RESEARCH ARTICLE

Polynomial probability distribution estimation using the method of moments

Joakim Munkhammar[1¤a]*, Lars Mattsson[2¤b], Jesper Rydén[3¤c]

1 BEESG, Department of Engineering Sciences/Uppsala University, Uppsala, Sweden, **2** Nordita, KTH Royal Institute of Technology and Stockholm University, Stockholm, Sweden, **3** Department of Mathematics, Uppsala University, Uppsala, Sweden

¤a Current address: Department of Engineering Sciences, Uppsala University, SE-751-21 Uppsala, Sweden
¤b Current address: Nordita, KTH Royal Institute of Technology and Stockholm University, SE-106 91 Stockholm, Sweden
¤c Current address: Department of Mathematics, Uppsala University, SE-751 06 Uppsala, Sweden
* joakim.munkhammar@angstrom.uu.se

Abstract

We suggest a procedure for estimating Nth degree polynomial approximations to unknown (or known) probability density functions (PDFs) based on N statistical moments from each distribution. The procedure is based on the method of moments and is setup algorithmically to aid applicability and to ensure rigor in use. In order to show applicability, polynomial PDF approximations are obtained for the distribution families Normal, Log-Normal, Weibull as well as for a bimodal Weibull distribution and a data set of anonymized household electricity use. The results are compared with results for traditional PDF series expansion methods of Gram–Charlier type. It is concluded that this procedure is a comparatively simple procedure that could be used when traditional distribution families are not applicable or when polynomial expansions of probability distributions might be considered useful approximations. In particular this approach is practical for calculating convolutions of distributions, since such operations become integrals of polynomial expressions. Finally, in order to show an advanced applicability of the method, it is shown to be useful for approximating solutions to the Smoluchowski equation.

Introduction

Estimating PDFs is essential in applied statistical analysis in many diverse fields of science, for instance for a few examples from our own experience in engineering [1, 2], physics [3, 4] and the Earth sciences [5, 6]. With knowledge of the PDF, further statistical problems can be tackled, such as for instance estimation of quantiles, and it can be used to construct stochastic models for various applications [1, 2].

In applied statistics, a distribution family is usually decided upon for a particular situation, and fitting a PDF then means determination of parameters by some estimation technique. Several common parametric methods such as the method of maximum likelihood and the method of moments are then well-known examples [7]. In case there is no prior knowledge of the

distribution family, other methods can be used, for instance kernel-density estimation (see e.g. [8]) or B-splines (see e.g [9], chapter 6). It is also possible to obtain approximation by means of series expansions, e.g. Gram–Charlier series and Edgeworth series [10, 11]. However, such expansions are often considered useful only in cases of moderate skewness [11]. Generally, the complexity of series approximation methods and the simplicity of traditional distribution fitting has inspired the use of series approximation methods in applied statistics, albeit to a limited level of use. See p.107-117 [12] for a review of approximation methods in statistics and [13] for a recent review of series approximation methods in statistics.

The present paper suggests an algorithm for estimating an N-order polynomial approximation of a PDF, based only on N moments of the PDF. Polynomial approximations for example distributions are given and a comparison with existing PDF series expansions is made. In addition to this, regarding more advanced problems, an example is used to illustrate that the procedure can be useful for approximating solutions to the Smoluchowski equation for such cases where moments of the solution are available.

Procedure

The method of moments estimates parameters of a predefined distribution f by equating moments of sample values and moments of the distribution, see p.467 [14]. This forms a system of equations which is often not analytically solvable, see p.467 [14], p.816 [15].

The procedure below estimates coefficients for a polynomial probability distribution approximation of the PDF f. This is done, via the method of moments, by equating the known moments of f with moments of the polynomial approximation of f.

The procedure is defined by the following algorithm:

1. Define a goodness-of-fit test $\mathcal{T}$ (for example a Kolmogorov-Smirnov test) for the PDF approximation.

2. Choose a real interval $[a', b']$ on which the approximation should hold good.

3. Choose a polynomial degree $N \in [0, 1, 2, 3, \ldots]$ for approximating a not necessarily known PDF f.

4. Ensure that nth order statistical moments for $n \in [0, \ldots, N]$ of f based on a real valued interval $[a, b]$ are available. ($[a, b]$ may equal $[a', b']$, but not necessarily)

5. As an approximation to f, define an Nth order polynomial P on the real interval $x \in [a', b']$:

$$P(N, x) \equiv w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \cdots + w_N x^N \tag{1}$$

where w_n, for each $n \in [0, \ldots, N]$, are unknown weights.

6. When solvable, the equation system resulting from equating the known ith moment of f on the interval $[a, b]$ with the unknown ith moment of the approximation P, for $i \in [0, \ldots, N]$ and interval $[a, b]$, is used to determine w_n for $n \in [0, \ldots, N]$. If not solvable, repeat from step 2 or 3 with increased (or decreased) N or changed $[a, b]$.

7. To qualify as a PDF, $P(N, x)$, defined on $[a', b']$, must be non-negative on $[a', b']$. If this is not met repeat from step 3 with increased (or decreased) N or changed $[a, b]$.

8. Also, to satisfy the conditions for a PDF on $[a', b']$, $P(N, x)$ should be normalized by $P(N, x) \rightarrow P(N, x) / \int_{a'}^{b'} P(N, x)\, \mathrm{d}x$ so that $\int_{a'}^{b'} P(N, x)\, \mathrm{d}x = 1$.

9. The procedure holds good if the polynomial approximation $P(N, x)$, for the predefined measure of goodness-of-fit $\mathcal{T}$, sufficiently approximates f on $[a', b']$. For improved fit,

repeat the procedure from step 3 with increased (or decreased) N or different choices of $[a, b]$.

