm{d}x = \sum_{i=1}^{M}\sum_{j=1}^{M}\sqrt{y_i y_j}\,s_{ij}. \tag{29}$$

The Lagrange polynomials satisfy, $q_{2(M-1)}(x_i) = y_i$, so they are subject to Runge's phenomenon [3]. In particular, the approximation error can be estimated as [3],

$$p(x) - q_{2(M-1)}(x) = \frac{p^{(M+1)}(\tilde{x})}{(M+1)!}\prod_{i=1}^{M}(x - x_i) \tag{30}$$

where $l \le \tilde{x} \le u$. Provided that $|p^{(M+1)}(x)| < \infty$, for $\forall M$, and, $\forall x \in (l,u)$, $q_{2(M-1)}(x)$ converges uniformly to $p(x)$, i.e., $\sup_x |p(x) - q_{2(M-1)}(x)| = 0$, as $M \to \infty$. The PDF, $p(x)$, is often only known as a sequence of sample points, $(x_i, y_i)$, so the derivatives, $p^{(M+1)}(x)$, cannot be determined. However, the approximation can be improved by using Chebyshev or extended Chebyshev points instead of equally spaced points [4].

The most common method for fitting a polynomial to a histogram is a linear regression [40]. Denote the vectors, $\boldsymbol{y} = [y_i]$, $i = 1,2,\ldots,M$, and, $\boldsymbol{a} = [a_j]$, $j = 0,1,\ldots,n$, and the matrix, $\boldsymbol{X} = [x_i^j]$. The constrained least squares (LS) problem is then formulated as

$$\min_{\boldsymbol{a}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{a}\|^2, \quad \text{s.t.} \quad \boldsymbol{w}^T\boldsymbol{a} = 1 \tag{31}$$

where the weights, $w_i = \frac{1}{i+1}(u^{i+1} - l^{i+1})$, assuming the support interval, $(l,u)$. The first derivative of the corresponding Lagrangian is set equal to zero, and the estimated coefficients, $\hat{\boldsymbol{a}}$, of the fitting polynomial, $p_n(x)$, are computed as [41],

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{a}}\mathcal{L}(\lambda) = 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} - 2\boldsymbol{X}^T\boldsymbol{y} + \lambda\boldsymbol{w}^T \overset{!}{=} \boldsymbol{0} \\ &\Rightarrow \quad \hat{\boldsymbol{a}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\Big(\boldsymbol{X}^T\boldsymbol{y} + \frac{\lambda}{2}\boldsymbol{w}^T\Big). \end{aligned} \tag{32}$$

In order to approximate a known continuous distribution, $f(x)$, over a finite interval, $(l,u)$, representing the full or truncated support of that distribution, the constrained least-squares (32) can be again used assuming the distribution samples, $f(l + \Delta_x i)$, $i = 0,1,\ldots$.

If $p_n(x)$ is the best polynomial fit of a histogram, or of a sampled known PDF, then it must be evaluated whether it is non-negative over the whole support of interest, $(l,u)$, for example, due to Runge's phenomenon as discussed in Section 2.2. This can be readily and reliably tested by numerically computing the integral, $I_1 = \mathrm{Im}\Big(\int_l^u \sqrt{p_n(x)}\,\mathrm{d}x\Big)$, or, $I_2 = \int_l^u p_n(x) - |p_n(x)|\,\mathrm{d}x$. If $p_n(x)$ contains negative values within the interval, $(l,u)$, then $I_1 \neq 0$, and $I_2 < 0$, respectively. It is also possible to assume a logarithm instead of the square root in the definition of the integral, $I_1$.

In this case, the polynomial fitted to a histogram contains negative values, a constant, $d > 0$, can be added to the observed data points, i.e., $y_i \mapsto \frac{y_i + d/\Delta_x}{1+Md}$, where $\Delta_x = x_{i+1} - x_i$, and the scaling ensures that, $\Delta_x \sum_{i=1}^{M} y_i = 1$. Correspondingly, the fitted polynomial is also shifted and scaled as $p_n(x) \mapsto \frac{p_n(x)+d/\Delta_x}{1+dM}$, so $\int_l^u p_n(x)\,\mathrm{d}x = 1$.

Furthermore, the roots of a polynomial, $p_n(x)$, can be constrained in order to guarantee that it is non-negative over a finite interval, $(l,u)$. This is formulated in the following theorem.

**Theorem 1.** *A Form II real-valued polynomial, $p_n(x)$, of order $n$ with $a_n > 0$ and the roots, $r_1 \le r_2 \le \ldots \le r_n$, is non-negative over the interval, $(l,u)$, provided that all its roots satisfy at least one of the following conditions:*

*(a)* *a root has even multiplicity;*

*(b)* *a root has a complex conjugate pair;*

(c) *a (real-valued) root is smaller than $l$;*
(d) *a real-valued root has odd multiplicity and is larger than $u$; the number of such roots must be even.*

**Proof.** Form II polynomial is a product of linear functions, $(x - r_i)$. Cases (a), (b), and (c) are trivial. Case (d) is a combinatorial problem. The roots with odd multiplicity cannot be smaller than $u$. Even if these roots are all larger than $u$, then their number must be even in order for their negative parts to cancel out for all values smaller than $u$. □

**Corollary 1.** *A Form II real-valued polynomial, $p_n(x)$, of order $n$ with $a_n > 0$ and the roots, $r_1 \leq r_2 \leq \ldots \leq r_n$, has negative values in the interval, $(l, u)$, provided that there is an odd-number of real-valued roots with odd multiplicities that are greater than $l$, or, there is an even number of real-valued roots with odd multiplicities, and at least one root is located between $l$ and $u$.*

Theorem 1 can also be used for Form I polynomials, provided that they are converted to Form II as indicated in Appendix B. Even though the roots cannot be obtained analytically for polynomials of order $n > 4$ (Abel–-Ruffini's theorem), it may be sometimes possible to consider a product, $\prod_j p_{n_j}(x)$, of polynomials of orders, $n_j \leq 4$, for $\forall j$.

### 2.3. Piecewise Polynomial PDF

In some applications, a piecewise polynomial curve fitting can be assumed. In particular, the following construction is proposed to fit a set of $(M + 1)$ points, $(x_i, y_i)$, $i = 1, \ldots, (M + 1)$, $x_i < x_{i+1}$, and $y_i \geq 0$, representing local extrema of a histogram, or of a known PDF. The construction yields a piecewise polynomial PDF, $p_n(x)$, of the same order $n$, over the interval, $(l, u)$, with $l = x_1$ and $u = x_{M+1}$, such that, exactly, $p_n(x_i) = y_i$. The data points, $(x_i, y_i)$, are referred to as control points of the piecewise polynomial $p_n$. It should be noted that univariate piecewise polynomial functions are generally referred to as splines, and their control points as knots.

The following construction defines a piecewise polynomial PDF by sample points representing an alternating sequence of local maxima and minima. The points between the subsequent extrema are then interpolated by increasing or decreasing polynomial segments with a defined continuity order between these segments. As long as the minima are non-negative, all segments are non-negative. However, the resulting curve must be normalized, so that it has a unit area, i.e., it is a PDF.

**Definition 2.** *Let $p_n(x)$ be piecewise continuous, and composed of $M$ non-overlapping polynomial segments, $q_{(i)n}(x)$, i.e.,*

$$p_n(x) = A \sum_{i=1}^{M} w_i \, q_{(i)n}(x). \tag{33}$$

*The segments, $q_{(i)n}(x)$, are increasing, i.e., $q'_{(i)n}(x) > 0$, over their support intervals, $(x_i, x_{i+1})$. The points, $x_i$, define the local minima and maxima, such that, if $y_i$ is a local minimum, then $y_{i+1}$ is a local maximum and vice versa. The weights, $w_i = +1$, if $y_i$ is a local minimum, and $w_i = -1$, if $y_i$ is a local maximum. The constant, $A > 0$, is chosen, so that, $p_n(x)$, integrates to unity over its interval of support, $(x_1, x_{M+1})$. In addition, a continuity (smoothness) of order $C$ requires that the first $C$ derivatives,*

$$\lim_{\epsilon \to 0^+} p_n^{(k)}(x + \epsilon) = \lim_{\epsilon \to 0^+} p_n^{(k)}(x - \epsilon), \quad \forall x \in (x_1, x_{M+1}), \text{ and}, \ \forall k = 0, 1, \ldots, C \tag{34}$$

*which needs to be true at all points of the local minima and maxima, i.e.,*

$$\lim_{\epsilon \to 0^+} q_{(i)n}^{(k)}(x_{i+1} - \epsilon) = \lim_{\epsilon \to 0^+} q_{(i+1)n}^{(k)}(x_{i+1} + \epsilon), \quad i = 1, \ldots, M - 1. \tag{35}$$

In order to construct the desired segment polynomials, $q_{(i)n}(x)$, consider two increasing polynomials, $u_n(x) = \sum_{i=0}^{n} a_i x^i$, and, $v_n(x) = \sum_{i=0}^{m} b_i x^i$, such that, for some $x_0$, the derivatives,

$$\begin{aligned} u_n^{(k)}(x_0) &= -v_n^{(k)}(x_0), \quad k = 0, 1, \ldots, C \\ \sum_{i=k}^{n} a_i x_0^{i-k} &= -\sum_{i=k}^{m} b_i x_0^{i-k} \end{aligned} \tag{36}$$

or, in matrix notation,

$$\underbrace{\begin{bmatrix} x_0^n & x_0^{n-1} & \cdots & x_0 & 1 \\ x_0^{n-1} & x_0^{n-2} & \cdots & 1 & 0 \\ & \ddots & & \ddots & \\ x_0^{n-C} & \cdots & 1 & 0 & 0 \end{bmatrix}}_{X_C(x_0)} \cdot \underbrace{\begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{bmatrix}}_{\boldsymbol{a}} = - \begin{bmatrix} x_0^m & x_0^{m-1} & \cdots & x_0 & 1 \\ x_0^{m-1} & x_0^{m-2} & \cdots & 1 & 0 \\ & \ddots & & \ddots & \\ x_0^{m-C} & \cdots & 1 & 0 & 0 \end{bmatrix} \cdot \underbrace{\begin{bmatrix} b_m \\ b_{m-1} \\ \vdots \\ b_0 \end{bmatrix}}_{\boldsymbol{b}}. \tag{37}$$

For $m = n$, Equation (37) can be rewritten as

$$\boldsymbol{X}_C(x_0)(\boldsymbol{a} + \boldsymbol{b}) = \boldsymbol{0} \tag{38}$$

so the coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$ are in the null space of $\boldsymbol{X}_C(x_0)$.

