


Knowable Moments in Stochastics: Knowing Their Advantages

Demetris Koutsoyiannis 

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, 15780 Zographou, Greece; dk@itia.ntua.gr

Abstract: Knowable moments, abbreviated as K-moments, are redefined as expectations of maxima or minima of a number of stochastic variables that are a sample of the variable of interest. The new definition enables applicability of the concept to any type of variable, continuous or discrete, and generalization for transformations thereof. While K-moments share some characteristics with classical and other moments, as well as with order statistics, they also have some unique features, which make them useful in relevant applications. These include the fact that they are knowable, i.e., reliably estimated from a sample for high orders. Moreover, unlike other moment types, K-moment values can be assigned values of distribution function by making optimal use of the entire dataset. In addition, K-moments offer the unique advantage of considering the estimation bias when the data are not an independent sample but a time series from a process with dependence. Both for samples and time series, the K-moment concept offers a strategy of model fitting, including its visualization, that is not shared with other methods. This enables utilization of the highest possible moment orders, which are particularly useful in modelling extremes that are closely associated with high-order moments.

Keywords: knowable moments; distribution function; probability density function; stochastic variables; extremes; estimation; bias; long-range dependence

MSC: 60G07; 60G70; 62A09; 62E17; 62G05; 62G07



Citation: Koutsoyiannis, D. Knowable Moments in Stochastics: Knowing Their Advantages. *Axioms* **2023**, *12*, 590. <https://doi.org/10.3390/axioms12060590>

Academic Editor: Palle E.T. Jorgensen

Received: 9 May 2023

Revised: 5 June 2023

Accepted: 10 June 2023

Published: 14 June 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Knowable moments, abbreviated as K-moments, were introduced by Koutsoyiannis in 2019 [1] and further explored by Koutsoyiannis (2022) [2,3]. They were applied to research [4–6] and engineering studies [7–9]. However, all these studies are authored or coauthored by Koutsoyiannis, who introduced them. Possible explanations for the reluctance of other researchers to use them are (a) that they are not useful or (b) they are not a new tool, as for example asserted by an anonymous negative reviewer of an initial submission of the already mentioned study [7], who stated that “the proposed ‘K-moments’ methods [. . .] is a renaming of the widely known and widely applied ‘Probability-Weighted Moments’ (PWM) and their combinations known as ‘L-moments’” [10,11]. Another explanation (c) is that their definition and their advantages are not paid attention to or are not well understood.

Conjecturing that the latter explanation (c) is correct, here we extend and clarify the meaning and features of K-moments, emphasizing their advantages, particularly those that are unique among all moment types. We also provide algorithmic details for the application of the framework. More specifically, we provide the following information.

1. We redefine K-moments (Section 2.1) as expectations of maxima or minima of a number of stochastic variables that are independent, identically distributed copies of the stochastic variable of interest. The new definition is clearer, more rigorous and more intuitive. In most cases of interest (but not in all), the new definition does not imply changes in the computational framework (Section 2.3).
2. We provide a more intuitive explanation of the statistical estimators of K-moments, which are adapted to the new definition, again without implying computational differences in

- most of the cases. Furthermore, we provide techniques to accelerate the calculations when the datasets are large, e.g., with million or billion values (Section 2.2).
3. We extend the scope of K-moments from continuous stochastic variables to also include discrete variables (Section 2.4).
 4. We discuss cases of generalization where the new definition implies some differences, notational and computational, to the existing one (Section 2.5).
 5. We discuss the advantages of the K-moment framework, starting from the information it provides about what a classical moment estimator determines, which actually is not the true value of that classical moment (Section 3.1); this is the reason why classical moments are unknowable from samples, for orders beyond 3–4.
 6. We show that K-moments for moment orders up to 4 replace the information contained in classical moments and L-moments, such as summary statistics (Section 3.2).
 7. We show that the real power of the K-moment framework is its ability to estimate moments reliably and unbiasedly for very high orders up to the sample size, even if this is several million or more. We also show the ability of the framework to readily assign, in a simple manner, a value of the distribution function to each K-moment—a property not shared by any other type of moments (Section 3.3).
 8. Exploiting the above features, we show how K-moments provide a sound and flexible framework for model fitting, making optimal use of the entire dataset, rather than relying on a few moments (as in classical moments and L-moments) or assigning probabilities to single data values (order statistics). This enables utilization of the highest possible moment orders, which are particularly useful in modelling extreme highs or lows that are closely associated with high-order moments. The model fitting concept with K-moments is a new strategy not shared with any other types of methods of moments and includes visualization of the goodness of fit (Section 3.4).
 9. We illustrate that K-moment estimators offer the ability to estimate the probability density function from a sample—a unique feature of the K-moment methodology.
 10. We show that K-moments offer the unique advantage of taking into account the estimation bias when the data are not an independent sample but a time series from a process with dependence, even of long range (Section 3.6).
 11. We provide algorithmic details of the computational framework of K-moments (particularly in Sections 2.2, 3.4 and 3.6)
 12. Finally, we discuss conceptual similarities and differences with other moment types (Section 4—Discussion) and summarize the overall conclusions (Section 5).

With respect to point 8, the K-moment framework offers the possibility of modelling extremes using merely the parent distribution, without reference to asymptotic Extreme Value distributions. This is a more reliable choice as it has been known that the convergence to the asymptotic Extreme Value distributions can be extraordinarily slow [2], while the non-asymptotic distributions of extremes can be quite difficult to determine, particularly when there is time dependence. Yet there is a theoretical connection of the K-moment framework with the asymptotic Extreme Value theory, related to assigning values of the distribution function to K-moments (point 7) at the distribution tails, as will be specified in Section 3.3.

Points 1–4, 6, and 12 are the original contributions of this work, while the remaining points provide a review of existing developments of the K-moment framework, with appropriate adaptations to the new definitions and with the aim to highlight its advantages and usefulness, which, as already mentioned, is the scope of the paper. In addition, all illustrations, including graphs and tables, and the appendix are original.

2. Definitions and Main Derivations

2.1. Definition and Meaning

Let \underline{x} be a stochastic variable with a distribution function $F(x)$ and a tail function $\bar{F}(x) := 1 - F(x)$. Notice that we have adopted the Dutch (van Dantzig–Hemelrijk [12]) convention of underlining stochastic variables while common variables are not underlined.

This convention is much clearer and more rigorous than its alternatives, which either (a) do not distinguish the stochastic from common variables or (b) use upper-case letters for stochastic variables and lower-case ones for common variables. Alternative (a), while being popular, particularly in Bayesian statistics texts, is ambiguous and can be misleading (see several examples in [2]). Alternative (b) is too restrictive (it prohibits using lower-case letters for stochastic variables), ambiguous (when the standard notation of a physical quantity is in upper case, it could be misinterpreted as a stochastic variable, while it is not), or both (for example, when using Greek letters, e.g., the Latin x and the Greek χ have the same appearance in upper case, i.e., X ; see additional remarks in [2]).

If \underline{x} is of continuous type, we define its probability density function as $f(x) := dF(x)/dx$. If it is of discrete type, taking on the values $x_j, j = 0, 1, \dots, J$, where possibly $J = \infty$, we define its probability mass function as $P_j := P\{\underline{x} = x_j\} = F(x_j) - F(x_{j-1})$.

We consider a sample of \underline{x} , i.e., a number p of independent copies of the stochastic variable \underline{x} , i.e., $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$. If we arrange the variables in ascending order, the i th smallest, denoted as $\underline{x}_{(i:p)}, i = 1, \dots, p$ is termed the i th order statistic. The largest (p th) order statistic is:

$$\underline{x}_{(p)} := \underline{x}_{(p:p)} = \max(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p) \tag{1}$$

and the smallest (first) is

$$\underline{x}_{(1:p)} = \min(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p) \tag{2}$$

Now we define the K-moments in terms of expectations of these variables, denoted as $E[\]$, in the following manner.

Definition 1. The expectation of the largest of the p variables $\underline{x}_{(p)}$:

$$K'_p := E[\underline{x}_{(p)}] = E[\max(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)] \tag{3}$$

is called the upper knowable moment (K-moment) of order p .

Definition 2. The expectation of the smallest of the p variables $\underline{x}_{(1:p)}$:

$$\overline{K}'_p := E[\underline{x}_{(1:p)}] = E[\min(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)] \tag{4}$$

is called the lower knowable moment (K-moment) of order p .

Furthermore, we generalize the definition to transformations $g(\underline{x})$ of the stochastic variable of interest \underline{x} . Thus, by setting $g(\underline{x}) = \underline{x}^q$, where q is an integer, we obtain:

Definition 3. The expectation:

$$K'_{pq} := E[\underline{x}_{(p)}^q] = E[\max(\underline{x}_1^q, \underline{x}_2^q, \dots, \underline{x}_p^q)] \tag{5}$$

is called the upper K-moment of orders p, q .

Definition 4. The expectation:

$$\overline{K}'_{pq} := E[\underline{x}_{(1:p)}^q] = E[\min(\underline{x}_1^q, \underline{x}_2^q, \dots, \underline{x}_p^q)] \tag{6}$$

is called the lower K-moment of orders p, q .

The above notation implies that we omit the subscript q when $q = 1$. Likewise, we can define central K-moments by setting $g(\underline{x}) = (\underline{x} - \mu)^q$.

Definition 5. *The expectation:*

$$K_{pq} := E \left[\left(x_{(p)} - \mu \right)^q \right] = E \left[\max \left((x_1 - \mu)^q, (x_2 - \mu)^q, \dots, (x_p - \mu)^q \right) \right] \tag{7}$$

is called the upper central K-moment of orders p, q .