In step 1 it is instructive to adopt a goodness-of-fit test such as for example a Kolmogorov-Smirnov test or Akaike's information criterion in order to ensure goodness-of-fit and applicability compared with other distributions, and set proper limits for pass and fail depending on application, see e.g. [16] and [17] for more information on that.

Given a defined goodness-of-fit test and a real interval $[a', b']$ for the distribution steps 3 and 4 certify the existence of N statistical moments. Consider a random variable X with probability density function $f(x)$, then the statistical moment of order k is defined by:

$$E_f[X^k] \equiv \int_a^b x^k f(x)\mathrm{d}x, \tag{2}$$

where a, b are real valued limits for the statistical moment. With only a data set $\{x_1, x_2, x_3, \ldots, x_M\}$ available, the sample moment of order k is given by:

$$E_f[X^k] \equiv \frac{1}{M}\sum_{i=1}^{M} x_i^k. \tag{3}$$

Although the distribution f is considered continuous, it is not a requirement of the procedure. However, step 5 in the algorithm is motivated by the use of a continuous distribution since the Weierstrass approximation theorem shows that a continuous function on an interval $[a, b]$ can be approximated arbitrarily good with a polynomial [18, 19]. With step 5 and Eq (2) it is possible to obtain the statistical moment of order n for the approximate polynomial $P(N, x)$:

$$E_P[X^k] = \int_a^b x^k P(N, x)\mathrm{d}x =$$
$$= \left(w_0 \frac{x^{k+1}}{k+1} + w_1 \frac{x^{k+2}}{k+2} + \cdots + w_N \frac{x^{k+N+1}}{k+N+1} \right) \Big|_a^b, \tag{4}$$

Via step 6 the moments $E_f[X^n]$ of f are equated with the moments $E_P[X^n]$ of the approximation P. With the polynomial approximation this can be setup in a linear equation system in w_n:

$$\mathbf{Mw} = \mathbf{E}, \tag{5}$$

where:

$$\mathbf{M} = \begin{bmatrix} b-a & \dfrac{b^2-a^2}{2} & \cdots & \dfrac{b^{N+1}-a^{N+1}}{N+1} \\ \dfrac{b^2-a^2}{2} & \dfrac{b^3-a^3}{3} & \cdots & \dfrac{b^{N+2}-a^{N+2}}{N+2} \\ \dfrac{b^3-a^3}{3} & \dfrac{b^4-a^4}{4} & \cdots & \dfrac{b^{N+3}-a^{N+3}}{N+3} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{b^N-a^N}{N} & \dfrac{b^{N+1}-a^{N+1}}{N+1} & \cdots & \dfrac{b^{N+N}-a^{N+N}}{2N} \end{bmatrix}, \; \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}, \; \mathbf{E} = \begin{bmatrix} E_f[X^0] \\ E_f[X^1] \\ E_f[X^2] \\ \vdots \\ E_f[X^N] \end{bmatrix}. \tag{6}$$

By inspection $\mathbf{M}$ is symmetric and a Hankel matrix [20]. The unknown weights w_n for $n \in [0, \ldots, N]$ constituting the vector $\mathbf{w}$ can now be obtained by computing the inverse of the

matrix $\mathbf{M}$:

$$\mathbf{w} = \mathbf{M}^{-1}\mathbf{E}. \tag{7}$$

It should be noted that since the matrix $\mathbf{M}$ is a Hankel matrix there exists particular algorithms for computing the inverse of the matrix as long as it is finite [21]. Also, the special case of using $a = 0$ and $b = 1$ makes $\mathbf{M}$ a Hilbert matrix, for which there exists an explicit expression for inversion, see [22]. Inserting $\mathbf{w}$ from Eq (7) in the polynomial approximation $P(N, x)$ in Eq (1) brings the polynomial approximation for f:

$$P(N, x) = \mathbf{w} \cdot [1 \; x \; x^2 \; x^3 \ldots \; x^N]. \tag{8}$$

The fact that $P(N, x)$ is supposed to approximate a PDF (Step 7) requires that $P(N, x)$ is non-negative on the interval $[a', b']$, which is not generally certified. As described in more detail in step 9 increased (or decreased) N or changed $[a, b]$ could solve this issue. When step 7 is checked, step 8 normalizes the PDF on $[a', b']$, and if $P(N, x)$ satisfies the goodness-of-fit test $\mathcal{T}$ (step 9), the procedure is finished.

Step 9 is important since there is no explicit check for convergence inside the procedure, a property not uncommon for series approximations in statistics, see p.114-115 [12]. Also, it is not unlikely that the set of moments might not uniquely describe a distribution, see [23].

With the computed weights w_n the CDF and the statistical moments of the polynomial distribution $P(N, x)$ is in turn possible to calculate analytically using Eq (4).

From a statistical point of view, an interesting property of a polynomial distribution approach is the convolution of distributions. The convolution of PDFs $g(x)$ and $h(x)$, both valid for some domain $x \in [c, d]$, is defined as:

$$(g * h)(x) = \int_c^d g(t)h(x - t)\,\mathrm{d}t. \tag{9}$$

Convolution of probability distributions is a complicated mathematical procedure for most applications. Also, only few distributions can be convoluted analytically and often this is instead handled numerically or approximated with e.g. the central limit theorem, such as in [2]. With the use of polynomial distributions $P(N, x) \equiv w_0 + w_1 x + w_2 x^2 + \cdots + w_N x^N$ and $Q(M, x) \equiv z_0 + z_1 x + z_2 x^2 + \cdots + z_M x^M$ the convolution Eq (9) in turn becomes a polynomial expression:

$$(P * Q)(x) = (v_0 + v_1 x + v_2 x^2 + \cdots + v_{N+M} x^{N+M})\big|_{a'}^{b'} \tag{10}$$

for some computable weights $v_0, v_1, \ldots, v_{N+M}$.