Provided that $\boldsymbol{a}_{(i)}$ denotes the coefficients of $q_{(i)n}(x) = \sum_{i=0}^{n} a_i x^i$, it is required that,

$$\begin{aligned} \boldsymbol{X}_C(x_2)\left(\boldsymbol{a}_{(1)} + \boldsymbol{a}_{(2)}\right) &= \boldsymbol{0} \\ \boldsymbol{X}_C(x_3)\left(\boldsymbol{a}_{(2)} + \boldsymbol{a}_{(3)}\right) &= \boldsymbol{0} \\ &\vdots \\ \boldsymbol{X}_C(x_M)\left(\boldsymbol{a}_{(M-1)} + \boldsymbol{a}_{(M)}\right) &= \boldsymbol{0}. \end{aligned} \tag{39}$$

Note that the matrices, $\boldsymbol{X}_C(x_i)$, are computed assuming the control points, $x_i$.

Given the first vector of coefficients, $\boldsymbol{a}_{(1)}$, the other coefficient vectors, $\boldsymbol{a}_{(i)}$, $i = 2, 3, \ldots, M-1$, can be computed iteratively using the underdetermined sets of Equation (39). The numerical feasibility of this problem requires that the order, $n \gg C$.

The vector, $\boldsymbol{a}_{(1)}$, must be selected, so that $q_{(1)n}(x_1) = y_1$, and, $q_{(1)n}(x_{M+1}) = y_{M+1}$, and $q_{(1)n}(x) > 0$ is $C$-continuous for $x \in (x_1, x_2)$. Let sample $q_{(1)n}(x)$ at $K$ equidistant points between $x_1$ and $x_2$. The coefficients $\boldsymbol{a}_{(1)}$ are then the solution of the quadratic program,

$$\begin{aligned} &\min \left\langle \boldsymbol{a}_{(1)}, \boldsymbol{a}_{(1)} \right\rangle \\ s.t. \quad &\left\langle \boldsymbol{X}_0(x_1), \boldsymbol{a}_{(1)} \right\rangle = y_1, \quad \left\langle \boldsymbol{X}_0(x_2), \boldsymbol{a}_{(1)} \right\rangle = y_2 \\ &\left\langle w_1 \boldsymbol{X}_0 \Delta_1, \boldsymbol{a}_{(1)} \right\rangle > 0 \end{aligned} \tag{40}$$

where $\boldsymbol{X}_0(x) = [x^n, x^{n-1}, \ldots, x, 1]$, and $\Delta_1 = (x_2 - x_1)/(K-1)$ is the equidistant sampling step. The last condition in (40) enforces $q_{(1)n}(x)$ to be approximately increasing between the points $x_1$ and $x_2$. The other coefficients, $\boldsymbol{a}_{(i)}$, $i > 1$, are computed similarly to the program (40), but with an additional constraint due to (39). The extended quadratic program to compute these coefficients is defined as

$$\begin{aligned} &\min \left\langle \boldsymbol{a}_{(i)}, \boldsymbol{a}_{(i)} \right\rangle \\ s.t. \quad &\left\langle \boldsymbol{X}_0(x_i), \boldsymbol{a}_{(i)} \right\rangle = y_i, \quad \left\langle \boldsymbol{X}_0(x_{i+1}), \boldsymbol{a}_{(i)} \right\rangle = y_{i+1} \\ &\left\langle w_i \boldsymbol{X}_0 \Delta_i, \boldsymbol{a}_{(i)} \right\rangle > 0 \\ &\boldsymbol{X}_C(x_i)(\boldsymbol{a}_{(i-1)} + \boldsymbol{a}_{(i)}) = \boldsymbol{0} \end{aligned} \tag{41}$$

where $\Delta_i = (x_{i+1} - x_i)/(K-1)$, and $i = 2, 3, \ldots, M$.

More importantly, quadratic programs (40) and (41) require that the constraints are sufficiently underdetermined, i.e., $n \gg C$, otherwise, the solution may be difficult to find, or even may not exist. Moreover, the solution is less numerically stable for a linear program than for a quadratic program and, therefore, the quadratic programs should be considered.

## 3. Derived Characteristics of a Polynomial Distribution

The cumulative distribution function (CDF) can be readily obtained for Form I polynomial PDF as shown in Appendix A, i.e.,

$$\begin{aligned} P_n(x) &= \int_l^x p_n(v)\,\mathrm{d}v = \sum_{i=0}^{n} \frac{a_i}{i+1}(x^{i+1} - l^{i+1}), \quad x \in (l, u) \\ &= x\sum_{i=0}^{n} \frac{a_i}{i+1} x^i - l\sum_{i=0}^{n} \frac{a_i}{i+1} l^i = x\tilde{p}_n(x) - l\tilde{p}_n(l). \end{aligned} \tag{42}$$

Note also that, for symmetric support interval, when $u = -l$, $P_n(u) = \sum_{i=0}^{n} \frac{a_i}{i+1}(u^{i+1} - (-u)^{i+1})$, so the normalization of PDF to unity is only affected by even-index coefficients, $a_i$.

In the case of a Form II polynomial PDF, it is best to convert it to Form I first as shown in Appendix B.

The median ($q = 1/2$), and more generally, the quantile, $X_q$, of a polynomial distribution is defined as

$$P_n(X_q) = X_q\tilde{p}_n(X_q) - l\tilde{p}_n(l) \stackrel{!}{=} q, \quad 0 < q < 1. \tag{43}$$

Denoting, $P_0(q) = l\tilde{p}_n(l) + q$, the quantile is the unique root of the polynomial, $x\tilde{p}_n(x) - P_0(q) = a(q)(x - X_q)$.

The expressions for general moments and characteristic or moment-generating functions are derived in Appendices A–C, respectively.

The Kullback–Leibler (KL) divergence or relative entropy between two polynomial distributions, $p_n(x) = \sum_{i=0}^{n} a_i x^i = a_n \prod_{i=1}^{n}(x - r_i)$, and, $q_m(x) = b_m \prod_{j=1}^{m}(x - s_j)$, is defined as

$$\begin{aligned} \mathrm{KL}(p_n \| q_m) &= \int_l^u p_n(x) \log\frac{p_n(x)}{q_m(x)}\,\mathrm{d}x = \int_l^u \sum_{i=0}^{n} a_i x^i \log\frac{a_n \prod_{l=1}^{n}(x - r_l)}{b_m \prod_{j=1}^{m}(x - s_j)}\,\mathrm{d}x \\ &= \sum_{i=0}^{n} a_i \int_l^u x^i \left(\log\frac{a_n}{b_m} + \sum_{l=1}^{n}\log(x - r_l) - \sum_{j=1}^{m}\log(x - s_j)\right)\mathrm{d}x \\ &= \sum_{i=0}^{n} a_i \left(\frac{u^{i+1} - l^{i+1}}{i+1}\log\frac{a_n}{b_m} + \sum_{l=1}^{n}\int_l^u x^i\log(x - r_l)\,\mathrm{d}x - \sum_{j=1}^{m}\int_l^u x^i\log(x - s_j)\,\mathrm{d}x\right). \end{aligned} \tag{44}$$

The inner integral, $\int_l^u x^i \log(x - r_l)\,\mathrm{d}x$, can be expressed in terms of the hypergeometric, ${}_2F_1$, functions with the help of, for example, Mathematica software.

Differential entropy of a polynomial distribution, $p_n(x)$, is defined as

$$\begin{aligned} H(p_n) &= -\int_l^u p_n(x)\log p_n(x)\,\mathrm{d}x = -\int_l^u \sum_{i=0}^{n} a_i x^i \log a_n \prod_{i=1}^{n}(x - r_i)\,\mathrm{d}x \\ &= \sum_{i=0}^{n} a_i \left(\frac{\log a_n}{i+1}(u^{i+1} - l^{i+1}) + \sum_{l=1}^{n}\int_l^u x^i \log(x - r_l)\,\mathrm{d}x\right). \end{aligned} \tag{45}$$

Finally, the sum, $Z = X + Y$, of two independent random variables, $X$, and, $Y$, having the polynomial distributions, $p_n(x)$, and, $q_m(y)$, respectively, with the same interval of support, $(l, u)$, also has the polynomial distribution, $f_m(z)$, given by the convolution,

$$\begin{aligned} f_m(z) &= \int_l^u p_n(x) q_m(z-x)\,\mathrm{d}x = \int_l^u \sum_{i=0}^{n} a_i x^i \sum_{j=0}^{m} b_j (z-x)^j\,\mathrm{d}x \\ &= \sum_{i=0}^{n}\sum_{j=0}^{m} a_i b_j \int_l^u x^i (z-x)^j\,\mathrm{d}x = \sum_{i=0}^{n}\sum_{j=0}^{m} a_i b_j \sum_{k=0}^{j} \binom{j}{k} z^{j-k} \int_l^u x^{i+k}\,\mathrm{d}x \\ &= \sum_{i=0}^{n}\sum_{j=0}^{m} a_i b_j \sum_{k=0}^{j} \binom{j}{k} z^{j-k} \frac{1}{i+k+1}\left(u^{i+k+1} - l^{i+k+1}\right) \\ &= \sum_{i=0}^{n}\sum_{j=0}^{m}\sum_{k=0}^{j} \binom{j}{k} \frac{a_i b_j}{i+k+1}\left(u^{i+k+1} - l^{i+k+1}\right) z^{j-k}. \end{aligned} \tag{46}$$

## 4. Estimation Problems Involving Polynomial Distributions

The parameter estimation of the polynomial distributions is subject to the following equality and inequality constraints, respectively,

$$\begin{aligned} \boldsymbol{a}^T \boldsymbol{X}_0 - 1 &= 0 \\ -\boldsymbol{a}^T \boldsymbol{X}_k &\leq 0, \quad k = 1, 2, \ldots, K \end{aligned} \tag{47}$$

where the column vectors, $\boldsymbol{X}_0 = [(u^{i+1} - l^{i+1})/(i+1)]$, and, $\boldsymbol{X}_k = [(l + (k-1)(u-l)/(K-1))^i]$, $i = 0, 1, \ldots, n$. The equality constraint guarantees that the estimated polynomial integrates to unity. The second constraint requires that the estimated polynomial is non-negative at $K$ equidistant points within the interval of support, $(l, u)$. The value of $K$ must be determined empirically.