Definition 6. *The expectation:*

$$\bar{K}_{pq} := E \left[\left(x_{(1:p)} - \mu \right)^q \right] = E \left[\min \left((x_1 - \mu)^q, (x_2 - \mu)^q, \dots, (x_p - \mu)^q \right) \right] \tag{8}$$

is called the lower central K-moment of orders p, q .

As we will see below, K-moments can alternatively be written as expectations of products of x^q (or $(x - \mu)^q$), the distribution function or tail function raised to the power $p - 1$, and some adjustment factors ($A(x, p)$ or $\bar{A}(x, p)$). Specifically, they are expressed as

$$\begin{aligned} K'_{pq} &= pE \left[A(x, p) F(x)^{p-1} x^q \right], & \bar{K}'_{pq} &= pE \left[\bar{A}(x, p) \bar{F}(x)^{p-1} x^q \right] \\ K_{pq} &= pE \left[A(x, p) F(x)^{p-1} (x - \mu)^q \right], & \bar{K}_{pq} &= pE \left[\bar{A}(x, p) \bar{F}(x)^{p-1} (x - \mu)^q \right] \end{aligned} \tag{9}$$

where, in the most common cases, the adjustment factors $A(x, p)$ and $\bar{A}(x, p)$ can be omitted as $A(x, p) = \bar{A}(x, p) = 1$. However, in some of the cases, they can take different values, as will be specified below.

Equation (9) justifies their name *moments*. As will be seen, for $p = 1$, $A(x, 1) = 1$ in all cases. Therefore, by setting $p = 1$, we recover the classical noncentral (raw) moments, i.e.,

$$K'_{1q} = \bar{K}'_{1q} = E[x^q] =: \mu'_q, \quad K_{1q} = E[(x - \mu)^q] =: \mu_q \tag{10}$$

However, whenever we have only an observed sample, and we do not know $F(x)$ from theoretical reasoning, the classical moments are unknowable, except if q is very low, as thoroughly justified [1,2]. On the other hand, if we choose a low q , namely $q = 1, 2$ (and occasionally 3), then we can estimate in a reliable manner a moment of total order $p + q - 1$, which can be made very high by choosing a high p . This enables knowing the high-order properties of a distribution from a sample, which justifies the name *knowable moments*. In the next section, we will see that the estimation of the K-moments can be easily made using order statistics.

While the K-moment framework allows for studying transformations of x , the simplest case of the untransformed variable suffices for most statistical tasks. In this case, Equation (9) is simplified to

$$\begin{aligned} K'_p &= pE \left[A(x, p) F(x)^{p-1} x \right], & \bar{K}'_p &= pE \left[\bar{A}(x, p) \bar{F}(x)^{p-1} x \right] \\ K_p &= K'_p - \mu, & \bar{K}_p &= \bar{K}'_p - \mu \end{aligned} \tag{11}$$

where the last two relationships that give the central moments are a direct consequence of the fact that $\max(x_1 - \mu, x_2 - \mu, \dots, x_p - \mu) = \max(x_1, x_2, \dots, x_p) - \mu$ (and likewise for the minimum).

2.2. Estimation

We assume that we have a sample of size n arranged in ascending order, $x_{(1:n)} \leq x_{(2:n)} \leq \dots \leq x_{(n:n)}$. We chose $p \leq n$ of them at random and we wish to find the probability that a specific one, say $x_{(i:n)}$, is the maximum of all p . Apparently, if $i < p$ then this probability is zero. If $i \geq p$, then the total number of combinations in which the $x_{(i:n)}$ is the maximum among the p variables equals the number of ways of choosing

the $p - 1$ remaining variables among the possible $i - 1$ of them, i.e., $\binom{i - 1}{p - 1}$. The total number of ways to choose any p variables out of n is $\binom{n}{p}$. Thus, the sought probability, call it b_{inp} , is

$$b_{inp} = \binom{i - 1}{p - 1} / \binom{n}{p} \tag{12}$$

Note that

$$\sum_{i=p}^n \binom{i - 1}{p - 1} = \binom{n}{p} \tag{13}$$

and hence

$$\sum_{i=p}^n b_{inp} = 1 \tag{14}$$

as expected.

Given that the probability of $x_{(i:n)}$ being the maximum of the p out of n variables is b_{inp} , the estimator of the expectation of the K-moment K'_p will be

$$\hat{K}'_p = \sum_{i=1}^n b_{inp} x_{(i:n)} \tag{15}$$

To find the estimator of the lower K-moments, it suffices to reverse the order of the sample, i.e., to replace $x_{(i:n)}$ with $x_{(n-i+1:n)}$. Hence,

$$\hat{K}'_p = \sum_{i=1}^n b_{inp} x_{(n-i+1:n)} = \sum_{i=1}^n b_{n-i+1,n,p} x_{(i:n)} \tag{16}$$

If we use all integer values of p from 1 to n , from the ordered sample $x_{(i:n)}$ we can calculate n upper moments \hat{K}'_p and n lower moments \hat{K}'_p , a total of $2n$ quantities. However, the information contained in each of the sequences \hat{K}'_p and \hat{K}'_p is equivalent to each other:

Proposition 1. *The equivalence relationships of upper and lower moment estimators are:*

$$\hat{K}'_p = - \sum_{i=1}^p (-1)^i \binom{p}{i} \hat{K}'_i, \quad \hat{K}'_p = - \sum_{i=1}^p (-1)^i \binom{p}{i} \hat{K}'_i \tag{17}$$

The proof is given in [2].

The above formulation is for integer moment order p . We can readily generalize for real p by replacing factorials with the Gamma function. After algebraic manipulations, we obtain the final expression:

$$b_{inp} = \begin{cases} 0, & i < p \\ p \frac{\Gamma(n-p+1)}{\Gamma(n+1)} \frac{\Gamma(i)}{\Gamma(i-p+1)}, & i \geq p \geq 0 \end{cases} \tag{18}$$

In the algorithmic evaluation of Equation (18), it is suggested to first calculate the logarithms of the gamma functions and add them (or subtract, as appropriate), and then exponentiate the sum.

For $p = 1$, Equation (18) results in $b_{in1} = 1/n$ and thus we recover the estimate of the mean, $\hat{K}'_1 = \hat{K}'_1 = \hat{\mu}$. For $p = n$, Equation (18) results in $b_{nnn} = 1$ with all other $b_{inn} = 0$, and thus only the maximum (or the minimum) value of the sample is taken into account in the estimation. For $p > n$, estimation becomes impossible. Thus, the method permits the

estimation of moment orders as high as the sample size, n . As p increases from 1 to n , fewer sample items are used in the estimation until it remains only one for $p = n$.

The following result is important to note:

Proposition 2. *The estimators (15)–(16) with b_{inp} determined from (18), are unbiased.*

The proof is given in [2]. By replacing \underline{x} with a transformation $g(\underline{x})$ we readily obtain an estimator of transformed K-moments but unbiasedness cannot be assured in all cases.

The calculations to estimate K-moments are easy and follow the steps listed below, which are also illustrated in a spreadsheet that accompanies the paper as Supplementary Information:

1. We sort the sample x_j into ascending order, i.e., we designate $x_{(i:n)}, i = 1, \dots, n$.
2. For a specific moment order p , we employ Equation (18) to find b_{inp} for $i = 1, \dots, n$.
3. From Equation (15), we calculate \hat{K}'_p .
4. From Equation (16) we calculate $\hat{\bar{K}}'_p$.
5. We repeat steps 2–4 for all required orders p .

Often the data include ties either because the stochastic variable of interest is of discrete type or because they are too many and are summarized in more manageable form. In this case, the calculations can be simplified by applying a single coefficient to each value appearing in the sample. Assuming that a certain value x_j appears j times in the sample, namely from positions j_1 to j_2 in the ordered sample, i.e., $x_{(j_1:n)} = \dots = x_{(j_2:n)} = x_j$, with $j = j_2 - j_1 + 1$, the value x_j should be multiplied by the sum $\sum_{i=j_1}^{j_2} b_{inp}$. This sum is easy to calculate analytically, resulting in a concise expression:

$$b_{(j_1, j_2), n, p} := \sum_{i=j_1}^{j_2} b_{inp} = \frac{j_2}{p} b_{j_2 np} - \frac{j_1 - p}{p} b_{j_1 np} \tag{19}$$

It is easy to verify that for $j_2 = j_1$, the result is $b_{(j_1, j_2), n, p} = b_{j_1 np}$ as it should. Moreover, if $j_2 = j_1 - 1$, which means that there is no appearance of a particular value ($j = 0$), then the result is $b_{(j_1, j_2), n, p} = 0$, as required.

An illustration of the good performance of the estimation tactic provided by Equation (19) follows. A sample of 10,000 values was generated from the generalized Pareto distribution with mean $\mu = 1$ and tail index $\zeta = 0.1$ (so that its scale parameter is $\lambda = 0.9$). Its distribution function is shown in Table 1, along with the expressions of the classical moments, and upper and lower K-moments, all of which are closed analytical expressions (see [2]). These expressions are evaluated for the above parameter values for order p from 1 to 10,000, and the results are plotted in Figure 1. Note that for better legibility, the quantities $\mu'_p{}^{1/p}$ were plotted for the classical moments, and that these diverge to infinity for $p \geq 1/\zeta = 10$. The classical moments were also estimated by the standard statistical estimators; naturally, these do not diverge for any p but rather for the maximum value $p = 10,000$ the quantity $\hat{\mu}'_p{}^{1/p}$ equals the maximum sample value, which happens to be 15.4. As a result, the estimates start to depart for the theoretical values even for low p (>4) and the difference is infinite for $p \geq 10$.

Table 1. Equations of the moments of the generalized Pareto distribution (adapted from [2])¹.