For data on the sum of correlated stochastic variables $X_1, X_2, \ldots$ it is possible to obtain a polynomial approximation of the resultant distribution by using moments from the sum of the stochastic variables. For such situations, which might be multimodal, and generally for more complex distributions, the polynomial estimation procedure simplifies the process which otherwise involves using complicated methods for determining parameters of mixture distributions. See [24] for information on the use of mixture distributions, and [1] for an applied example of mixture distributions in solar engineering.

Polynomial expansions of known probability distributions

In Fig 1, we present examples of the procedure applied to commonly used PDFs. For each of three distribution families (Normal, Weibull, Log-Normal), four parameter settings were considered. The moments were directly calculated from explicit expressions for each distribution.
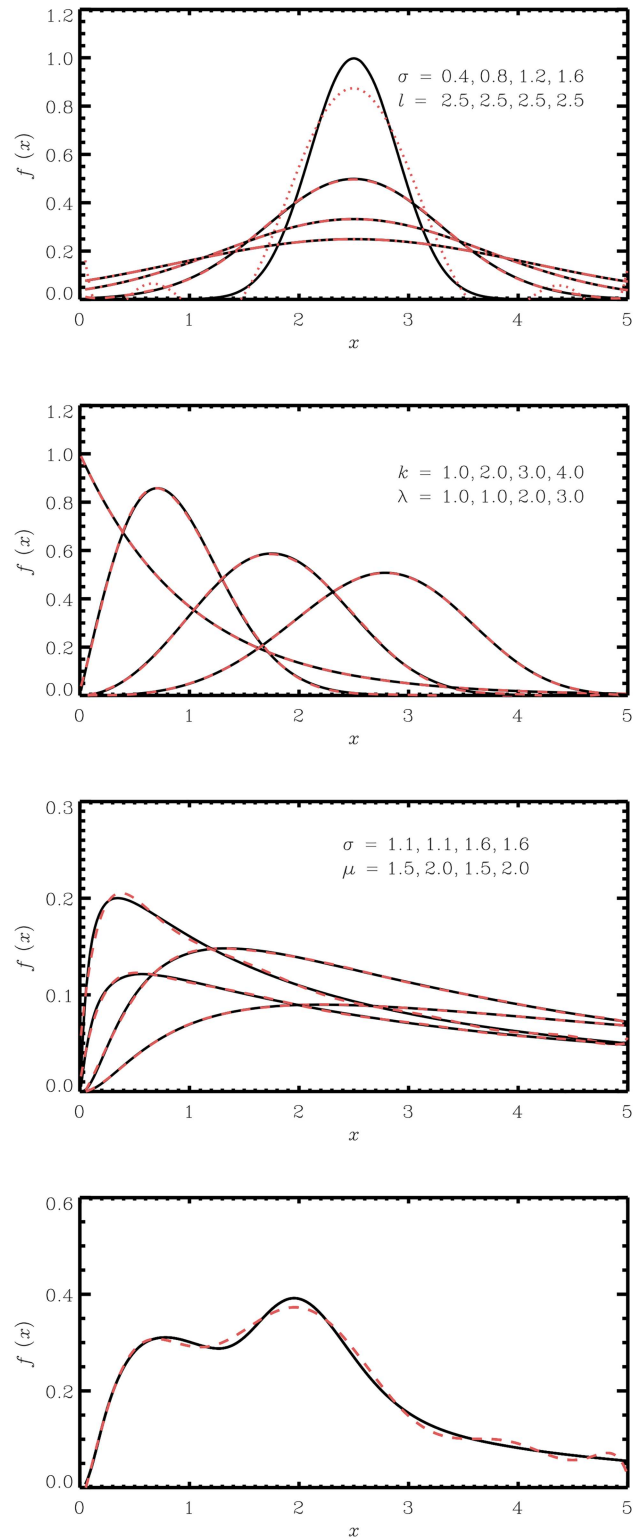
Fig 1. A set of distributions (black line) and polynomial approximations (dashed line) to level 10 for the common distributions: Normal, Weibull, Log-Normal, Bimodal. The polynomial distributions were computed from numerically integrated moments on the interval $a = 0$, $b = 5$.
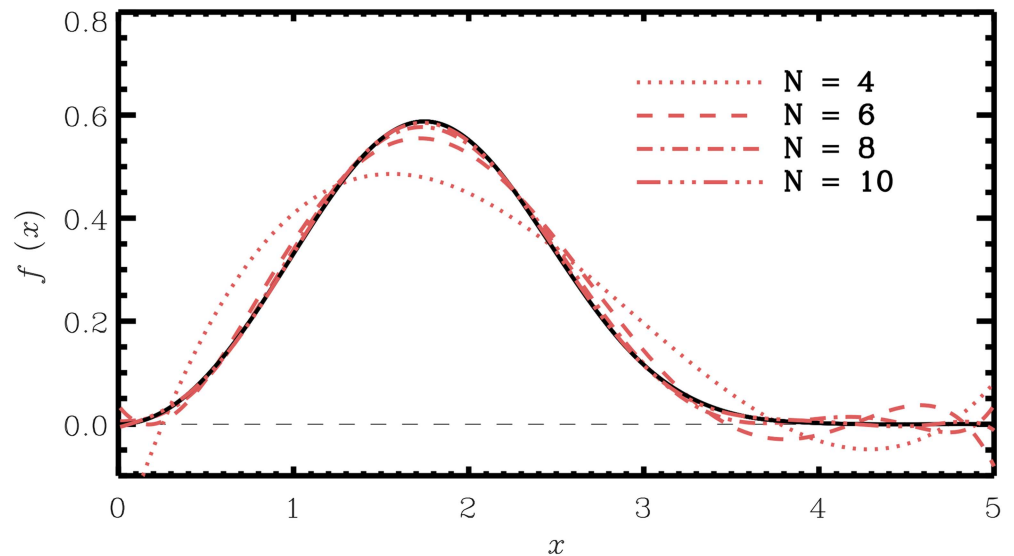
Fig 2. A Weibull distribution and polynomial distributions for certain number of terms *N*. The polynomial distributions were computed from numerically calculated moments on the interval $a = 0$, $b = 5$. In the polynomial distributions, note the wiggle of and negative values when the PDF is close to zero.