Consider the problem of estimating the coefficients, $\boldsymbol{a}$, of a polynomial PDF, $p_n(x; \boldsymbol{a})$. For $M$ independent measurements, $x_m$, the likelihood function is,

$$L(\boldsymbol{x}; \boldsymbol{a}) = \prod_{m=1}^{M} p_n(x_m; \boldsymbol{a}) = \prod_{m=1}^{M} \sum_{i=0}^{n} a_i x_m^i = \prod_{m=1}^{M} \boldsymbol{a}^T \boldsymbol{x}_m \tag{48}$$

where the column vector, $\boldsymbol{x}_m = [x_m^i]$, $i = 0, 1, \ldots, n$. The Karush—Kuhn-–Tucker (KKT) function representing the constrained maximum (log-) likelihood (ML) estimation is [42],

$$\mathcal{K}(\boldsymbol{a}; \boldsymbol{x}; \{\mu_k\}) = \log L(\boldsymbol{x}; \boldsymbol{a}) + \mu_0(\boldsymbol{a}^T \boldsymbol{X}_0 - 1) + \sum_{k=1}^{K} \mu_k\, \boldsymbol{a}^T \boldsymbol{X}_k \tag{49}$$

where $\mu_k \leq 0$, and, $\mu_k \boldsymbol{a}^T \boldsymbol{X}_k = 0$, for $\forall k \geq 1$. The first vector-derivative of $\mathcal{K}$ by $\boldsymbol{a}$ [41] must be equal to zero, i.e.,

$$\frac{\partial}{\partial \boldsymbol{a}} \mathcal{K}(\boldsymbol{a}; \boldsymbol{x}; \{\mu_k\}) \stackrel{!}{=} \boldsymbol{0} \quad \Leftrightarrow \quad \frac{\partial}{\partial \boldsymbol{a}} \sum_{m=1}^{M} \log(\boldsymbol{a}^T \boldsymbol{x}_m) = \sum_{m=1}^{M} \frac{\boldsymbol{x}_m^T}{\boldsymbol{a}^T \boldsymbol{x}_m} = -\sum_{k=0}^{K} \mu_k\, \boldsymbol{X}_k^T. \tag{50}$$

Expression (50) together with constraint (47) represent the set of $(n + K + 2)$ non-linear equations with the same number of unknowns, which must be solved numerically.

The cosine theorem can be assumed instead of logarithm in maximizing (49) and (50), since,

$$\underset{\boldsymbol{a}}{\operatorname{argmax}} \prod_{m=1}^{M} \boldsymbol{a}^T \boldsymbol{x}_m = \underset{\boldsymbol{a}}{\operatorname{argmax}} \prod_{m=1}^{M} \frac{\boldsymbol{a}^T}{\|\boldsymbol{a}\|} \frac{\boldsymbol{x}_m}{\|\boldsymbol{x}_m\|} = \underset{\boldsymbol{a}}{\operatorname{argmax}} \prod_{m=1}^{M} \cos\phi_m \tag{51}$$

where $\phi_m$ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{x}_m$. The unconstrained maximization of (51) can be performed using the geometric-arithmetic mean inequality [43]. Specifically, the likelihood (51) is maximized when the coefficient vector, $\boldsymbol{a}$, is aligned with all the observations, $\boldsymbol{x}_m$. It can be approximated by assuming that the distances between the normalized vectors, $\boldsymbol{a}^T/\|\boldsymbol{a}\|$, and, $\boldsymbol{x}_m/\|\boldsymbol{x}_m\|$, are constant, for $\forall m$. Then, the estimate is the centroid of observations, i.e., $\hat{\boldsymbol{a}} = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{x}_m/\|\boldsymbol{x}_m\|$.

The estimator complexity can be greatly reduced if constraint (47) is ignored. In this case, the estimated coefficients, $\hat{\boldsymbol{a}}$, may not satisfy the PDF conditions (1). Assuming that the estimate, $\hat{\boldsymbol{a}}$, is not too far from the true vector, $\boldsymbol{a}$, the first estimate can be improved by the subsequent estimator,

$$\hat{\hat{\boldsymbol{a}}} = \underset{\boldsymbol{a}}{\operatorname{argmin}} \|\boldsymbol{a} - \hat{\boldsymbol{a}}\| \quad s.t. \quad \text{constraints } (47). \tag{52}$$

The estimator (52) minimizes the distance between the first estimate, $\hat{\boldsymbol{a}}$, and the subsequent estimate, $\hat{\hat{\boldsymbol{a}}}$, under constraint (47). More importantly, the corresponding set of equations to solve (52) is now linear.

In the sequel, we drop the inequality constraints in (47) since closed-form expressions can be obtained with the equality constraint. For $M = 2$, or, equivalently, only two out of $M$ measurements are considered at a time, the ML estimator can be constrained as in (31), i.e.,

$$\hat{\boldsymbol{a}}_{12} = \underset{\boldsymbol{a}}{\operatorname{argmax}}\, \boldsymbol{a}^T \boldsymbol{X}_{12} \boldsymbol{a}, \quad \text{s.t.} \quad \boldsymbol{w}^T \boldsymbol{a} = 1 \tag{53}$$

where $\boldsymbol{x}_i = [x_i^j]$, $j = 0, 1, \ldots, n$, and $\boldsymbol{X}_{12} = \boldsymbol{x}_1 \boldsymbol{x}_2^T$ is a $(n+1) \times (n+1)$ square matrix. The first derivative of the corresponding Lagrangian must be equal to zero, i.e., [41]

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{a}}\left(\boldsymbol{a}^T \boldsymbol{X}_{12} \boldsymbol{a} + \lambda(\boldsymbol{w}^T \boldsymbol{a} - 1)\right) \overset{!}{=} \boldsymbol{0} \\ &\Rightarrow \quad \boldsymbol{a} = -\frac{\lambda}{2} \boldsymbol{X}_{12}^{-1} \boldsymbol{w}, \quad \lambda = \frac{-2}{\boldsymbol{w}^T \boldsymbol{X}_{12}^{-1} \boldsymbol{w}}. \end{aligned} \tag{54}$$

Consequently, the ML estimate is,

$$\hat{\boldsymbol{a}}_{12} = \frac{\boldsymbol{X}_{12}^{-1} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{X}_{12}^{-1} \boldsymbol{w}} \tag{55}$$

and its likelihood is equal to, $\left(\boldsymbol{w}^T \boldsymbol{X}_{12}^{-1} \boldsymbol{w}\right)^{-1}$. Finally, the observation pairs, $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, $(\boldsymbol{x}_3, \boldsymbol{x}_4)$, …, are independent, so the final ML estimate is,

$$\hat{\boldsymbol{a}} = \frac{2}{M} \sum_{i=1}^{M/2} \frac{\boldsymbol{X}_{2i-1,2i}^{-1} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{X}_{2i-1,2i}^{-1} \boldsymbol{w}}. \tag{56}$$

The Cramer–Rao bound has been defined to lower-bound the covariance matrix of the estimation error, $\hat{\boldsymbol{a}} - \boldsymbol{a}$, of any unbiased estimator, i.e.,

$$\operatorname{cov}[\hat{\boldsymbol{a}} - \boldsymbol{a}] \overset{\mathrm{E}[\hat{\boldsymbol{a}}] = \boldsymbol{a}}{=} \operatorname{var}[\hat{\boldsymbol{a}}] \geq \boldsymbol{J}^{-1}(\boldsymbol{a}) \tag{57}$$

where $\boldsymbol{J}(\boldsymbol{a})$ is the Fisher information matrix. In order to calculate the elements of this matrix, it is better to assume Form II of the polynomial distribution, $p_n(x; \boldsymbol{r}) = a_n \prod_{i=1}^{n}(x - r_i)$, and the problem of estimating the parameters, $\boldsymbol{r}$, i.e.,

$$
\begin{aligned}
[\boldsymbol{J}]_{ij} &= \mathrm{E}\left[\left(\frac{\partial}{\partial r_i}\log p_n(x;\boldsymbol{r})\right)\left(\frac{\partial}{\partial r_j}\log p_n(x;\boldsymbol{r})\right)\right] \\
&= \mathrm{E}\left[\left(\frac{\partial}{\partial r_i}\log(x-r_i)\right)\left(\frac{\partial}{\partial r_j}\log(x-r_j)\right)\right] \\
&= \mathrm{E}\left[\frac{1}{(x-r_i)}\frac{1}{(x-r_j)}\right] = \int_l^u a_n \prod_{\substack{k=1\\k\neq i,j}}^{n}(x-r_k)\,\mathrm{d}x.
\end{aligned}
\tag{58}
$$

The last integral in (58) can be computed by converting the Form II polynomial into Form I.