Characteristic	Equation	Equation No.
Distribution function	$F(x) = 1 - (1 + \zeta x / \lambda)^{-1/\zeta}$	(20)
Classical noncentral moment	$\mu'_p := E[x^p] = p(\lambda/\zeta)^p B(p, 1/\zeta - p)$	(21)
Upper K-moment	$K'_p := E[x_{(p)}] = (\lambda/\zeta)(pB(p, 1 - \zeta) - 1)$	(22)
Lower K-moment	$\bar{K}'_p := E[x_{(1:p)}] = \lambda/(p - \zeta)$	(23)

¹ $B(a, b)$ is the beta function.

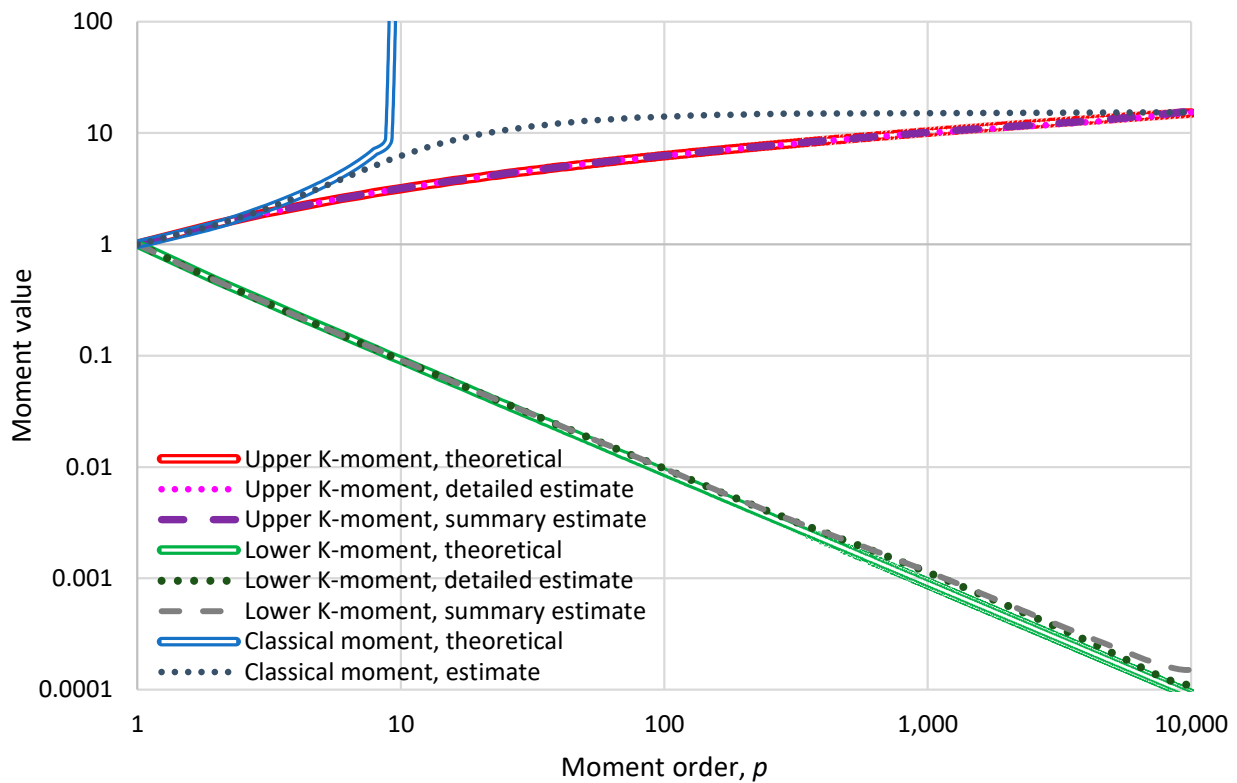


Figure 1. Comparison of theoretical moments of the Pareto distribution with parameters $\zeta = 0.1$ and $\lambda = 0.9$ with their estimates from a synthetic sample of 10,000 values, detailed (based on Equations (15)–(18)) and summary (based on Equation (19)). For the classical moments, for better legibility, the quantities $\mu_p'^{1/p}$ are plotted. The upper and lower K-moments are K_p' and \bar{K}_p' , respectively.

The upper and lower K-moments were estimated in two ways. In the detailed estimation, the entire synthetic sample of 10,000 values was used with Equations (15)–(18). In the summary estimation, the 10,000 values were grouped into 110 classes, and for each class $i = 1, \dots, 110$ with x values ranging in $[a_i, a_{i+1})$ the x values were replaced by the midpoint $x_i = (a_i + a_{i+1})/2$ and the estimation was made using Equation (19). It can be observed in Figure 1 that (a) the summary estimates are almost indistinguishable from the detailed estimates and (b) both estimates are almost indistinguishable from the theoretical values. These observations support the reliability of the K-moment concept. Furthermore, they suggest that when the sample size is large, a summary estimation is equally reliable and algorithmically much faster.

For the reader’s convenience, the calculations to produce Figure 1 as well as the figure per se are also provided in a spreadsheet that accompanies the paper as Supplementary Information.

Apparently, if the stochastic variable of interest is of discrete type, the estimation will always be made using Equation (19).

2.3. Theoretical Calculations for Continuous Stochastic Variables, $q = 1$

If \underline{x} is a continuous stochastic variable, then the maximum of p variables, i.e., the variable $\underline{x}_{(p)} = \max(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)$, will have distribution and density functions, respectively,

$$F^{(p)}(x) = F(x)^p, \quad f^{(p)}(x) = pf(x)F(x)^{p-1} \tag{24}$$

where the former is the product of p instances of $F(x)$ (justified by the fact that the variables \underline{x}_i are independent copies of \underline{x} , by definition of the sample concept), while the latter is the derivative of $F^{(p)}(x)$ with respect to x . It is then readily verified that the following proposition holds true.

Proposition 3. For a continuous stochastic variable \underline{x} , the upper and lower K-moments of order p , are, respectively:

$$K'_p = E[\underline{x}_{(p)}] = E[\max(x_1, x_2, \dots, x_p)] = pE[(F(\underline{x}))^{p-1}\underline{x}] \tag{25}$$

$$\overline{K}'_p = E[\underline{x}_{(1;p)}] = E[\min(x_1, x_2, \dots, x_p)] = pE[(\overline{F}(\underline{x}))^{p-1}\underline{x}] \tag{26}$$

Hence, with reference to Equation (11), the adjustment factors are $A(\underline{x}, p) = \overline{A}(\underline{x}, p) = 1$ for continuous variables. It is easy to see that the sequence of K'_p is non-decreasing and that of \overline{K}'_p is non-increasing as p increases.

We note that the calculation of K'_p can be made by

$$K'_p = pE[(F(\underline{x}))^{p-1}\underline{x}] = p \int_{-\infty}^{\infty} (F(x))^{p-1} x f(x) dx = p \int_0^1 x(F)F^{p-1} dF \tag{27}$$

where $x(F)$ is the inverse of function $F(x)$, known as the quantile function. Therefore, we can write

$$G(p) = \int_0^1 x(F)F^{p-1} dF, \quad G(p) := \frac{K'_p}{p} \tag{28}$$

In other words, the function $G(p)$ is the finite Mellin transform [13] of the quantile function $x(F)$. Consequently, if we know the K-moments K'_p for any p , and hence the function $G(p)$, by inverting the transform we can find the quantile function $x(F)$ and hence the distribution function $F(x)$. A similar result can be obtained for the lower K-moments \overline{K}'_p .

2.4. Theoretical Calculations for Discrete Stochastic Variables, $q = 1$

If \underline{x} is a discrete stochastic variable taking on the values $x_j, j = 0, 1, \dots, J$, with probability mass function $P_j := P\{\underline{x} = x_j\} = F(x_j) - F(x_{j-1})$, then $P\{\underline{x}_{(p)} \leq x_i\} = F(x_i)^p$ and, hence, the probability mass function of $\underline{x}_{(p)}$ will be

$$P_j^{(p)} := P\{\underline{x}_{(p)} = x_j\} = P\{\underline{x}_{(p)} \leq x_j\} - P\{\underline{x}_{(p)} < x_j\} = F(x_j)^p - F(x_{j-1})^p \tag{29}$$

where we use the convention $F(x_{-1}) = 0$. Consequently, the upper K-moment of order p is

$$K'_p = E[\underline{x}_{(p)}] = \sum_{j=0}^J (F(x_j)^p - F(x_{j-1})^p) x_j \tag{30}$$

By expanding and making algebraic manipulations, we find

$$K'_p = x_J - \sum_{j=0}^{J-1} F(x_j)^p (x_{j+1} - x_j) \tag{31}$$

In the most common case in which $x_j = j$ the following proposition holds true:

Proposition 4. For a discrete stochastic variable \underline{x} , taking on values $j = 0, \dots, J$, where J is finite or infinite, the upper K-moment of order p , is, respectively:

$$K'_p = J - \sum_{j=0}^{J-1} F(j)^p \tag{32}$$

$$K'_p = \sum_{j=0}^{\infty} (1 - F(j))^p \tag{33}$$

For infinite J and for large j , $F(j) \rightarrow 1$. Thus, the expression in parenthesis in Equation (33) tends to zero. Therefore, we can easily evaluate K'_p numerically from (32), and by choosing a large (but finite) upper limit J , the convergence is fast.