Moreover, order $N = 10$ of the polynomial was chosen to ensure a high level goodness-of-fit, however in principle any N terms can be used, with varying degree of goodness-of-fit as a result. By visual inspection of the plots, the agreement between the theoretical curve and the polynomial approximation seems satisfactory. Moreover, in Fig 1, we present an example with a bimodal distribution constructed from two Weibull distributions. Also here, the polynomial approximation adequately reproduces the distribution. Note that $[a', b']$ is not explicitly defined for the results presented in Fig 1, but could be used for at least the entire interval displayed in the plots (which equals $[a, b]$), in all cases except for the normal distribution with mean 2.5 and variance 0.4 which is negative for some parts of the interval. In that case changed $[a, b]$ or N, or if changed $[a', b']$ would be admissible (such as e.g. $a' = 2$ and $b' = 3$), could aid this issue.

The convergence rate with increasing order of polynomials is illustrated with an example distribution in Fig 2. In this example a two-parameter Weibull distribution with shape parameter $k = 3$ and scale parameter $\lambda = 2$ was chosen. Moreover, the values $N = 4, 6, 8, 10$ were chosen for the approximating polynomials. In this case, as N increases the goodness-of-fit increases. However, note the fluctuations and negative values of the polynomial approximation for different choices of N on the displayed interval. This shows that the possible choices of interval $[a', b']$, for which the polynomial approximation needs to be positive, depends on N.

However, note the fluctuations and negative values for the polynomial approximation for certain choices of $[a', b']$ for each choice of N.

Comparison with Gram–Charlier series approximations

A common type of polynomial-series expansion for PDFs is the Gram–Charlier type. In short, when the true PDF $f(x)$ of a random variable X is unknown, it is approximated with a PDF of the form:

$$g(x) = p_N(x)\phi(x) \tag{11}$$

where $\phi(x)$ is the PDF for a Normal distribution with zero mean and unit variance. The

polynomial $p_N(x)$ is then of the form:

$$p_N(x) = \sum_{i=1}^{N} c_i H_i(x) \tag{12}$$

where $H_i(x)$ represents Hermite polynomial of order i.

The expansions, in terms of the coefficients c_i, can be presented in various equivalent ways, using cumulants or moments ([11], Chapter 6). Introducing γ_1 and γ_2, the skewness and (excess) kurtosis (with μ_2, μ_3 and μ_4 being the second, third and fourth central moments, giving the relations $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2^2 - 3$.) of the distribution, we can write in a compact way, when X is a standardized random variable (zero mean and unit variance):

$$p_4(x) = 1 + \frac{\gamma_1}{6} H_3(x) + \frac{\gamma_2}{24} H_4(x). \tag{13}$$

This is referred to as Gram–Charlier type A. The related Edgeworth expansion involves one more Hermite polynomial, while keeping the number of parameters constant:

$$p_6(x) = 1 + \frac{\gamma_1}{6} H_3(x) + \frac{\gamma_2}{24} H_4(x) + \frac{\gamma_1^2}{72} H_6(x). \tag{14}$$

In the sequel, we use the former type, Eq (13), as will be commented upon shortly. Examples of successful modelling are found in various applications, see for instance [25, 26].

These expansions have a serious drawback, also apparent for the polynomial distribution: the approximation could result in negative values. Conditions on ensuring positive outcome were given by Barton and Dennis [27], where it was shown that for the Edgeworth expansion, Eq (14), the range for γ_1 and γ_2 over which positivity of the approximation is guaranteed is smaller than for the Gram–Charlier one [28].

Moreover, from a practical point of view, the series must have a finite number of terms. Use of higher-order terms does not necessarily guarantee a better result [11, 29]. For instance, in [11] a numerical example is given where actually four- and five-term series are worse than the three-term; each giving a negative frequency at a high value and a second mode at a low value (contrary to the data).

We consider cases with the Gram–Charlier expansions applied to some of the distributions and parameter settings presented in Fig 3. Note that these examples do not represent a situation with zero mean and unit variance; a transformation (and back transformation) had to be done to get the right scaling.

In Fig 3, we compare the cases of a two-parameter Weibull distribution (shape parameter $k = 1$, scale parameter $\lambda = 1$) and a Normal distribution with mean 2.5 and standard deviation 0.4. For each distribution, expansions with 3, 4 or 5 Hermite terms were considered. For the Weibull case, we note that for this particular parameter combination, the behavior at origin is difficult to capture for the various expansions and we get contributions of probability also at negative values, in addition to negative values of the density function itself. Note that the polynomial approximation proposed in the present paper approximates the true density more closely, cf. Fig 1. For the Normal distribution, we note that the overall mode behavior is found, but there are negative values in the resultant density.