The coefficients, $\boldsymbol{a}$, can also be estimated by the method of moments [7,44]. In particular, the $k$-th general moment of a polynomial distribution, $p_n(x)$, is, (cf. Appendix A)

$$
\mathcal{M}_k = \int_l^u x^k \sum_{i=0}^{n} a_i x^i\,\mathrm{d}x = \sum_{i=0}^{n}\frac{a_i}{i+k+1}\left(u^{i+k+1}-l^{i+k+1}\right). \tag{59}
$$

Observing a vector of the first $K$ general moments, $\boldsymbol{M} = [\mathcal{M}_k]$, $k = 1, 2, \ldots, K$, and precomputing the matrix, $\boldsymbol{B} = [(i+k+1)^{-1}(u^{i+k+1} - l^{i+k+1})]$, $i = 0, 1, \ldots, n$, and ignoring non-negativity constraint in (47), the estimation can be again defined as the constrained or unconstrained least-square regression, i.e.,

$$
\hat{\boldsymbol{a}} = \underset{\boldsymbol{a}}{\operatorname{argmin}}\|\boldsymbol{M} - \boldsymbol{B}\boldsymbol{a}\|, \quad \text{s.t.} \quad \boldsymbol{w}^T\boldsymbol{a} = 1 \tag{60}
$$

which can be efficiently solved as in (32).

*Other Estimation Problems*

In addition to estimating the coefficients, $a_i$, of the polynomial distribution, $p_n(x)$, an important task is to also decide the polynomial order, $n$. The usual strategy is to estimate a sequence of PDFs, $p_n(x)$, with increasing orders, $n = n_{\min}, n_{\min}+1, \ldots$, and then choose the one minimizing the Akaike information criterion (AIC),

$$
\mathrm{AIC}(n) = 2(n+1) - 2L(\boldsymbol{x};\boldsymbol{a}) \tag{61}
$$

which penalizes the model complexity, $\propto (n+1)$, while maximizing the likelihood $L$ defined in (48).

However, the AIC cannot be used with non-parametric and other likelihood-free estimation methods. In the case of the polynomial distributions, the empirical moments can be readily computed from the observed random samples as well as estimated using the inferred distribution parameters. In particular, denote $\boldsymbol{M}_{\mathrm{emp}}(\boldsymbol{x})$, the vector of the first $K$ empirical moments. Let, $\boldsymbol{M}_{\mathrm{est}}(\hat{\boldsymbol{a}})$ be the vector of the first $K$ moments computed assuming the estimated coefficients of the polynomial distribution, $p_n(x)$. Similarly to the AIC definition (61), we propose that the goodness of fit (GoF) of a polynomial distribution to observed data can be estimated as

$$
\mathrm{GoF}(n) = \left\|\boldsymbol{M}_{\mathrm{emp}}(\boldsymbol{x}) - \boldsymbol{M}_{\mathrm{est}}(\hat{\boldsymbol{a}})\right\|_2^2 - n - 1. \tag{62}
$$

Thus, the squared Euclidean distance between the two-moment vectors is penalized by the number of model coefficients, i.e., $(n+1)$. In our numerical examples, we observed that $n$ should be chosen as the smallest polynomial order causing a sudden significant drop in the value of GoF.

Furthermore, Bayesian estimation methods for estimating the coefficients, $\boldsymbol{a}$, require adopting a prior, $p(\boldsymbol{a})$. Since the coefficients are likely to be mutually correlated, defining the prior distribution may be challenging, unless a Gaussian prior can be assumed. On the other hand, consider a general probabilistic model with observations $X$ and a parameter,

$\theta$, which is described by the likelihood, $p(x|\theta)$, and the prior, $p(\theta)$. If the likelihood and the prior are both polynomially distributed, then the corresponding posterior, $p(\theta|x) \propto p(x|\theta)p(\theta)$, is also polynomially distributed.

Finally, polynomial distributions can also be considered in variational Bayesian inference to approximate the posterior, $p(\theta|\boldsymbol{x})$, and lower-bound the evidence, $p(\boldsymbol{x})$. The parameters of the polynomial distribution can also be inferred without computing the likelihood by assuming the approximate Bayesian computations.

## 5. Numerical Examples

This section briefly explores the numerical properties of the polynomial distributions by means of several examples, since the practical approximations of functions is a rather extensive subject [45]. The examples are presented in Sections 5.1–5.4, including constructing a piecewise polynomial PDF, generating samples of a polynomially distributed random variable, estimating the parameters of a polynomial PDF from observed random samples, and approximating a given PDF by a polynomial PDF.

### *5.1. Constructing Piecewise Polynomial PDF*

The first example in Figure 1 demonstrates the construction of a piecewise polynomial PDF as described in Section 2.3. Given the continuity order $C = 2$, the given set of 5 control points can be connected by 4 increasing or decreasing polynomial functions of order $n = 5$ or $n = 6$. For polynomial orders smaller than 5 or larger than about 8, it was observed that Runge's phenomenon starts to appear, which indicates that the polynomial order cannot be chosen completely arbitrarily.

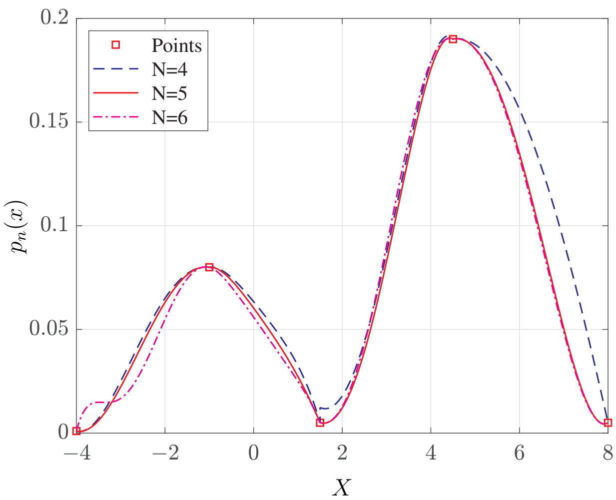


**Figure 1.** The piecewise polynomial distributions, $p_n(x)$, consisting of 4 segments with 5 control points (red squares), when the continuity order $C = 2$. The segments were computed using 20 equally spaced samples within each segment.

### *5.2. Generating Random Samples Using Polynomial PDF*

Consider now the methods how to generate polynomially distributed random samples. Similar to most other distributions, polynomial distributions cannot be easily inverted. It is then challenging to use the inverse method for generating random variables. On the other hand, the CDF of a polynomially distributed random variable is another polynomial as shown in Appendix A. A CDF discretization can be used as a general strategy for implementing the inverse method of generating random samples from the distribution with a known CDF. In particular, consider approximating the CDF by a piecewise linear function between the samples, $(x_i, F(x_i))$, $i = 1, 2, \ldots$. The inverse value, $X = F^{-1}(U)$, where $U \in (0,1)$ is a uniformly distributed random variable, is then approximated by a piecewise linear function as

$$x = x_i + \frac{(x_{i+1} - x_i)(u - F(x_i))}{F(x_{i+1} - F(x_i))}. \tag{63}$$

Moreover, polynomial distributions can also be assumed to obtain a proposal distribution for the rejection and importance of sampling methods. Assuming the latter, denote the expected value, $\theta = \mathrm{E}_q[g(x)]$, assuming a random observation, $X$, from a complex distribution, $q(x)$. The mean can be computed by assuming instead the distribution, $p(x)$, as

$$\theta = \int_l^u g(x)\frac{q(x)}{p(x)}p(x)\,\mathrm{d}x = \mathrm{E}_p\left[g(x)\frac{q(x)}{p(x)}\right]. \tag{64}$$

The corresponding difference in variances is,

$$\mathrm{var}_q[g(x)] - \mathrm{var}_p\left[g(x)\frac{q(x)}{p(x)}\right] = \int_l^u g^2(x)\left(1 - \frac{q(x)}{p(x)}\right)q(x)\,\mathrm{d}x. \tag{65}$$

Consequently, the better $p(x)$ approximates $g(x)q(x)$, the larger the variance reduction by sampling from $p(x)$ instead of $q(x)$. It is easy to show that, if $p(x) = g(x)q(x)/\theta$, then $\mathrm{var}_p\left[g(x)\frac{q(x)}{p(x)}\right] = 0$. Consequently and importantly, it is likely that the flexibility of the polynomial PDF, $p_n(x)$, allows approximating $g(x)q(x)$ with much better accuracy than any other canonical distribution.

Figure 2 assumes the piecewise polynomial distribution, $p_5(x)$, designed in the example in Figure 1 in order to generate samples from the target distribution, $q(x)$. The distribution, $q(x)$, is chosen to be a mixture of two truncated skewed-Gaussian distributions [46,47]. In particular, the rejection sampling method requires that, $q(x) \le Ap_5(x)$, for some $A > 0$ and $\forall x$. Note that the third and the fifth (counting from left) control points were moved slightly above zero in Figure 2 (and already also in Figure 1) in order to satisfy the required inequality.

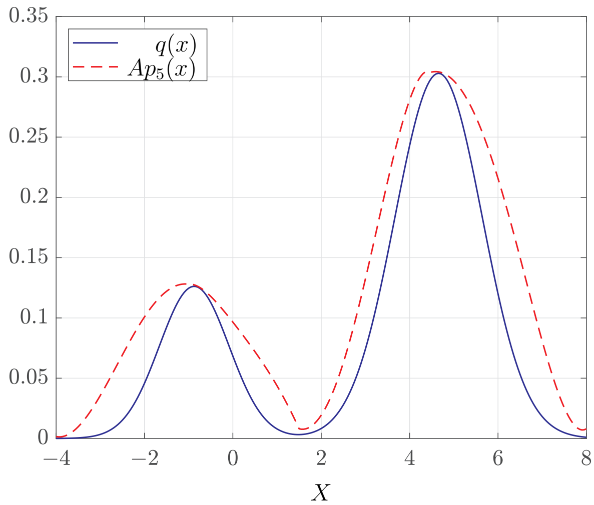


**Figure 2.** The rejection sampling from distribution $p_5(x)$ designed in Figure 1 to generate samples from the target distribution, $q(x) \le Ap_5(x)$, where $A = 1.6$.