The above relationships do not use the general formula (9). If we want to use it, we should evaluate $A(x_i, p)$. From Equation (30) we obtain

$$K'_p = E[x_{(p)}] = \sum_{j=0}^J F(x_j)^{p-1} \frac{(F(x_j) - F(x_{j-1}))^p / F(x_j)^{p-1}}{pP_j} P_j x_j \tag{34}$$

Thus, by comparing this with Equation (9), we find

$$A(x_j, p) = \frac{F(x_j) - F(x_{j-1})^p / F(x_j)^{p-1}}{pP_j} = \frac{F(x_j)}{p(F(x_j) - F(x_{j-1}))} \left(1 - \left(\frac{F(x_{j-1})}{F(x_j)}\right)^p\right) \tag{35}$$

and finally

$$A(x_j, p) = \frac{1 - (F(x_{j-1}) / F(x_j))^p}{p(1 - F(x_{j-1}) / F(x_j))} \tag{36}$$

As j increases, $F(x_{j-1}) / F(x_j) \rightarrow 1$, and it is easy to see that

$$\lim_{j \rightarrow \infty} A(x_j, p) = 1 \tag{37}$$

thus, approaching the behaviour seen in continuous variables. On the other hand, for varying p

$$\lim_{p \rightarrow 0} A(x_j, p) = \frac{-\ln(F(x_{j-1}) / F(x_j))}{1 - F(x_{j-1}) / F(x_j)}, \quad \lim_{p \rightarrow 1} A(x_j, p) = 1, \quad \lim_{p \rightarrow \infty} A(x_j, p) = 0 \tag{38}$$

If one sets $A(x_j, p) = 1$ to approximate K'_p , the error is prohibitively large. This is illustrated in Figure 2 for two of the most common discrete-variable distributions, Poisson and geometric.

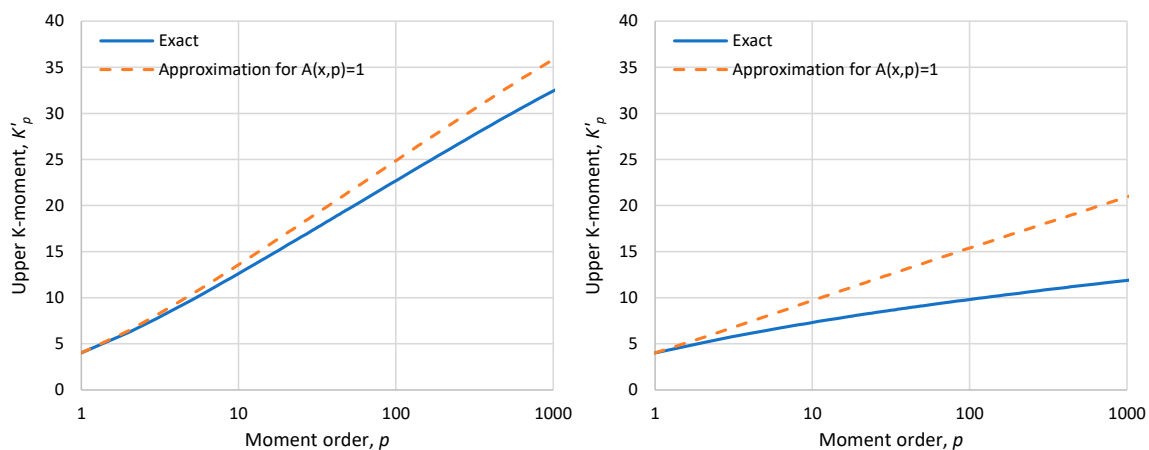


Figure 2. Comparison of exact upper K-moments and approximations thereof setting $A(x_j, p) = 1$, for the distribution (left) Poisson and (right) geometric, where in both cases the mean is 4.

The Poisson distribution has probability mass, distribution, and characteristic K-moments, respectively,

$$P_j = e^{-\lambda} \frac{\lambda^j}{j!}, F(x) = e^{-\lambda} \sum_{j=0}^{\lfloor x \rfloor} \frac{\lambda^j}{j!}, K'_1 = K_{12} = \lambda \tag{39}$$

and the geometric distribution

$$P_j = \lambda(1 - \lambda)^j, F(x) = 1 - (1 - \lambda)^{\lfloor x \rfloor + 1}, K'_1 = \frac{1-\lambda}{\lambda}, K'_2 = \frac{(1-\lambda)(3-\lambda)}{\lambda(2-\lambda)}, \bar{K}'_2 = \frac{(1-\lambda)^2}{\lambda(2-\lambda)}, K_{12} = \frac{1-\lambda}{\lambda^2} \tag{40}$$

Coming now to the lower K-moments, the probability mass function of $\underline{x}_{(1:p)}$ will be

$$\bar{P}_j^{(p)} := P\{\underline{x}_{(1:p)} = x_j\} = P\{\underline{x}_{(1:p)} \geq x_j\} - P\{\underline{x}_{(p)} > x_j\} = \bar{F}(x_{j-1})^p - \bar{F}(x_j)^p \tag{41}$$

where we set $\bar{F}(x_{-1}) = 1$. Consequently, the upper K-moment of order p is

$$\bar{K}'_p = E[\underline{x}_{(1:p)}] = \sum_{j=0}^J (\bar{F}(x_{j-1})^p - \bar{F}(x_j)^p) x_j \tag{42}$$

By expanding and making algebraic manipulations, we find

$$\bar{K}'_p = x_0 + \sum_{j=0}^{J-1} \bar{F}(x_j)^p (x_{j+1} - x_j) \tag{43}$$

and in the most common case in which $x_j = j$, the following proposition holds true:

Proposition 5. For a discrete stochastic variable \underline{x} , taking on values $j = 0, \dots, J$, where J is finite or infinite, the lower K-moment of order p , is, respectively:

$$\bar{K}'_p = \sum_{j=0}^{J-1} \bar{F}(j)^p \tag{44}$$

$$\bar{K}'_p = \sum_{j=0}^{\infty} \bar{F}(j)^p \tag{45}$$

Again, in the latter case, the convergence is fast and therefore we can easily evaluate \bar{K}'_p numerically by choosing a large J ; notice that for large J , $\bar{F}(J) \rightarrow 0$, and the sum will not change if we choose an even larger J .

2.5. Theoretical Calculations for Continuous Stochastic Variables, $q > 1$

As already mentioned, it is possible to generalize the K-moments for transformations of the original variable \underline{x} . Here we consider the transformation $\underline{y} := g(\underline{x}) = (\underline{x} - c)^q$, where c is a real constant and $q \geq 1$ is an integer. This transformation will allow recovering the classical moments, raw and central, as special cases of the K-moments for $p = 1, q > 1$, which has some usefulness as will be seen in Section 3.1.

There are two distinct cases, depending on whether q is odd or even. We start with the former case, which is easier because the transformation is a monotonic (increasing) function. In this case, as demonstrated in Appendix A, we have

$$K'_{py} = p \int_{-\infty}^{\infty} F(x)^{p-1} (x - c)^q f(x) dx \tag{46}$$

For $c = 0$ and for $c = \mu$, we obtain the following results.

Proposition 6. For a continuous stochastic variable \underline{x} , the upper and lower K-moments of orders p, q where q is an odd integer, are, respectively:

$$K'_{pq} = pE[F(\underline{x})^{p-1}\underline{x}^q], \quad \bar{K}'_{pq} = pE[\bar{F}(\underline{x})^{p-1}\underline{x}^q] \tag{47}$$

and the respective central K-moments are, respectively:

$$K_{pq} = pE[F(\underline{x})^{p-1}(\underline{x} - \mu)^q], \quad \bar{K}_{pq} = pE[\bar{F}(\underline{x})^{p-1}(\underline{x} - \mu)^q] \tag{48}$$

This means that, again, $A(\underline{x}, p) = \bar{A}(\underline{x}, p) = 1$.

Next, we consider the case where q is even, in which, as shown in Appendix A, we have

$$K'_{py} = p \int_{-\infty}^{\infty} (F(c + |x - c|) - F(c - |x - c|))^{p-1} (x - c)^q f(x) dx \tag{49}$$

For $c = 0$, we obtain the following result.

Proposition 7. For a continuous stochastic variable \underline{x} , the upper K-moment of orders p, q where q is an even integer, is:

$$K'_{pq} = pE[(F(|\underline{x}|) - F(-|\underline{x}|))^{p-1}\underline{x}^q] \tag{50}$$

and in the case that the variable is non-negative, it simplifies to

$$K'_{pq} = pE[F(\underline{x})^{p-1}\underline{x}^q] \tag{51}$$

This means that for non-negative variables, again, $A(\underline{x}, p) = 1$.

For $c = \mu$, we find the upper central moments as follows:

Proposition 8. For a continuous stochastic variable \underline{x} , the upper central K-moment of orders p, q where q is an even integer, is:

$$K_{pq} = pE[(F(\mu + |\underline{x} - \mu|) - F(\mu - |\underline{x} - \mu|))^{p-1}(\underline{x} - \mu)^q] \tag{52}$$

This means that, in this case, $A(\underline{x}, p) \neq 1$. Specifically,

$$A(\underline{x}, p) = \left(\frac{F(\mu + |\underline{x} - \mu|) - F(\mu - |\underline{x} - \mu|)}{F(\underline{x})} \right)^{p-1} \tag{53}$$

In the special case that the distribution is symmetric about μ , as happens, e.g., in the normal distribution, it holds that $F(\mu - |\underline{x} - \mu|) = 1 - F(\mu + |\underline{x} - \mu|)$. Hence, we can write the following:

Proposition 9. For a continuous stochastic variable \underline{x} with a symmetric distribution, the upper central K-moment of orders p, q where q is an even integer, is:

$$K_{pq} = pE\left[|2F(\underline{x}) - 1|^{p-1}(\underline{x} - \mu)^q\right] = 2p \int_{\mu}^{\infty} (2F(x) - 1)^{p-1}(x - \mu)^q f(x) dx \quad (54)$$

Illustrations of the above results are given in Figure 3 for $q = 2$, for the normal and exponential distributions, and for both noncentral and central upper K-moments. The exact values are compared to approximations obtained by the simplification $A(x, p) = 1$. In the case of noncentral upper K-moments of the exponential distribution, $A(x, p) = 1$ is exact, but this also gives good approximations in other cases, unless the distribution is symmetric (such as the normal distribution) and the moments central.