Data application example

As an example of an application of the procedure, we consider a data set on measured household electricity use with ten-minute resolution for a detached house over one year, see [2, 30]. In power systems modeling it is conventional to fit a unimodal distribution family such as a
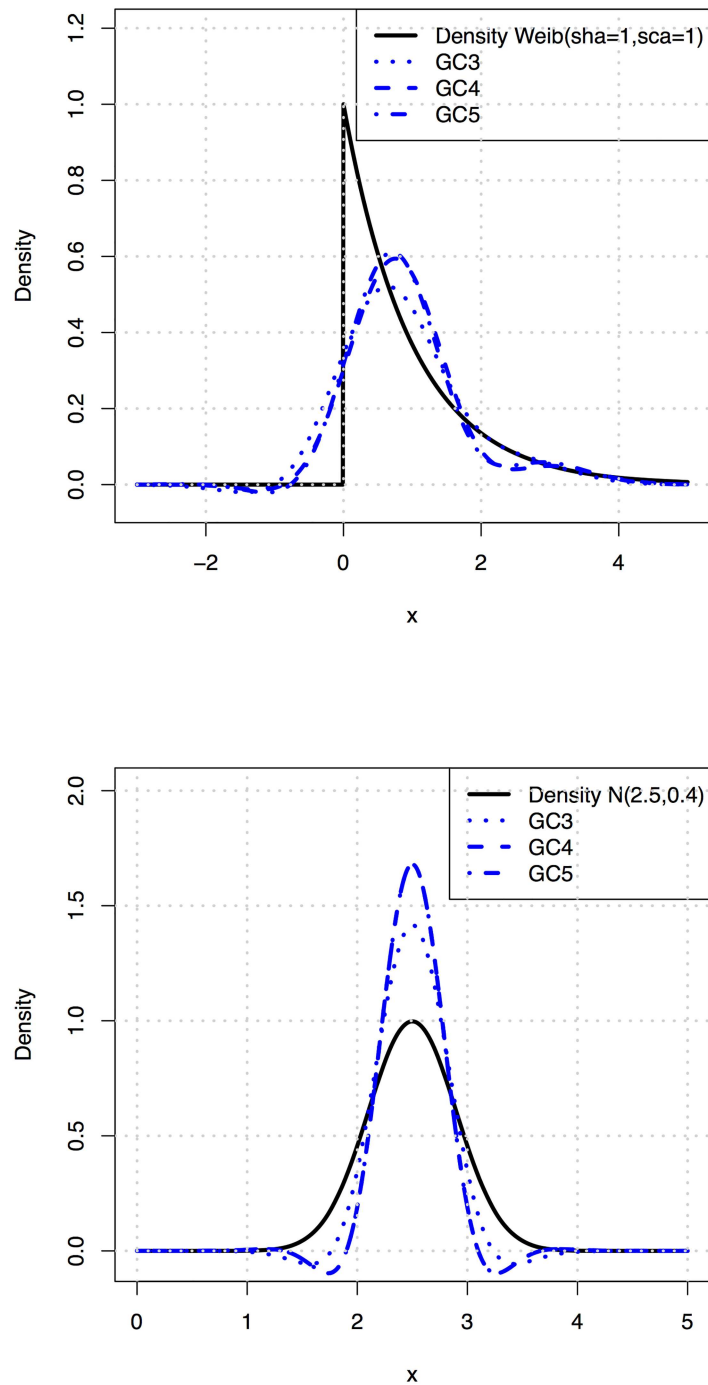
Fig 3. Examples of Gram-Charlier series expansions of a Weibull distribution (left) and a Normal distribution (right).

Log-Normal or Weibull distribution to this type of data, although the data set does not always correspond to an ideal unimodal distribution [2]. The data set is here fitted with Weibull, Gram–Charlier and polynomial distributions, and the resulting PDFs are shown in Fig 4. The moments for the data set were obtained using Eq (3). For the Gram–Charlier expansion,
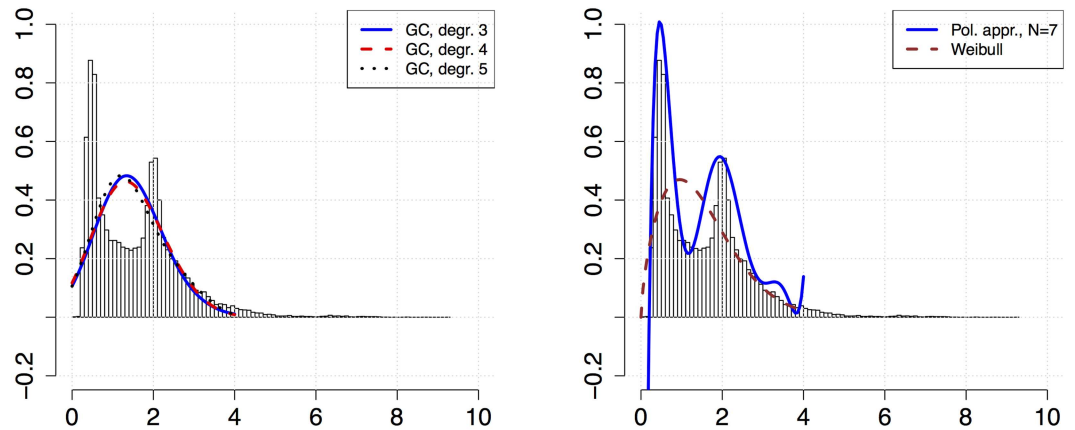
Fig 4. Histograms of household electricity use data with Gram–Charlier PDF (left), Weibull PDF (right) and polynomial PDF (right) fit. The polynomial distributions were computed from numerically calculated moments on the interval $a = 100$, $b = 5000$. A Weibull or Log-Normal PDF is usually a typical choice for this type of data [2].

moments were estimated from data by the routine `emm` from the R package `actuar`. The parameters of the Weibull distribution were estimated with the maximum likelihood estimate routine `wblfit` in Matlab. For the computation $a = 100$ and $b = 4000$ were used, which were also assumed as upper and lower limits for the polynomial approximation, as to represent reasonable lowest and highest power use for the household.