In the sequel, the examples assume the polynomial PDF,

$$p_4(x) = \frac{75}{896}\left(-\frac{x^4}{30} + \frac{x^3}{5} + \frac{x^2}{10} - \frac{26x}{15} + 2\right), \quad x \in (-3,5). \tag{66}$$

The random samples from this distribution were generated by discretizing the support interval, $(-3, 5)$, into 30 equally sized bins. The distribution and the empirical histogram for $10^4$ generated samples are shown in Figure 3.

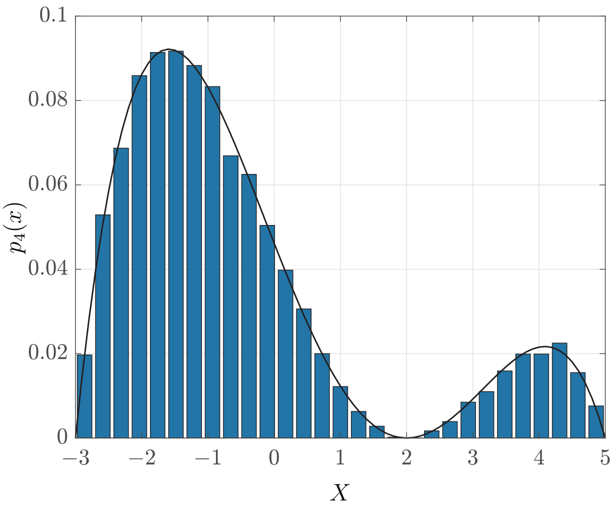

**Figure 3.** The PDF (66) and the empirical histogram for $10^4$ generated random samples.

*5.3. Estimating Parameters of Polynomial PDF*

Estimating the coefficients of the polynomial PDF must be constrained in order to satisfy the PDF conditions (1). The numerical experiments revealed that ML methods described in Section 4 are generally less suitable for estimating longer vectors of parameters. The ML methods require a relatively large number of observations, which quickly leads to the accumulation of numerical errors. On the other hand, non-parametric methods, such as the method of moments and the LS fitting appear to be much more robust, and they work well, even if the number of observations is relatively small.

Figure 4 shows the absolute error between the PDF (66) and the PDF, $\hat{p}_4(x)$, estimated using the method of moments with 500 data samples. It can be observed that at least the first $K = 5$ moments are required to achieve a good estimation accuracy.

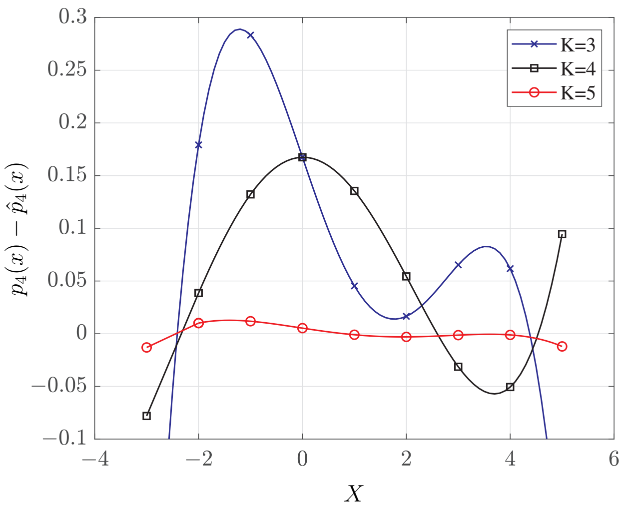


**Figure 4.** The estimation error, $p_4(x) - \hat{p}_4(x)$, for the method of moments with 500 data samples.

Figure 5 shows the distributions of the estimated coefficients for the PDF (66) using a constrained LS fit of the histogram with 10 bins and 50 data samples. The little triangles at the sides of the boxes in Figure 5 indicate the true values of the coefficients. Interestingly, the smallest accuracy in terms of the largest variance occurs for the lower-order coefficients,

$a_1$, and, $a_0$. If the polynomial order, $n$, is not known a priori, it can be determined by plotting the GoF measure (62) as shown in Figure 6. It can be observed that the value of GoF drops sharply when $n$ is increased to 4, and then it remains nearly constant even if $n$ is increased further.

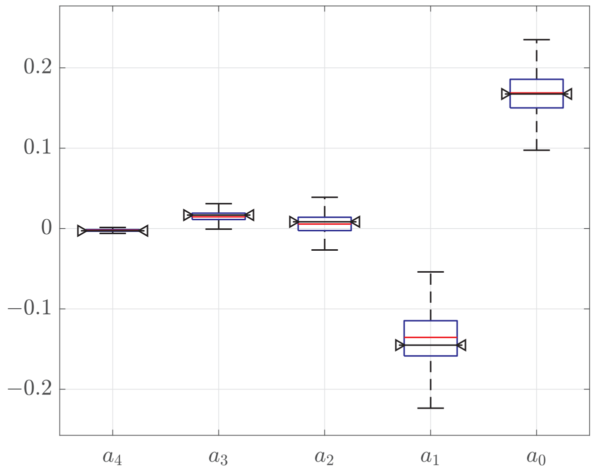


**Figure 5.** The box plots of estimated coefficients of the polynomial PDF (66) from only 50 samples, and repeated 1000 times. The triangles on the side of the boxes indicate the true coefficient values.

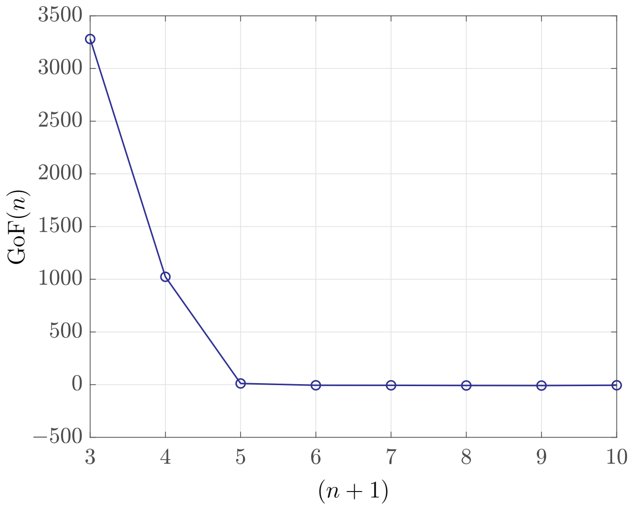


**Figure 6.** The goodness-of-fit measure (62) as a function of the polynomial order, $n$.

### *5.4. Approximating PDF by Polynomial PDF*

Even if the closed-form expression of a given PDF is available, the expression may be too complex to use in mathematical derivations [46,47]. In this case, a polynomial approximation of the PDF can be assumed. The approximation of a given PDF by a polynomial PDF is studied in Figure 7. In particular, Figure 7 compares the ordinary (unconstrained) LS fit (panel A), and the constrained LS fit (panel B) in order to satisfy the PDF constraints (1). It can be observed that assuming the constraints suppresses Runge's phenomenon. However, in all cases considered, the polynomial PDF struggles to approximate the peak value of the given PDF, even when the polynomial order is increased.

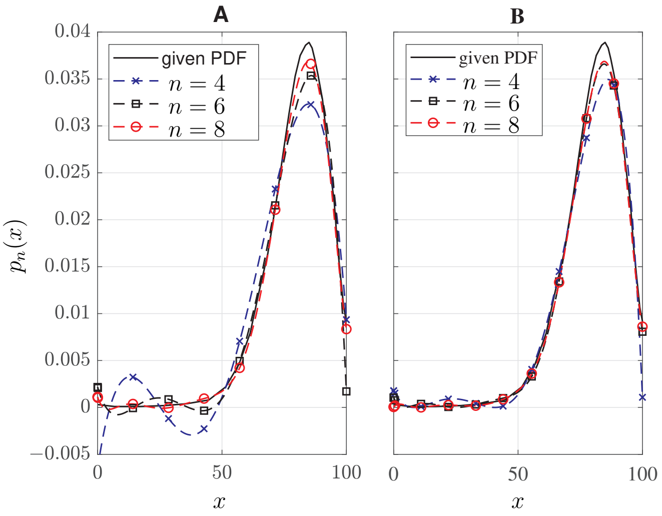


**Figure 7.** The approximations of a given PDF by polynomial PDFs, $p_n(x)$, with $n = 3, 5$, and 7 using the unconstrained (**A**), and constrained (**B**) LS approximation, respectively.

Finally, the approximation of a given PDF by Lagrange polynomials defined in Section 2.2 is investigated in Figure 8. Recall that Lagrange approximation is an interpolating polynomial function having the order given by the number of sampling points. Specifically, the left panel (A) in Figure 8 assumes equidistant sampling points, whereas the right panel (B) in Figure 8 shows the approximations with Chebyshev sampling points [4]. It can be observed that the latter case is clearly more accurate including the peak of the given PDF. More importantly, for the PDF example in Figure 8, the polynomial PDF approximation consisting of Lagrange polynomials is much easier to compute than performing the constrained LS curve fitting or using other statistical estimation methods, such as the method of moments.

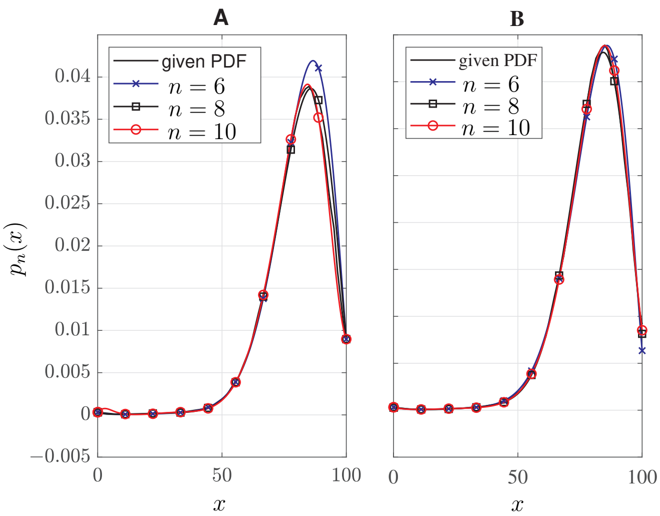


**Figure 8.** The approximations of a given PDF by Lagrange polynomials assuming $N = 6, 8$, and 10 equidistant samples (**A**), and the same number of Chebyshev samples (**B**).