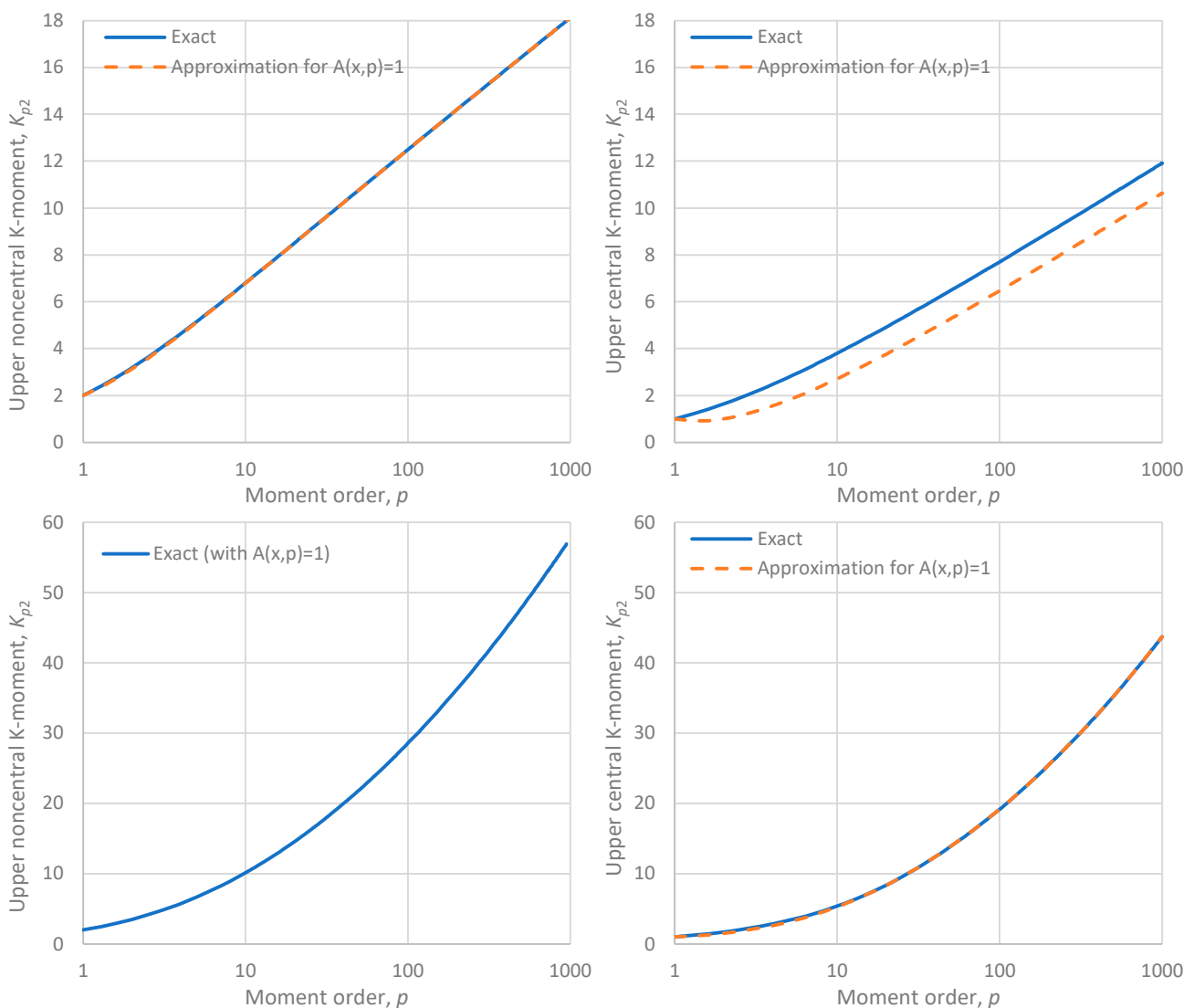


Figure 3. Comparison of exact upper K-moments of orders $p, 2$ and approximations thereof setting $A(x, p) = 1$, for the distribution (**upper**) normal and (**lower**) exponential, where, in both cases, the mean and standard deviation are 1; (**left**) noncentral and (**right**) central upper K-moments.

3. Applications and Advantages

3.1. Evaluation of Classical Moment Estimates

As we have seen (Equation (10)), the classical moments can be recovered as special cases of K-moments. Undoubtedly, the classical moments (and tools related to them such as cumulants) are very useful theoretical concepts. If the distribution function is specified, their true values can be derived easily. However, their estimation from samples is problematic for moment order beyond 3–4 (hence the term unknowable). In contrast, K-moments can be estimated reliably for high orders p (hence their name knowable), provided that the order q is low. Furthermore, K-moments can also predict the value of the estimates of the classical moments, which in fact is not the true value of the classical moment. This may sound paradoxical, given that the classical estimator

$$\hat{\mu}'_p := \frac{1}{n} \sum_{i=1}^n x_i^p \tag{55}$$

is unbiased for order p , however large. In practice, however, the convergence of $\hat{\mu}'_p$ to μ'_p is extraordinarily slow. Thus, the value of $\hat{\mu}'_p$ deviates from μ'_p . This deviation has been termed [2] the slow convergence bias, because, theoretically speaking, based on the bias definition, there is no bias per se. The K-moments can give us an indication of what we can anticipate for the value of $\hat{\mu}'_p$. By examining the moment estimators, we establish relationships between K- and classical moments broader than Equation (10).

Specifically, as shown by Koutsoyiannis [2], the following relationship links the estimated with the true classical moment via the K-moments:

$$\hat{\mu}'_p \approx \frac{K_{n1}^p}{K_{np}'} \mu'_p \tag{56}$$

If the classical moment μ'_p is estimated as the average of m independent samples, each of size n , then the product mn should be substituted for n in Equation (56).

We illustrate the relationship of K- and classical moments using Monte Carlo simulation for the exponential distribution with a lower bound of zero and a scale parameter of 1 and the normal distribution with a mean of $\mu = 1$ and a standard deviation of $\sigma = 1$. The exponential distribution has simple expressions of its moments, i.e.,

$$\mu'_p = K'_{1p} = p!, \quad K'_p = H_p, \quad \bar{K}'_p = 1/p \tag{57}$$

where H_p denotes the p th harmonic number. For $q > 2$, the K-moment K'_{pq} does not have a closed analytical expression but its calculation can be easily performed by numerical integration. For the normal distribution, the classical noncentral moments can be determined by a recursive relationship involving cumulants [2], which, in this particular case, takes the form

$$\mu'_p = \mu \mu'_{p-1} + (p-1)\sigma^2 \mu'_{p-2} \tag{58}$$

For the K-moments of the normal distribution and for $p \leq 4$, the related expressions are contained in Table 4. For $p > 4$, no analytical solution exists but numerical integration can readily give the K-moment values.

Figure 4 shows the simulation results for sample size $n = 100$. The theoretical (true) moments, of orders 1 to 100, of the two distributions are compared to the empirical estimations from a single sample, as well as from an ensemble of m samples (by averaging m estimates where $m = 1000$ and 10 for the exponential and normal distribution, respectively). For the exponential distribution, the single sample estimates are lower than the true moments for $p \geq 4$ and the deviation increases with p , approaching one order of magnitude as p approaches n . For the normal distribution, the deviations are somewhat smaller.

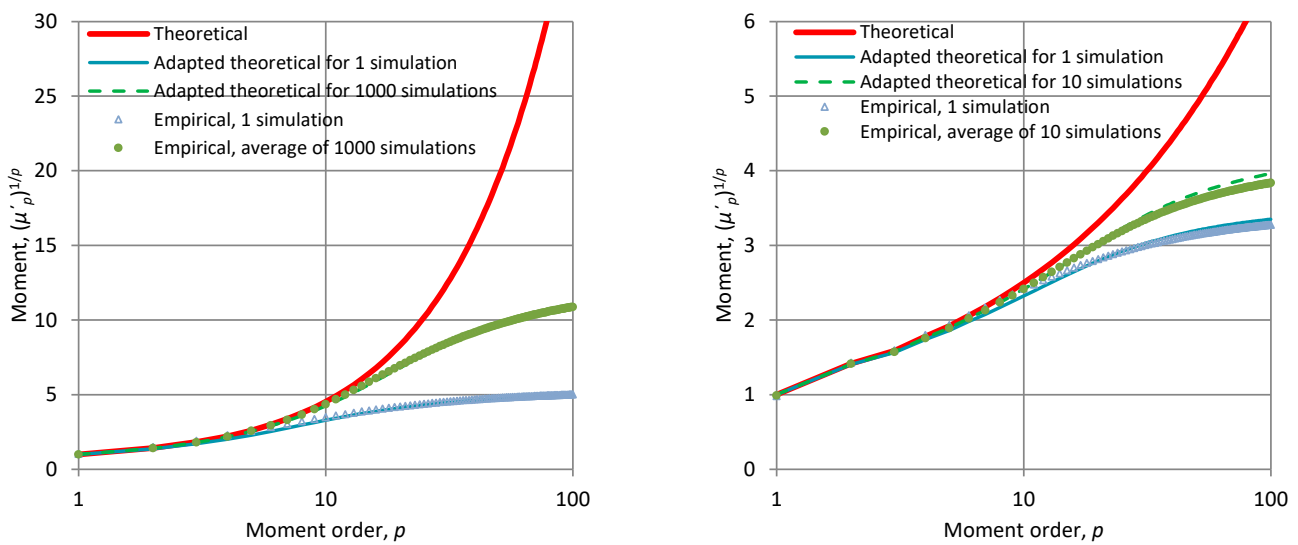


Figure 4. Comparison of the empirical estimates to theoretical values of classical noncentral moments from the indicated number m of simulations of independent samples of size $n = 100$ each from **(left)** the exponential distribution with a mean of 1 and **(right)** the normal distribution with a mean and standard deviation of 1. The empirical estimates are compared to the true moments (theoretical; Equation (55)) and to the values determined by Equation (56) (adapted theoretical).