Based on the calculations we may conclude that the results from the Gram–Charlier series approach is similar to the Weibull distribution by displaying unimodal characteristics, albeit with a different location of the peak. For higher number of terms in the Gram–Charlier series approach the resulting PDF becomes negative in the interval and displays a rugged behavior. The polynomial distribution, on the other hand, captures both peaks of this particular data set. However, it is negative for values near zero, but this is not an issue as the valid interval $[a', b']$ of the polynomial distribution is chosen to be $[100, 4000]$, on which it is positive. The polynomial distribution also has some extra fluctuations between 3 kW and 4 kW, which is similar to Runge's phenomenon [31], and stems from the polynomial approximation characteristics of the approach, which is analogous to polynomial interpolation.

Integro-differential equation application: The Smoluchowski coagulation equation

The procedure may be used to approximate solutions to certain differential and integro-differential equations whose solutions are probability density distributions. The formal criteria are that the solution is a probability density distribution and that N moments for the solution are available.

One example of this is the Smoluchowski coagulation equation (SCE) [32], for which moments of the solution can be obtained for certain cases, and thus the moment transform can be used to approximate the resulting probability distribution. The Smoluchowski coagulation equation is [33]:

$$\frac{\partial f}{\partial t} = \frac{1}{2} \int_0^m K(m - m', m') f(m - m', t) f(m', t) \mathrm{d}m' -$$

$$- f(m, t) \int_0^\infty K(m, m') f(m', t) \mathrm{d}m', \tag{15}$$

which describes the process of a particle of mass m joining with another particle of mass m', given an initial mass distribution $f(m, 0)$ and a coagulation kernel $K(m, m')$. Here it is implicitly assumed that the kernel $K(m, m')$ is such that exact equations can be formulated for integer-order moments, although the equations may, in principle, include non-integer order moments.

With the moments M_n of $f(m, t)$ as previously defined, the moment equations corresponding to the SCE are

$$\frac{\mathrm{d}M_n}{\mathrm{d}t} = \frac{1}{2} \int_0^\infty \int_0^\infty K(m, m') f(m, t) f(m', t) \left[(m + m')^n - \right.$$
$$\left. m^n - (m')^n \right] \mathrm{d}m' \, \mathrm{d}m. \tag{16}$$

For kernels K which are expressible as a combination of powers of the particle masses m and m', the right hand side will only contain moments of $f(m, t)$. However, due to the term on the right-hand side containing the $(m + m')^n - m^n$ factor, it is only possible to construct equations for integer-order moments regardless of the kernel K.

The SCE has three known exact analytical solutions. These are for a constant ($K = K_0$), linear ($K \propto m + m'$) and product ($K \propto mm'$) kernel [34]. With $K(m, m') = \frac{1}{2}(m + m')$, $M_0(0) = M_1 = 1$ and an initial configuration corresponding to infinitesimally small initial clusters (note that $f(m, 0)$ has a singularity at $m = 0$), the solution of the SCE is [33, 34]:

$$f(m, t) = \frac{1}{\sqrt{2\pi}\, m^{3/2}} \exp\left(-\frac{m}{2}\, \mathrm{e}^{-2t} - t \right). \tag{17}$$

Multiplying with m^n and integrating over $m \in [0, a]$, an expression for the truncated moments $\tilde{M}_n$ is obtained,

$$\tilde{M}_n(a, t) = \int_0^a m^n f(m, t)\, \mathrm{d}m =$$
$$\frac{2^{n-1}\, \mathrm{e}^{2(n-1)\,t}}{\sqrt{\pi}} \left[\Gamma\left(n - \frac{1}{2}\right) - \Gamma\left(n - \frac{1}{2}, \frac{a}{2}\, \mathrm{e}^{-2t}\right) \right] \tag{18}$$

which provides the time evolution of any truncated moment of arbitrary order n. Using the above expression for $\tilde{M}_n$ a polynomial approximation of $f(m, t)$ should be possible to obtain for a given time t. Since an exact solution exists for this particular case (see Eq (17)), it is possible to test the reliability of the method as a way of obtaining approximate solutions to the SCE.

Before the results of this example are presented, a known but not yet fully understood weakness of the procedure outlined in this paper, is mentioned. Recreating power-laws is problematic and usually leads to the occurrence of "wiggles" in the polynomial approximation. This is analogous to Runge's phenomenon and probably related to the same [31]. However, if a power-law component can be identified, it can be removed before the transformation is made. In the case considered above, this would correspond to multiplication by $m^{3/2}$, before the moments $\tilde{M}_n$ are calculated. Then, dividing the resultant polynomial with $m^{3/2}$ will yield an approximation of $f(m, t)$ which does not suffer from Runge's phenomenon, except at the upper end of the mass scale where $f(m, t) \rightarrow 0$.