### *5.5. Summary of Observations*

In general, the ML estimation of a vector of parameters of a polynomial distribution becomes numerically problematic when the polynomial orders are larger than only about 2 (and because numerical errors quickly accumulate with the number of observations). On the other hand, non-parametric and likelihood-free methods appear to be much more robust and numerically stable. For instance, the method of moments and the constrained LS fit can estimate the parameters of polynomial distributions efficiently, even if the number of observations is relatively small.

It is always useful to inspect the estimated polynomial coefficient values since some of those values can be near zero. The corresponding coefficients can then be rounded to zero, i.e., removed. Provided that these coefficients are the ones with the largest indices, the order of the estimated polynomial is reduced.

If the likelihood is not computed, the AIC cannot be used to determine the best polynomial order. In this paper, we propose defining the goodness of the polynomial fit as the squared Euclidean distance between the empirical and estimated general moments, which are penalized by the number of parameters to be estimated.

The Lagrange interpolation yields a polynomial function of the same order as the number of distribution samples. This method is usually less accurate when the samples are equidistant. Better accuracy can be achieved with Chebyshev or extended Chebyshev samples. Consequently, the bins of a histogram should be optimized rather than assumed with equal sizes in order to improve the interpolation accuracy.

In some cases, Runge's phenomenon can be avoided or mitigated by extending the given finite support interval with zero samples since oscillations of the approximating function tend to concentrate near the approximation interval boundaries.

## 6. Conclusions

Polynomials are often used for approximating univariate and multivariate functions, including probability distributions in order to enable mathematical tractability and simulation efficiency. This paper defined polynomial distributions, which can also be used to approximate other canonical and empirically estimated distributions over a finite interval of support. Polynomial distributions can be considered as more flexible alternatives to commonly used canonical distributions. More importantly, in this paper, many key properties, as well as limitations of the polynomial distributions, were derived and presented as closed-form expressions.

There is a need for defining a family of distributions, such as polynomial distributions that can be used more universally for solving problems in probability, statistics, and data analysis. The polynomial distributions considered in this paper are univariate and continuous; the extension to multivariate and discrete polynomial distributions may be the subject of future work. Polynomials could be generalized as weighted linear sums of non-linear functions in the same variables. A number of research problems remain open. For example, given a polynomial, it would be useful to algebraically identify all sub-intervals where it is non-negative. Or, given a polynomial order and an interval of support, the task is to find all polynomials that represent a PDF. This problem can be further constrained by the desired number of modes, the smoothness and/or sparsity conditions, and assuming other statistical and algebraic properties of the polynomial distributions. Moreover, the problem of determining the minimum polynomial degree or sparsity to satisfy the given constraints was not fully considered in this paper. It would be very useful to investigate how to interpret polynomial distributions, especially as they may arise naturally when observing some stochastic phenomena.

**Author Contributions:** Conceptualization, P.L.; Formal analysis, Y.Y. and P.L.; Investigation, Y.Y. and P.L.; Supervision, P.L.; Project administration, P.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a research grant from Zhejiang University.

**Data Availability Statement:** No new data were created or analyzed in this study.

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

## Appendix A. Basic Properties of Form I Polynomials

Definition:

$$p_n(x) = \sum_{i=0}^{n} a_i\, x^i, \quad a_i \in \mathcal{R},\ a_n \neq 0 \tag{A1}$$

Roots, $n = 1$:

$$p_1(x) = 0 \quad \Leftrightarrow \quad x_1 = -\frac{a_0}{a_1} \tag{A2}$$

Roots, $n = 2$:

$$p_2(x) = 0 \quad \Leftrightarrow \quad \begin{array}{ll} x_{1,2} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2a_0}}{2a_2} & a_1^2 > 4a_2a_0 \\ x_1 = x_2 = -\frac{a_1}{2a_2} & a_1^2 = 4a_2a_0 \\ x_{1,2} \notin \mathcal{R} & a_1^2 < 4a_2a_0 \end{array} \tag{A3}$$

Roots, $n = 3$:

$$p_3(x) = 0 \quad \Leftrightarrow \quad \begin{array}{ll} x_1 = \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} + \sqrt[3]{-\frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}},\ x_{2,3} \notin \mathcal{R} & D < 0 \\ x_1 = x_2 = x_3 = -\frac{a_2}{3a_3} & D = 0,\ a_2^2 = 3a_3a_1 \\ x_1 = x_2 = \frac{9a_3a_0 - a_2a_1}{2(a_2^2 - 3a_3a_1)},\ x_3 = \frac{4a_3a_2a_0 - 9a_3^2a_0 - a_2^3}{a_3(a_2^2 - 3a_3a_1)} & D = 0,\ a_2^2 \neq 3a_3a_1 \\ x_k = -\frac{1}{3a}(b + \xi^{k-1}C + \frac{\Delta_0}{\xi^{k-1}C}),\ k = 1,2,3 & D > 0 \end{array} \tag{A4}$$

where

$$a = \frac{-a_2}{3a_3}, \quad b = a^3 + \frac{a_2a_1 - 3a_3a_0}{6a_3^2}, \quad D = \frac{4(a_2^2 - 3a_3a_1) - (2a_2^3 - 9a_3a_2a_1 + 27a_3^2a_0^2)}{27a_3^2}$$

$$\Delta_0 = a_2^2 - 3a_3, \quad \Delta_1 = 2a_2^3 - 9a_3a_2a_1 + 27a_3^2a_0, \quad C = \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}, \quad \xi = \frac{-1 + \sqrt{-3}}{2}$$

Roots, general case:

- By Abel—Ruffini's theorem, closed-form expressions for roots of a polynomial exist for polynomials of degree at most $n = 4$, and there is no algebraic solution for the polynomial roots for degree $n > 4$.
- The total number of real roots of a polynomial within a given interval or overall real numbers can be determined by Sturm's theorem.
- Other relationships between polynomial coefficients and roots can be obtained, such as the Vieta's formulas:

$$\sum_{i_1 \neq i_2 \neq \cdots \neq i_k} r_{i_1} r_{i_2} \cdots r_{i_k} = (-1)^i\, \frac{a_{n-i}}{a_n}, \quad k \leq i = 1, 2, \ldots, n \tag{A5}$$

Indefinite integral:

$$\tilde{p}_n(x) \equiv \int p_n(x)\, \mathrm{d}x = \sum_{i=0}^{n} \frac{a_i}{i+1} x^{i+1} \tag{A6}$$

Definite integral:

$$P_n(u) \equiv \int_{-\infty}^{u} p_n(x)\, \mathrm{d}x \tag{A7}$$

$$\begin{aligned}\int_l^u p_n(x)\,\mathrm{d}x &= \int_{-\infty}^u p_n(x)\,\mathrm{d}x - \int_{-\infty}^l p_n(x)\,\mathrm{d}x = P_n(u) - P_n(l), \quad l < u \\ &= \sum_{i=0}^n \frac{a_i}{i+1}\left(u^{i+1} - l^{i+1}\right) \equiv \tilde{p}(u) - \tilde{p}(l)\end{aligned} \tag{A8}$$

Indefinite $k$-fold integral, $k > 1$:

$$\underbrace{\int \cdots \int}_{k} p_n(x)\,\mathrm{d}x^k = \sum_{i=0}^n \frac{a_i}{(i+1)(i+2)\ldots(i+k)}\, x^{i+k} = \sum_{i=0}^n a_i\,\frac{i!}{(i+k)!}\, x^{i+k} \tag{A9}$$

Derivative:

$$\dot{p}_n(x) = \frac{\mathrm{d}}{\mathrm{d}x} p_n(x) = \sum_{i=1}^n i\, a_i\, x^{i-1} \tag{A10}$$

$k$-th derivative, $1 < k \leq n$:

$$\begin{aligned}p_n^{(k)}(x) &= \sum_{i=k}^n i(i-1)\ldots(i-k+1)\, a_i\, x^{i-k} \\ &= \sum_{i=k}^n a_i\,\frac{i!}{(i-k)!}\, x^{i-k} = \sum_{i=k}^n a_i\, k!\binom{i}{k}\, x^{i-k}\end{aligned} \tag{A11}$$

$k$-th moment, $k \geq 1$:

$$\begin{aligned}\int x^k\, p_n(x)\,\mathrm{d}x &= \sum_{i=0}^n \frac{a_i}{i+k+1} x^{i+k+1} \\ \int_l^u x^k\, p_n(x)\,\mathrm{d}x &= \sum_{i=0}^n \frac{a_i}{i+k+1}\left(u^{i+k+1} - l^{i+k+1}\right)\end{aligned} \tag{A12}$$

Characteristic function:

$$\begin{aligned}\phi_X(t) = \mathrm{E}\left[\mathrm{e}^{\mathrm{j}tX}\right] &= \int_{-1}^{+1} p_n(x)\,\mathrm{e}^{\mathrm{j}tx}\,\mathrm{d}x = \sum_{i=0}^n a_i \int_{-1}^{+1} x^i\,\mathrm{e}^{\mathrm{j}tx}\,\mathrm{d}x \\ &= \sum_{i=0}^n a_i\,\frac{\mathrm{j}}{t}\int_{-t/\mathrm{j}}^{t/\mathrm{j}} \left(\frac{x\mathrm{j}}{t}\right)^i \mathrm{e}^{-x}\,\mathrm{d}x \\ &= \sum_{i=0}^n a_i \left(\frac{\mathrm{j}}{t}\right)^{i+1}\left(\int_{-t/\mathrm{j}}^{\infty} x^i\,\mathrm{e}^{-x}\,\mathrm{d}x - \int_{t/\mathrm{j}}^{\infty} x^i\,\mathrm{e}^{-x}\,\mathrm{d}x\right) \\ &= \sum_{i=0}^n \frac{a_i}{(-\mathrm{j}t)^{i+1}}\left(\Gamma(1+i,\mathrm{j}t) - \Gamma(1+i,-\mathrm{j}t)\right)\end{aligned} \tag{A13}$$

where $\Gamma(a,z)$ is the incomplete Gamma function

## Appendix B. Basic Properties of Form II Polynomials

Definition:

$$p_n(x) = a_n \prod_{i=1}^n (x - r_i), \quad a_n \neq 0,\ r_i \in \tag{A14}$$

Recursive form:

$$\begin{aligned} p_n(x) &= \frac{a_n}{a_{n-1}}\,(x - r_n)\,p_{n-1}(x), \quad n > 1 \\ p_1(x) &= x - r_1 \end{aligned} \tag{A15}$$