From a practical point of view—and despite the theoretical guarantee that the classical moment estimates are unbiased—Figure 4 shows that these estimates depart from the true moments. Even the average estimates from 1000 different synthetic samples deviate substantially from the true moments for order $p > 10$. On the other hand, Equation (56), also plotted in Figure 4, very aptly represents the behaviour of the estimates in all cases.

In brief, the classical moment estimators do not estimate the classical moments but rather some hybrid quantities involving both classical and K-moments, as seen in Equation (56).

It should be noted that Equation (56) was extracted for non-negative stochastic variables. However, it can also be applied to variables that can also take negative values, provided that the mean is positive. This was the case for the illustrated normal distribution with $\mu = \sigma = 1$ (Figure 4, right panel), which takes negative values with a probability of 16%.

3.2. Summary Statistics

Because the classical moments beyond order 3–4 are unknowable (from a sample), it has been a customary practice to use those for orders p from 1 to 4 only. These indicate the location, variability, skewness and kurtosis of a distribution, often called summary statistics. For $p = 2, 3, 4$, the central classical moments are used after standardization by the variance raised in a proper power so that a nondimensional metric is obtained.

In using K-moments, as we have seen, we can estimate moments of very high orders, which are particularly useful for the study of extremes [2], as well as for simulation from non-normal distributions [14]. Yet we can use the K-moments of orders p from 1 to 4, to derive summary statistics, such as in the classical moments. To this aim, we enrol the notion of the discrete derivative of K-moments at zero. Earlier studies [1,2] have used the notion of hypercentral K-moments, which for parsimony we avoid defining and using here.

We set $k_j = K'_{j-1}$ for $j \geq 0$ and $k_j = \bar{K}'_{-j+1}$ for $j \leq 0$ and we determine the (centred) discrete derivatives of k_j at $j = 0$. The discrete derivatives are linear functions of k_j terms, with coefficients given in [15] and reproduced in columns 2–8 of Table 2. The calculations are presented in Table 2, where to find the final results we eliminated the lower K-moments as these can also be expressed in terms of upper K-moments. We notice in Table 2 that the second derivative is, by identity, zero and the third derivative is a multiple of the first. Therefore, these are skipped as they do not add any information. Our final result is this:

Proposition 10. *The mean and dispersion of a probability distribution are expressed by the zeroth, and first discrete derivatives of k_j at $j = 0$ and are equal to K'_1 and K_2 , respectively. The skewness and kurtosis of a probability distribution are expressed by the fourth and fifth discrete derivatives, respectively, of k_j at $j = 0$, standardized by K_2 . These four summary statistics are expressed in terms of K-moments of first to fourth orders.*

Table 2. Calculation of discrete derivatives of K-moments.

Derivative	Coefficients of $k_j = K'_{j-1}$ or $k_j = \bar{K}'_{j-1}$ for Lag j							Derivative Expression *	Result *
	-3	-2	-1	0	1	2	3		
0				1				$K'_1 = \bar{K}'_1 = \mu$	$K'_1 = \mu$
1			-1/2	0	1/2			$\frac{1}{2}(K'_2 - \bar{K}'_2)$	K_2
2			1	-2	1			$[K'_2 + \bar{K}'_2 - 2K'_1]$	[0]
3		-1/2	1	0	-1	1/2		$[\frac{1}{2}(K'_3 - \bar{K}'_3 - 2K'_2 + 2K'_1)]$	$[-\frac{1}{2}K_2]$
4		1	-4	6	-4	1		$K'_3 + \bar{K}'_3 - 4K'_2 - 4\bar{K}'_2 + 6K'_1$	$2K_3 - 3K_2$
5	-1/2	2	-5/2	0	5/2	-2	1/2	$\frac{1}{2}(K'_4 - \bar{K}'_4 - 4K'_3 + 4\bar{K}'_3 + 5K'_2 - 5\bar{K}'_2)$	$K_4 - 2K_3 + 2K_2$

* The expressions in square brackets are not used as they do not contain additional information.

These summary statistics are shown in Table 3. We can also recover the classical summary statistics by increasing q from 1 to 4, keeping $p = 1$ and using the zeroth derivative alone (or else not taking derivatives at all). These are also shown in Table 3 and are identical to the classical summary statistics. Interestingly, the summary statistics for $q = 1$ are related to those produced by L-moments [16], as it can be shown that

$$K_2 = \lambda_2, \quad 2\frac{K_3}{K_2} - 3 = \frac{\lambda_3}{\lambda_2}, \quad \frac{K_4}{K_2} - 2\frac{K_3}{K_2} + 2 = \frac{1}{5}\frac{\lambda_4}{\lambda_2} + \frac{4}{5} \tag{59}$$

Table 3. Standardized summary statistics derived by discrete derivatives of K-moments for $q = 1$ or the zeroth derivative for $p = 1$; the latter coincide with the classical summary statistics.

Characteristic	$q = 1$	$p = 1$
Location		$K'_1 = \bar{K}'_1 = \mu$
Dispersion	K_2	K_{12}
Skewness	$2\frac{K_3}{K_2} - 3$	$\frac{K_{13}}{K_{12}^{3/2}}$
Kurtosis	$\frac{K_4}{K_2} - 2\frac{K_3}{K_2} + 2$	$\frac{K_{14}}{K_{12}^2}$

Apparently, as illustrated in Section 3.1, the sample estimates in the case $q = 1$ are more reliable (closer to their theoretical values) than those for $p = 1$. Some examples of theoretical values for customary distributions (normal, exponential, Pareto) are listed in Table 4.

Table 4. Examples of summary statistics for customary distributions.

Case	Symbol	Normal (μ, σ)	Exponential (λ)	Generalized Pareto (λ, ξ)
	$F(x)$	$\frac{1}{2} \operatorname{erfc}\left(\frac{\mu-x}{\sqrt{2}\sigma}\right)$	$1 - e^{-x/\lambda}$	$1 - \left(1 + \xi \frac{x}{\lambda}\right)^{-1/\xi}$
$q = 1$	K'_1	μ	λ	$\frac{\lambda}{1-\xi}$
	K_2	$\frac{\sigma}{\sqrt{\pi}} = 0.56419\sigma$	$\frac{\lambda}{2} = 0.5\lambda$	$\frac{\lambda}{(1-\xi)(2-\xi)}$
	K_3	$\frac{3\sigma}{2\sqrt{\pi}} = 0.84628\sigma$	$\frac{5\lambda}{6} = 0.83333\lambda$	$\frac{(5-\xi)\lambda}{(1-\xi)(2-\xi)(3-\xi)}$
	K_4	$\frac{6\operatorname{ArcTan}[\sqrt{2}]}{\pi^{3/2}}\sigma = 1.02938\sigma$	$\frac{13\lambda}{12} = 1.08333\lambda$	$\frac{(26-9\xi+\xi^2)\lambda}{(1-\xi)(2-\xi)(3-\xi)(4-\xi)}$
	Skewness, $2\frac{K_3}{K_2} - 3$	0	$\frac{1}{3} = 0.33333$	$\frac{1+\xi}{3-\xi}$
Kurtosis, $\frac{K_4}{K_2} - 2\frac{K_3}{K_2} + 2$	$\frac{6}{\pi} \arctan\sqrt{2} - 1 = 0.82452$	$\frac{5}{6} = 0.83333$	$\frac{10-5\xi+\xi^2}{(3-\xi)(4-\xi)}$	
$p = 1$	$K_{12} = \mu_2$	σ^2	λ^2	$\frac{\lambda^2}{(1-\xi)^2(1-2\xi)}$
	$K_{13} = \mu_3$	0	$2\lambda^3$	$\frac{2(1+\xi)\lambda^3}{(1-\xi)^3(1-2\xi)(1-3\xi)}$
	$K_{14} = \mu_4$	$3\sigma^4$	$9\lambda^4$	$\frac{3(3+\xi+2\xi^2)\lambda^4}{(1-\xi)^4(1-2\xi)(1-3\xi)(1-4\xi)}$
	Skewness, $\frac{K_{13}}{K_{12}^{3/2}}$	0	2	$\frac{2\sqrt{1-2\xi}(1+\xi)}{1-3\xi}$
	Kurtosis, $\frac{K_{14}}{K_{12}^2}$	3	9	$\frac{3(1-2\xi)(3+\xi+2\xi^2)}{(1-3\xi)(1-4\xi)}$

3.3. Estimation of Distribution Function

In an observed sample of size n , the i th smallest value is an estimate of the n th order statistic, $x_{(i:n)}$. It is well known in statistics that the stochastic variable $u := F(x_{(i:n)})$ has beta distribution with parameters i and $n - i + 1$ and hence its expected value is:

$$E[u] = E\left[F(x_{(i:n)})\right] = \frac{i}{n+1} \tag{60}$$

Equation (60) constitutes the most widely known and the most popular way of empirically assigning values of the distribution function (and return periods) to observed sample values. It is known as the *Weibull plotting position* (Weibull, 1939; [17]). Namely, to the observed value $x_{(i:n)}$ we assign a distribution function estimate $\hat{F}(x_{(i:n)}) = i/(n+1)$, which is unbiased for $F(x_{(i:n)})$ but not for transformations thereof or for $x_{(i:n)}$ per se. However, Equation (60) is not the earliest, as Hazen (1914; [18]) had proposed a different formula, $\hat{F}(x_{(i:n)}) = (i - 0.5)/n$, which looks similar but, as far as the maximum observation is concerned, the difference in the assigned exceedance probability (and return period) is dramatic, at a ratio of 2:1. Several other formulae have been proposed in the 20th century, which are listed in [2] (Table 5.4) and can be written in the general form

$$\hat{F}(x_{(i:n)}) = \frac{i + A - 1}{n + 2A - 1} \tag{61}$$

where A is a constant ($0 \leq A \leq 1$). The formulae by Weibull and Hazen correspond to $A = 1$ and $A = 1/2$, respectively. Another common value is $A = 5/8$, which was proposed by Blom (1958; [19]) for the normal distribution (see also [20]). The lowest possible value $A = 0$ yields $\hat{F}(x_{(i:n)}) = (i - 1)/(n - 1)$ and thus assigns the values 0 and 1 to the lowest and highest sample values, respectively.