The results for polynomial approximation of the solutions for constant (A) and linear kernel (B) is shown in Fig 5. By visual inspection, the polynomial approximation is reasonable fit to the analytically calculated results. For time 0.04 in the constant kernel case and time 0.4 in the linear kernel case there is some deviation, which is likely due to Runge's phenomena. For other choices of mass-range and time steps the deviations could be dominant. This suggests that
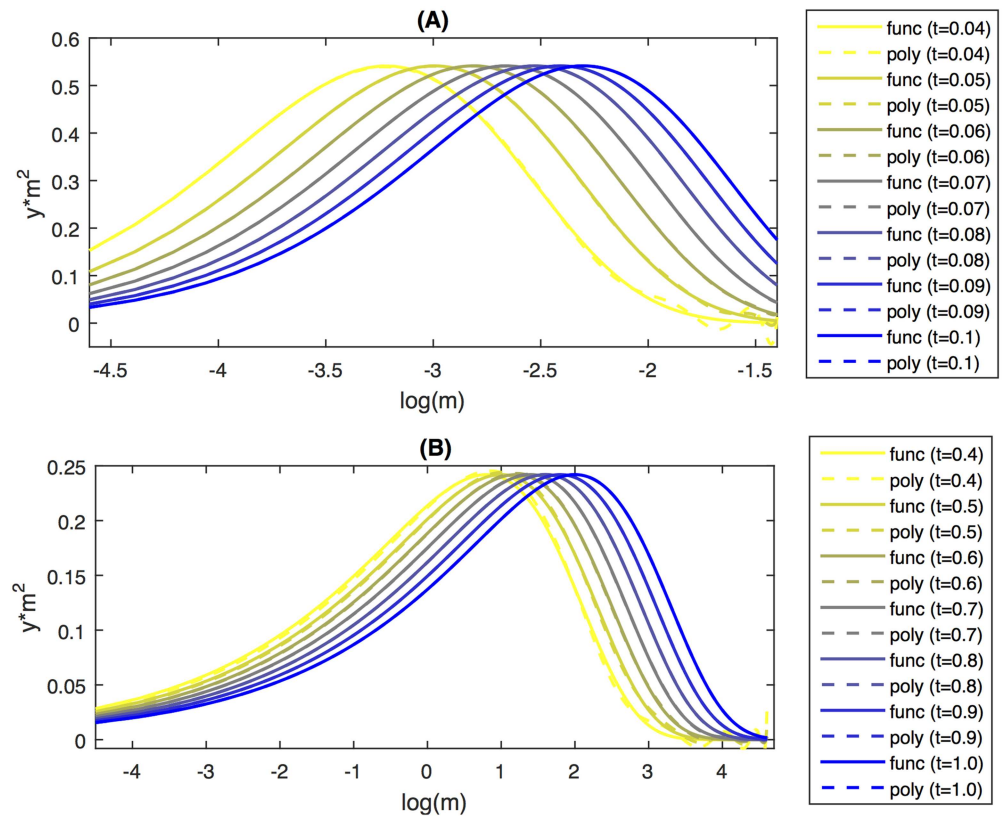
Fig 5. A plot of approximate solutions to the SCE for constant kernel (A) and for linear kernel (B). The approximation was based on the moment transform and polynomial degree $N = 8$.

developing an internal check for convergence in the procedure would be helpful, in particular when there are no analytical solutions or even numerical simulations available to compare with.

Practical issues

From computations with the polynomial distribution procedure it is possible to conclude two practical issues of particular interest:

1. Numerical computations of the inversion of **M** can become problematic as N increases.

2. When the PDF is close to zero there is a high tendency for mismatch.

Issue 1 arises since the Hankel matrix, and in particular the Hilbert matrix, are ill-conditioned [35]. This can be alleviated by using the algorithm for inverting Hankel matrices [21], or using the inversion formula for the Hilbert case of $a = 0$ and $b = 1$, when applicable [22]. The second issue has characteristics similar to Runge's phenomenon, see [31] for more information on that. This issue appears not to be easily resolved, it is instead instructive to use the algorithmic step 9 and adapt N and the interval $[a, b]$ for the moments, or the interval $[a', b']$ for the distribution if admissible, so that the mismatch can be minimized.

Discussion

We have suggested a simple procedure for estimating an N-order polynomial approximation to a known or unknown distribution based on N statistical moments of the distribution. The

procedure offers a possibility to use polynomial distributions as simple approximations of distributions which alleviates the necessity for identifying and defining distribution families or even particular parameters of interest. This could be particularly useful for approximating non-unimodal distributions such as the aforementioned examples of a bimodal distribution and the distribution of household electricity use.

The procedure is based on the perhaps simplest possible polynomial approximation. For a potentially better fit it is possible—without changing the framework of the procedure—to adopt alternative polynomial expansions, such as a Chebychev polynomial or a Hermite polynomial, like in the Gram-Charlier series expansion. Alternatively, expansions and moments that are not centered at the origin could be used. It is suitable to let the goodness-of-fit and applicability determine the choice of polynomial expansion. However, with alternative polynomial expansions there is no guarantee that the matrix remains Hankel type, or even that it is invertible. Another possibility for improved goodness-of-fit is to use different order (e.g., non integer-orders) of moments, since there is no requirement within the framework of the procedure that the moment orders must be consecutive natural positive integers, even if integer-order moments are most commonly used in statistics.

Since the polynomial approximation only has satisfactory goodness-of-fit within a limited interval $[a', b']$, the procedure is perhaps most suitable for applications which are defined on truncated probability distributions.

More advanced analysis of the convergence of the procedure is needed, in particular results which provide an explicit internal check for convergence.

Detailed investigations regarding the overall fitness of the procedure compared with other nonparametric probability density estimators such as kernel-density estimation and B-splines could be interesting to pursue for various types of distributions. For such investigations model complexity would be an interesting measure alongside approximation fitness, where perhaps the Akaike information criterion (AIC) could be a useful combined measure of this [17].

Generally, the applicability to differential or integro-differential equations is largely unknown, since only an example was given in this paper. More detailed investigations into this could prove fruitful. It is conjectured that this procedure can be generally useful for finding approximations to differential equation solutions when moments of the solution are available.