Converting Form I to Form II, general case:

$$\sum_{i=0}^{n} a_i x^i = a_n \prod_{i=1}^{n} (x - r_i) \tag{A16}$$

$$\begin{aligned} a_n \equiv a_n, \quad a_{n-1} = a_n(-1)^1 \sum_{i=1}^{N} r_i, \quad a_{n-2} = a_n(-1)^2 \sum_{i=1, j=1, i \neq j}^{N} r_i r_j \\ a_{n-3} = a_n(-1)^3 \sum_{i=1, j=1, k=1, i \neq j \neq k}^{N} r_i r_j r_k, \quad \cdots \quad , a_0 = a_n(-1)^n \prod_{i=1}^{n} r_i \end{aligned} \tag{A17}$$

Converting Form I to Form II, $n = 2$:

$$a_2 \equiv a_2, \quad a_1 = -a_2(r_1 + r_2), \quad a_0 = a_2 r_1 r_2 \tag{A18}$$

Converting Form I to Form II, $n = 3$:

$$a_3 \equiv a_3, \quad a_2 = -a_3(r_1 + r_2 + r_3), \quad a_1 = a_3(r_1 r_2 + r_1 r_3 + r_2 r_3), \quad a_0 = -a_3 r_1 r_2 r_3 \tag{A19}$$

Converting Form I to Form II, $n = 4$:

$$\begin{aligned} a_4 \equiv a_4, \quad a_3 &= -\,a_4(r_1 + r_2 + r_3 + r_4), \quad a_2 = a_4(r_1 r_2 + r_1 r_3 + r_1 r_4 + r_2 r_3 + r_2 r_4 + r_3 r_4) \\ a_1 &= -\,a_4(r_1 r_2 r_3 + r_1 r_2 r_4 + r_1 r_3 r_4 + r_2 r_3 r_4), \quad a_0 = a_4 r_1 r_2 r_3 r_4 \end{aligned} \tag{A20}$$

Indefinite integral: (recursive form)

$$\begin{aligned} I_n(x) &= \frac{1}{a_n} \int p_n(x)\,\mathrm{d}x, \quad n > 1 \\ &= (x - r_n) I_{n-1}(x) - \int I_{n-1}(x)\,\mathrm{d}x \\ I_1(x) &= \frac{1}{a_1} \int p_1(x)\,\mathrm{d}x = \int (x - r_1)\,\mathrm{d}x = \frac{1}{2}x^2 - r_1 x \end{aligned} \tag{A21}$$

Indefinite $k$-fold integral:

$$\begin{aligned} I_n(x) &= (x - r_n) I_{n-1}(x) - \int I_{n-1}(x)\,\mathrm{d}x \\ \int I_n(x)\,\mathrm{d}x &= \int (x - r_n) I_{n-1}(x)\,\mathrm{d}x - \iint I_{n-1}(x)\,\mathrm{d}x^2 \\ &= (x - r_n) \int I_{n-1}(x)\,\mathrm{d}x - 2 \iint I_{n-1}(x)\,\mathrm{d}x^2 \\ \underbrace{\int \cdots \int}_{k} I_n(x)\,\mathrm{d}x^k &= (x - r_n) \underbrace{\int \cdots \int}_{k} I_{n-1}(x)\,\mathrm{d}x^k - (k+1) \underbrace{\int \cdots \int}_{k+1} I_{n-1}(x)\,\mathrm{d}x^{k+1} \end{aligned} \tag{A22}$$

Derivative:

$$\begin{aligned}
\dot{p}_n(x) &= a_n \prod_{i=1}^{n-1}(x - r_i) + \frac{a_n}{a_{n-1}}(x - r_n)\dot{p}_{n-1}(x), \quad n > 1 \\
&= \frac{a_n}{a_{n-1}} p_{n-1}(x) + \frac{a_n}{a_{n-1}}(x - r_n)\dot{p}_{n-1}(x) \\
\dot{p}_1(x) &= a_1
\end{aligned} \tag{A23}$$

$k$-th derivative, $1 < k \leq n$:

$$\begin{aligned}
\dot{p}_n(x) &= \frac{a_n}{a_{n-1}}\, p_{n-1}(x) + \frac{a_n}{a_{n-1}}(x - r_n)\, \dot{p}_{n-1}(x) \\
p_n^{(k)}(x) &= k\frac{a_n}{a_{n-1}}\, p_{n-1}^{(k-1)}(x) + \frac{a_n}{a_{n-1}}(x - r_n)\, p_{n-1}^{(k)}(x)
\end{aligned} \tag{A24}$$

$k$-th moment, $k \geq 1$:

$$\begin{aligned}
\tilde{p}_n(x) \equiv \int p_n(x)\,\mathrm{d}x, &\quad \overset{k}{\tilde{p}}_n(x) \equiv \underbrace{\int \cdots \int}_{k} p_n(x)\,\mathrm{d}x \\
\int x^k\, p_n(x)\,\mathrm{d}x &= x^k\, \tilde{p}_n(x) - k \int x^{k-1} \tilde{p}_n(x)\,\mathrm{d}x^2 \\
\int x^{k-1}\, \tilde{p}_n(x)\,\mathrm{d}x &= x^{k-1} \overset{2}{\tilde{p}}_n(x) - (k-1) \int x^{k-2}\, \overset{2}{\tilde{p}}_n(x)\,\mathrm{d}x \\
&\vdots \\
\int x\, \overset{k-1}{\tilde{p}}_n(x)\,\mathrm{d}x &= x\overset{k}{\tilde{p}}_n(x) - \overset{k+1}{\tilde{p}}_n(x) \\
\int_l^u x^k\, p_n(x)\,\mathrm{d}x &= \left[x^k \tilde{p}_n(x)\right]_l^u - k\int_l^u x^{k-1} \tilde{p}_n(x)\,\mathrm{d}x
\end{aligned} \tag{A25}$$

Moment-generating function:

$$\begin{aligned}
M(t) = \int_l^u e^{tx} p_n(x)\,\mathrm{d}x &= \frac{e^{tx}}{t} p_n(x) - \frac{1}{t}\int_l^u e^{tx} \overset{1}{\dot{p}}_n(x)\,\mathrm{d}x \\
\int_l^u e^{tx} \overset{1}{\dot{p}}_n(x)\,\mathrm{d}x &= \frac{e^{tx}}{t} p_n'(x) - \frac{1}{t}\int_l^u e^{tx} \overset{2}{\dot{p}}_n(x)\,\mathrm{d}x \\
&\vdots \\
\int_l^u e^{tx} \overset{n}{\dot{p}}_n(x)\,\mathrm{d}x &= \int_l^u e^{tx} a_n n!\,\mathrm{d}x = \frac{a_n n!}{t} e^{t^u - t^l}
\end{aligned} \tag{A26}$$

## Appendix C. Basic Properties of Form III Polynomials

Definition:

$$p_n(x) = \frac{s_m(x)}{q_n(x)} = \frac{s_m(x)}{c_n \prod_{i=1}^n (x - r_i)} = \sum_{i=1}^n \frac{a_i}{x - r_i}, \quad m < n,\ c_n \neq 0,\ r_i \neq r_j\ \forall i \neq j \tag{A27}$$

where the residuals, $a_i = \frac{s_m(r_i)}{\dot{q}_n(r_i)} \neq 0$

Indefinite integral:

$$r_i \in \mathcal{R}: \int_l^u \frac{1}{x-r_i}\,\mathrm{d}x = \begin{cases} \ln\frac{u-r_i}{l-r_i}, & r_i < l \text{ or } r_i > u \\ n.c. & \text{otherwise} \end{cases}$$
$$r_i \in: \int_l^u \frac{1}{x-r_i}\,\mathrm{d}x = \begin{cases} \ln\frac{u-r_i}{l-r_i}, & \mathrm{Re}(r_i) < l \text{ or } \mathrm{Re}(r_i) > u \text{ or } \mathrm{Im}(r_i) \neq 0 \\ n.c. & \text{otherwise} \end{cases} \quad \text{(A28)}$$
$$\int \sum_{i=1}^{n} \frac{c_i}{x-r_i}\,\mathrm{d}x = \sum_{i=1}^{n} c_i \ln(x-r_i)$$

Definite integral:

$$\int_l^u \sum_{i=1}^{n} \frac{c_i}{x-r_i}\,\mathrm{d}x = \sum_{i=1}^{n} c_i \frac{\ln(u-r_i)}{\ln(l-r_i)} \quad \text{(A29)}$$

$k$-th derivative, $1 \le k \le n$:

$$p_n^{(k)}(x) = (-1)^k k! \sum_{i=1}^{n} \frac{c_i}{(x-r_i)^{k+1}} \quad \text{(A30)}$$

$k$-fold integral, $k \ge 1$:

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k}\left(\frac{x^{k-1}\ln x}{(k-1)!}\right) = \frac{1}{x} \quad \text{(A31)}$$

$$\underbrace{\int \cdots \int}_{k} p_n(x)\,\mathrm{d}x^k = \sum_{i=1}^{n} c_i \frac{(x-r_i)^{k-1}\ln(x-r_i)}{(k-1)!} \quad \text{(A32)}$$