The results for the highest and lowest values are

$$\hat{F}(x_{(n)}) = \frac{n + A - 1}{n + 2A - 1}, \hat{F}(x_{(1:n)}) = \frac{A}{n + 2A - 1} \tag{62}$$

These can also assign distribution function values to K-moments values, which are the expectations of the highest and lowest values:

$$\hat{F}(K'_p) = \frac{p + A - 1}{p + 2A - 1}, \hat{F}(\bar{K}_p) = \frac{A}{p + 2A - 1} \tag{63}$$

However, Koutsoyiannis [2] provided more accurate relationships for K-moment values, based on the following quantities:

Definition 7. The quantities:

$$\Lambda_p := \frac{1}{pF(K'_p)}, \bar{\Lambda}_p := \frac{1}{pF(\bar{K}_p)} \tag{64}$$

are termed the Λ -coefficients.

For given p and distribution function $F(x)$, the quantities $K'_p, \bar{K}_p, \bar{F}(K'_p), \Lambda_p$ and $\bar{\Lambda}_p$ are analytically or numerically determined from their definitions, but this might be complicated. However, Λ_p and $\bar{\Lambda}_p$ exhibit small variation with p , which makes a very good approximation possible if we first accurately determine (a) the values Λ_1 and $\bar{\Lambda}_1$ for $p = 1$, and (b) the asymptotic values Λ_∞ and $\bar{\Lambda}_\infty$. The values Λ_1 and $\bar{\Lambda}_1$ are very easy to determine, as they refer to the probability of non-exceedance of the mean:

$$\Lambda_1 = \frac{1}{1 - F(\mu)}, \bar{\Lambda}_1 = \frac{1}{F(\mu)} = \frac{\Lambda_1}{\Lambda_1 - 1} \tag{65}$$

and can also be reliably estimated from a sample by finding the proportion of sample values that are smaller than μ . For symmetric distributions, $F(\mu) = 1/2$ and thus $\Lambda_1 = \bar{\Lambda}_1 = 2$.

The asymptotic values can be found by utilizing results by Koutsoyiannis [2] and depend on the tail behaviour of the distribution, as shown in Table 5. From the limits also listed in Table 5, we infer that when the distribution is upper or lower bounded, the upper or lower Λ -coefficient is in the range $(0, e^\gamma = 1.78107)$. Otherwise (when the distribution is unbounded), the Λ -coefficient is greater than e^γ . In unbounded variables that have finite variance (as happens with several natural processes), the tail index should be bounded by $1/2$ ($\xi, \xi' < 1/2$) and thus the upper or lower Λ -coefficient is in the range $(e^\gamma = 1.78107, \pi = 3.14159)$.

Table 5. Asymptotic values of Λ -coefficients (based on [2]).

Characteristic	Definition of Tail Index ¹	Asymptotic Λ	Limit for $\xi, \xi', \bar{\xi}, \bar{\xi}' =$				Equation No.
			0	1/2	1	∞	
Upper bounded by c_U , tail index ξ'	$\lim_{x \rightarrow c_U} (c_U - x)^{-\xi'} \bar{F}(x) = l_1$	$\Lambda_\infty = \Gamma(1 + 1/\xi')^{-\xi'}$	0	$1/\sqrt{2}$	1	e^γ	(66)
Upper unbounded, tail index ξ	$\lim_{x \rightarrow \infty} x^{1/\xi} \bar{F}(x) = l_2$	$\Lambda_\infty = \Gamma(1 - \xi)^{1/\xi}$	e^γ	π	∞	-	(67)
Lower bounded by c_L , tail index ξ	$\lim_{x \rightarrow c_L} (x - c_L)^{-\xi} F(x) = l_3$	$\bar{\Lambda}_\infty = \Gamma(1 + 1/\xi)^{-\xi}$	0	$1/\sqrt{2}$	1	e^γ	(68)
Lower unbounded, tail index ξ'	$\lim_{x \rightarrow -\infty} (-x)^{1/\xi'} F(x) = l_4$	$\bar{\Lambda}_\infty = \Gamma(1 - \xi')^{1/\xi'}$	e^γ	π	∞	-	(69)

¹ $l_i, i = 1, \dots, 4$ are constants $< \infty$; $\Gamma(a)$ is the gamma function; $\gamma = 0.577216$ is the Euler constant and thus $e^\gamma = 1.78107$.

For the most common case that the domain of the stochastic variable x is $[0, \infty)$ and the variance is finite, Λ_∞ will be in the range (e^γ, π) , depending on the upper tail index ξ , and $\bar{\Lambda}_\infty$ will be in the range $(0, e^\gamma)$, depending on the lower tail index ζ . Both tail indices ξ and ζ can be visualized in a double logarithmic plot of the ratio $F(x)/\bar{F}(x)$ vs. x . As $x \rightarrow 0$, given the definitions in Table 5, we will have $F(x) \propto x^\zeta$ and $F(x)/\bar{F}(x) \propto x^\zeta$. Therefore, ζ is the double logarithmic slope of $F(x)/\bar{F}(x)$ as $x \rightarrow 0$. On the other hand, as $x \rightarrow \infty$, we will have $1/\bar{F}(x) \propto x^{1/\xi}$ and $F(x)/\bar{F}(x) \propto x^{1/\xi}$. Therefore, ξ is the inverse of the double logarithmic slope of $F(x)/\bar{F}(x)$ as $x \rightarrow \infty$.

Once the tail indices ξ and ζ are known, the asymptotic Λ -coefficients can be determined from Table 5. Notably, most of the customary distributions, such as normal, lognormal, exponential, Gamma, Weibull, etc. (more specifically, those belonging to the domain of attraction of the Extreme Value Type I distribution; see [2], pp. 76–79 and 235–236), have an upper tail index $\xi = 0$ and thus $\Lambda_\infty = e^\gamma$.

Given the three quantities Λ_1 , Λ_∞ and $\bar{\Lambda}_\infty$ (with $\bar{\Lambda}_1$ determined from Equation (65)), we can approximate Λ_p for any order p by

$$\Lambda_p \approx \Lambda_\infty + \frac{(\Lambda_1 - \Lambda_\infty)}{p}, \quad \bar{\Lambda}_p \approx \bar{\Lambda}_\infty + \frac{\bar{\Lambda}_1 - \bar{\Lambda}_\infty}{p} \tag{70}$$

from which we find

$$\hat{F}(K'_p) \approx 1 - \frac{1}{\Lambda_\infty p + \Lambda_1 - \Lambda_\infty}, \quad \hat{F}(\bar{K}'_p) \approx \frac{1}{\bar{\Lambda}_\infty p + \bar{\Lambda}_1 - \bar{\Lambda}_\infty} \tag{71}$$

More accurate approximations are given in [2] (pp. 208–215), but these depend on the entire expression of distribution function—not only on its tail indices. Nonetheless, the approximation by Equation (71) suffices for most statistical tasks.

By comparing Equations (63) and (71), we deduce that the former is a special case of the latter, in which $\Lambda_1 = \bar{\Lambda}_1 = 2$ and $\Lambda_\infty = \bar{\Lambda}_\infty = 1/A$. We understand, thus, that the approximation by Equation (71), which is based on three independent parameters (Λ_1 , Λ_∞ and $\bar{\Lambda}_\infty$), is more accurate and flexible than that of plotting positions, i.e., Equations (61)–(63). There is an additional reason that makes the former even more accurate. Each K'_p or \bar{K}'_p (excepting the case $p = n$) is estimated by several sample values, while in Equations (61)–(63) each $x_{(i:n)}$ is a single data value. An additional advantage of using Equation (71) is that it can be applied for any desired p , integer or real (up to the sample size n) while the equations based on plotting positions work only for the individual values of the observed sample.

An illustration of the method for estimating the distribution function from samples is shown in Figure 5. This is for the generalized Pareto distribution and for a rather small sample size, $n = 100$. Monte Carlo simulation was employed to generate 100 series from the same distribution and find the median at each p and the uncertainty band. The figure includes a Cartesian plot of the probability distribution function $F(x)$ (left), as well as a double logarithmic plot of the ratio $F(x)/\bar{F}(x)$. As discussed above, the latter enables visualizing the tails of the distribution. Note that this ratio is closely associated with the return period, a quantity widely used in engineering design [2], by

$$\frac{F(x)}{\bar{F}(x)} = \frac{T - D}{D} = \frac{D}{\bar{T} - D} \tag{72}$$

where T and \bar{T} denote return periods of maxima and minima, respectively, and D is the time step of observations.