Supporting information

S1 Dataset. Data set on household electricity use used in this paper.
(TXT)

Acknowledgments

Author Contributions

Conceptualization: JM.

Formal analysis: JM LM JR.

Investigation: JM LM JR.

Project administration: JM.

Writing – original draft: JM LM JR.

References

1. Hollands KGT, Suehrcke H. A three-state model for the probability distribution of instantaneous solar radiation, with applications. Solar Energy 2013; 96: 103–112. https://doi.org/10.1016/j.solener.2013.07.007

2. Munkhammar J, Rydén J, Widén J. Characterizing probability density distributions for household electricity load profiles from high-resolution electricity use data. Applied Energy 2014; 135: 382–390. https://doi.org/10.1016/j.apenergy.2014.08.093

3. Chabrier G. Galactic Stellar and Substellar Initial Mass Function. Publications of the Astronomical Society of the Pacific 2013; 114: 763–795.

4. Mattsson L, Höfner S. Dust-driven mass loss from carbon stars as a function of stellar parameters. Astronomy & Astrophysics 2011; 533, A42. https://doi.org/10.1051/0004-6361/201015572

5. Monahan AH. Empirical Models of the probability distribution of sea surface wind speeds. J. of Climate 2007; 20: 5798–5814. https://doi.org/10.1175/2007JCLI1609.1

6. Ochi MK. Ocean Waves: The Stochastic Approach. Cambridge University Press; 1991.

7. Casella G, Berger RL. Statistical Inference. Duxbury press 2nd ed; 2002.

8. Wand MP, Jones MC. Kernel Smoothing. Chapman & Hall CRC; 1995.

9. Gallier J. Curves and Surfaces In Geometric Modeling: Theory And Algorithms. Morgan Kaufmann; 1st edition; 1999.

10. DasGupta A. Asymptotic Theory of Statistics and Probability. Springer-Verlag; 2008.

11. Kendall M, Stuart A, Ord JK. Kendall's Advanced Theory of Statistics. 5th ed. Vol. 1., Griffin, London; 1987.

12. Kotz S, Johnson NL. Encyclopedia of Statistical Sciences: Volume 1. John Wiley & Sons; 1982.

13. Kolassa JE. Series Approximation Methods in Statistics. Springer lecture notes in Statistics, third edition; 2006.

14. Kotz S, Johnson NL. Encyclopedia of Statistical Sciences: Volume 5, John Wiley & Sons; 1985.

15. Lovric M. International Encyclopedia of Statistical Science. Springer-Verlag Berlin Heidelberg; 2011.

16. Conover WJ. Practical Nonparametric Statistics, 3rd ed. Wiley; 1999.

17. Akaike H. A new look at the statistical model identification. Transactions on Automatic Control 1974; 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705

18. Jeffreys H, Jeffreys BS. Weierstrass's Theorem on Approximation by Polynomials, Mathematical Physics, 3rd ed. Cambridge, England: Cambridge University Press 1988; 446–448.

19. Weierstrass K. Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin; 1885.

20. Partington JR. An introduction to Hankel operators, LMS Student Texts 13, Cambridge University Press; 1988.

21. Trench WF. An algorithm for the inversion of finite Hankel matrices. J. Soc. Indust. Appl. Math. 1965; 13: 1102–1105. https://doi.org/10.1137/0113078

22. Choi MD. Tricks or Treats with the Hilbert Matrix, The Am. Math. Montly. 1983; 90: 301–312. https://doi.org/10.2307/2975779

23. Shohat JA, Tamarkin JD. The Problem of Moments, Mathematical Surveys and Monographs 1, American mathematical society: New York; 1943.

24. Titterington DM, Smith AFM, Makov UE. Statistical Analysis of Finite Mixture Distributions. Wiley, Chapter 3.3; 1985.

25. Mita K, Akataki K, Miyagawa T, Koyama K, Ishida N. Statistical approach for the estimation of daily physical activity levels using the probability density function of heart rate, Journal of Biomedical Engineering 1989; 11: 315–319. https://doi.org/10.1016/0141-5425(89)90066-6 PMID: 2755112

26. Yim JZ, Chou CR, Huang WP. Study on the distributions of the measured fluctuating wind velocity components, Atmospheric Environment 2000; 34: 1583–1590. https://doi.org/10.1016/S1352-2310(99)00414-8

27. Barton DE, Dennis KER. The conditions under which Gram–Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 1952; 39: 425–427. https://doi.org/10.2307/2334037

28. Jondeau E, Rockinger M. Gram–Charlier densities, J. of Economic Dynamics and Control 2001; 25: 1457–1483. https://doi.org/10.1016/S0165-1889(99)00082-2

29. Ochi MK. Applied Probability and Stochastic Processes in Engineering and Physical Sciences. Wiley; 1992.

30. Zimmermann JP. End-use metering campaign in 400 households in Sweden: Assessment of the potential electricity savings, Report, Swedish Energy Agency; 2009.

31. Runge C. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, Zeitschrift fur Mathematik und Physik 1901; 246: 224–243.

32. Smoluchowski M. Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen, Physik. Z. 1916; 17: 557–571.

33. Golovin A. The solution of the coagulating equation for cloud droplets in a rising air current. zv. Geophys. 1963; 5: 482–487.

34. Aldous D. J. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli 1999; 5: 3–48. https://doi.org/10.2307/3318611

35. Fasino D. Spectral properties of Hankel matrices and numerical solutions of finite moment problems, Journal of Computational and Applied Mathematics 1995; 65: 145–155. https://doi.org/10.1016/0377-0427(95)00108-5