$k$-th moment, $k \ge 1$:

$$\int_l^u x^k \sum_{i=1}^{n} \frac{c_i}{x-r_i}\,\mathrm{d}x = \sum_{i=1}^{n} c_i r_i^k \left(\beta_{l/r_i}(1+k,0) - \beta_{u/r_i}(1+k,0)\right),\ 0 \le l < u \le 1$$
$$\int_l^u \frac{x^k}{x-r_i}\,\mathrm{d}x = -r_i^{k-1}\int_l^u \left(\frac{x}{r_i}\right)^k \left(1-\frac{x}{r_i}\right)^{-1}\mathrm{d}x = -r_i^k \int_{l/r_i}^{u/r_i} x^k(1-x)^{-1}\,\mathrm{d}x \quad \text{(A33)}$$
$$= r_i^k \left(\beta_{l/r_i}(1+k,0) - \beta_{u/r_i}(1+k,0)\right),\ r_i \neq 0$$
$$\int_l^u \frac{x^k}{x}\,\mathrm{d}x = \frac{1}{k}(u^k - l^k),\ r_i = 0$$

where $\beta_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$ is incomplete $\beta$-function

Characteristic function:

$$\phi_X(t) = \mathrm{E}\left[\mathrm{e}^{\mathrm{j}tX}\right] = \sum_{i=1}^{n} c_i \int_l^u \frac{\mathrm{e}^{\mathrm{j}tx}}{x-r_i}\,\mathrm{d}x = \sum_{i=1}^{n} c_i \mathrm{e}^{\mathrm{j}r_i t}(\Gamma(0,\mathrm{j}t(r_i-l)) - \Gamma(0,\mathrm{j}t(r_i-u))),\ r \in (l,u) \quad \text{(A34)}$$

## References


1. Phillips, G.M.; Taylor, P.J. "Best"Approximation. In *Theory and Applications of Numerical Analysis*; Academic Press: London, UK, 1996; pp. 86–130. [CrossRef]
2. Cheney, E.W. *Introduction to Approximation Theory*, 2nd ed.; AMS Chelsea Publishing: Providence, RI, USA, 1982.
3. Epperson, J. On the Runge example. *Ameican Math. Mon.* **1987**, *94*, 329–341. [CrossRef]
4. Smith, S.J. Lebesgue constants in polynomial interpolation. *Ann. Math. Inform.* **2006**, *33*, 109–123.
5. Ibrahimoglu, B.A. Lebesgue functions and Lebesgue constants in polynomial interpolation. *J. Inequalities Appl.* **2016**, *93*, 93. [CrossRef]

6. Freedman, D.; Diaconis, P. On the histogram as a density estimator:L2 theory. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1981**, *57*, 453–476. [CrossRef]
7. Munkhammar, J.; Mattsson, L.; Rydén, J. Polynomial probability distribution estimation using the method of moments. *PLoS ONE* **2017**, *12*, e0174573. [CrossRef]
8. Badinelli, R.D. Approximating probability density functions and their convolutions using orthogonal polynomials. *Eur. J. Oper. Res.* **1996**, *95*, 211–230. [CrossRef]
9. Abdous, B.; Bensaid, E. Multivariate local polynomial fitting for a probability distribution function and its partial derivatives. *J. Nonparametric Stat.* **2007**, *13*, 77–94. [CrossRef]
10. Gasca, M.; Sauer, T. Polynomial interpolation in several variables. *Adv. Comput. Math.* **2000**, *12*, 377–410. . [CrossRef]
11. Ghasemi, M.; Marshall, M. Lower bounds for a polynomial in terms of its coefficients. *Arch. Math.* **2010**, *95*, 343–353. [CrossRef]
12. Forsythe, G.E. Generation and Use of Orthogonal Polynomials for Data-Fitting with a Digital Computer. *J. Soc. Ind. Appl. Math.* **1957**, *5*, 74–88. [CrossRef]
13. Cunis, T. The pwpfit Toolbox for Polynomial and Piece-wise Polynomial Data Fitting. In Proceedings of the International Federation of Automatic Control, Baku, Azerbaijan, 13–15 September 2018; pp. 682–687. [CrossRef]
14. Hiang, T.S.; Ali, J.M. Quartic and quintic polynomial interpolation. In Proceedings of the AIP Conference Proceedings, Kuala Lumpur, Malaysia, 24–26 July 2013; Volume 1522, pp. 664–675. [CrossRef]
15. Gao, J.; Ji, W.; Zhang, L.; Shao, S.; Wang, Y.; Shi, F. Fast Piecewise Polynomial Fitting of Time-Series Data for Streaming Computing. *IEEE Access* **2020**, *8*, 43764–43775. [CrossRef]
16. Guo, L.; Narayan, A.; Zhou, T. Constructing Least-Squares Polynomial Approximations. *SIAM Rev.* **2020**, *62*, 483–508. [CrossRef]
17. Han, H.; Liu, H.; Ji, X. Interpolation to Data Points in Plane with Cubic Polynomial Precision. In Proceedings of the Technologies for E-Learning and Digital Entertainment, Hong Kong, China, 11–13 June 2007; pp. 677–686. [CrossRef]
18. Melucci, M. A brief survey on probability distribution approximation. *Comput. Sci. Rev.* **2019**, *33*, 91–97. [CrossRef]
19. He, W.; Hao, P.; Li, G. A novel approach for reliability analysis with correlated variables based on the concepts of entropy and polynomial chaos expansion. *Mech. Syst. Signal Process.* **2021**, *146*, 106980. [CrossRef]
20. Chen, D.; Yuan, Z.; Hua, G.; Zheng, N.; Wang, J. Similarity Learning on an Explicit Polynomial Kernel Feature Map for Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015. [CrossRef]
21. Cotter, N.E. The Stone-Weierstrass theorem and its application to neural networks. *IEEE Trans. Neural Netw.* **1990**, *1*, 290–295. [CrossRef]
22. Tong, Y.; Yu, L.; Li, S.; Liu, J.; Qin, H.; Li, W. Polynomial Fitting Algorithm Based on Neural Network. *ASP Trans. Pattern Recognit. Intell. Syst.* **2021**, *1*, 32–39. [CrossRef]
23. Barbeau, E.J. *Polynomials*; Springer: New York, NY, USA, 1989.
24. Rahman, Q.I.; Schmeisser, G. *Analytic Theory of Polynomials*; Oxford University Press: New York, NY, USA, 2002.
25. Apostol, T.M. *Mathematical Analysis*, 2nd ed.; Addison-Wesley: Reading, MA, USA, 1974.
26. Rahman, S. An extended polynomial dimensional decomposition method for arbitrary probability distributions. *J. Eng. Mech.* **2009**, *135*, 1439–1451. [CrossRef]
27. Funaro, D. *Polynomial Approximation of Differential Equations*; Springer: Berlin/Heidelberg, Germany, 1992.
28. Guo, B.Y.; Shen, J.; Wang, L.L. Generalized Jacobi Polynomials/Functions and Their Applications. *Appl. Numer. Math.* **2009**, *59*, 1011–1028. [CrossRef]
29. Boas, R.P.; Klamkin, M.S. Extrema of Polynomials. *Math. Mag.* **1977**, *50*, 75–78. [CrossRef]
30. Hanzon, B.; Jibetean, D. Global Minimization of a Multivariate Polynomial using Matrix Methods. *J. Glob. Optim.* **2003**, *27*, 1–23. [CrossRef]
31. Qi, L.; Teo, K.L. Multivariate Polynomial Minimization and Its Application in Signal Processing. *J. Glob. Optim.* **2003**, *26*, 419–433. [CrossRef]
32. Uteshev, A.Y.; Cherkasov, T.M. The Search for the Maximum of a Polynomial. *J. Symb. Comput.* **1998**, *25*, 587–618. [CrossRef]
33. Pan, V.Y. Solving A Polynomial Equation: Some History And Recent Progress. *SIAM Rev.* **1997**, *39*, 187–220. [CrossRef]
34. Beji, S. Polynomial Functions Composed of Terms with Non-Integer Powers. *Adv. Pure Math.* **2021**, *11*, 791–806. [CrossRef]
35. Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*, 4th ed.; McGraw-Hill: New York, NY, USA, 2002.
36. Alzaatreh, A.; Lee, C.; Famoye, F. A new method for generating families of continuous distributions. *Metron* **2013**, *71*, 63–79. [CrossRef]
37. Gorav, J.; Pandey, A.; Shukla, H.; Zisopoulos, C. How many zeros of a random sparse polynomial are real? In Proceedings of the International Symposium on Symbolic and Algebraic Computation, Hong Kong, China, 14–18 December 2020; pp. 273–280. [CrossRef]
38. Bini, D.A. Numerical computation of polynomial zeros by means of Aberth's method. *Numer. Algorithms* **1996**, *13*, 179–200. [CrossRef]
39. Lang, M.; Frenzel, B.C. Polynomial root finding. *IEEE Signal Process. Lett.* **1994**, *1*, 141–143. [CrossRef]
40. Maulud, D.; Abdulazeez, A.M. A Review on Linear Regression Comprehensive in Machine Learning. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 140–147. [CrossRef]
41. Seber, G.A.F. *A Matrix Handbook for Statisticians*; John Wiley & Sons: Hoboken, NJ, USA, 2008.

42. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
43. Aldaz, J.M. Self–Improvement Of The Inequality Between Arithmetic And Geometric Means. *J. Math. Inequalities* **2009**, *3*, 213–216. [CrossRef]
44. Mnatsakanov, R.M.; Hakobyan, A.S. Recovery of Distributions via Moments. In Proceedings of the Optimality: The Third Erich L. Lehmann Symposium, Houston, TX, USA, 16–19 May 2007; Volume 57, pp. 252–265. [CrossRef]
45. Trefethen, L.N. *Approximation Theory and Approximation Practice*; SIAM-Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2019.
46. Flecher, C.; Allard, D.; Naveau, P. Truncated skew-normal distributions: Moments, estimation by weighted moments and application to climatic data. *Int. J. Stat.* **2012**, *68*, 331–345. [CrossRef]
47. Morán-Vásquez, R.A.; Zarrazola, E.; Nagar, D.K. Some Statistical Aspects of the Truncated Multivariate Skew-t Distribution. *Mathematics* **2022**, *10*, 2793. [CrossRef]