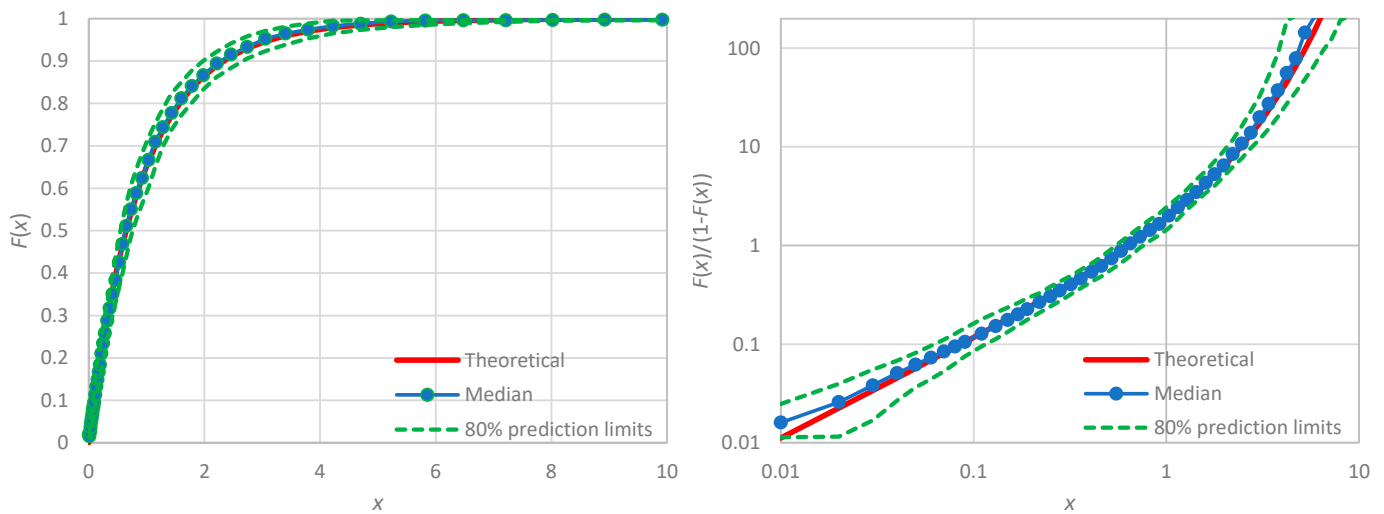


Figure 5. Illustration of the estimation, using K-moments, of the probability distribution function, plotted in a Cartesian plot of $F(x)$ (left) and a double logarithmic plot of $F(x)/\bar{F}(x)$ (right). 100 data series of $n = 100$ values each were generated from a Generalized Pareto distribution (see Table 1) with $\mu = 1$ and tail index $\zeta = 0.1$, and then processed to calculate the median and produce the uncertainty band.

The empirical estimates in Figure 5 are also compared with their theoretical values, also plotted in the figure. There is good agreement between the two, and the uncertainty band is fairly narrow.

When the stochastic variable of interest is discrete, the situation is more complicated [21,22]. In this case, the distribution function $F(x)$ is a discontinuous function with jumps at each possible realization of the stochastic variable, such as at the integers $j = 0, 1, \dots$. On the other hand, the K-moments, which are averages of maxima or minima, take on real values. We assume that there exists a distribution function, $F_C(x)$, of a continuous variable, which yields the same K-moments as $F(x)$. Theoretically, this can be determined as follows. We first estimate the K-moments K'_p from the discrete $F(x)$ and we equate them with those of $F_C(x)$. Knowing the latter, we use the inverse Mellin transform of (28) to estimate the quantile function $x_C(F)$. By inverting the latter, we find $F_C(x)$. However, this will be computationally cumbersome. Here we use the simplification:

$$F(x) = F_C(\lfloor x \rfloor + c) \tag{73}$$

assuming a constant $c = 0.25$, which was found optimal after numerical investigation. Based on this and assuming a broken-line form of $F_C(x)$, we write

$$F_C(x) = \begin{cases} F(0)x/c & x \leq c \\ F(\lfloor x - c \rfloor) + (F(\lfloor x - c \rfloor + 1) - F(\lfloor x - c \rfloor))(x - c - \lfloor x - c \rfloor), & x > c \end{cases} \tag{74}$$

This is illustrated for the Poisson distribution (Equation (39)) and geometric distribution (Equation (40)) in Figure 6. Both distributions belong to the domain of attraction of the Extreme Value I distribution and therefore $\Lambda_\infty = e^\gamma$. From Equation (74), the lower tail is $\zeta = 1$ and therefore $\bar{\Lambda}_\infty = 1$. Figure 6 is similar to Figure 5 and was constructed with the same method, except that the theoretical curve plotted is $F_C(x)$ instead of $F(x)$. The reason is that it is the former that is consistent with the K-moment estimates, which indeed compare well with the theoretical expectation. The uncertainty bands are again fairly narrow.

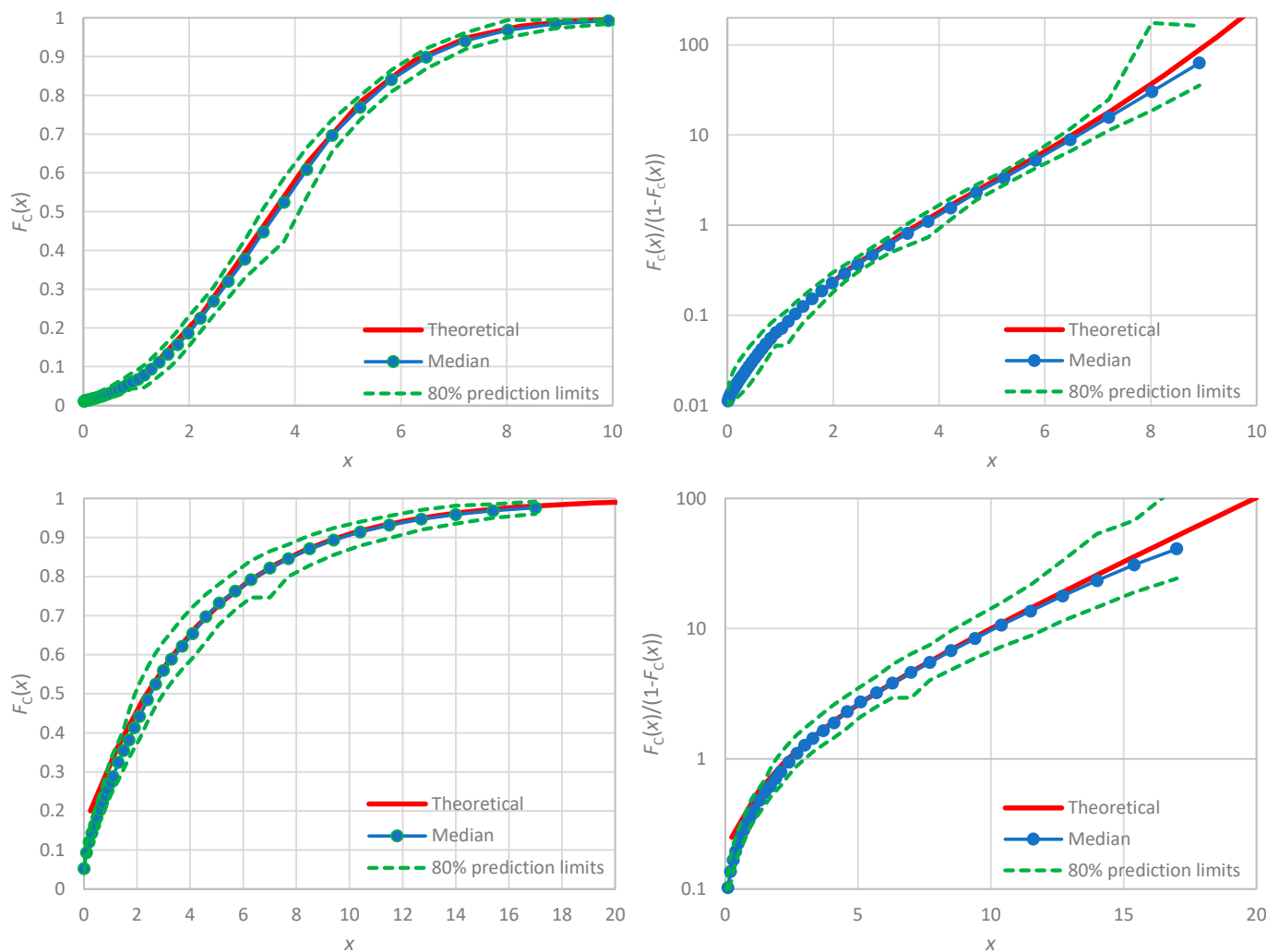


Figure 6. Illustration of the estimation, using K-moments, of the probability distribution function, plotted in a Cartesian plot of $F(x)$ (**left column**) and a logarithmic plot of $F(x)/\bar{F}(x)$ (**right column**). The distributions illustrated are the Poisson (**upper row**; Equation (39)) and the geometric (**lower row**; Equation (40)) for $\mu = 4$. For each of them, 100 data series of $n = 100$ values each were generated and then processed to calculate the median and produce the uncertainty band. The line marked “Theoretical” is $F_C(x)$, based on Equation (74).

3.4. Fitting a Distribution Function

The method of moments has been one of the standard techniques in fitting distribution functions to data. Another standard is the method of maximum likelihood, which is well reasoned and is based on an optimization logic. In contrast, the method of moments is based on solving equations and is not quite rigorously argued. Assuming that we fit a two-parameter model (say, a generalized Pareto distribution), the method of moments uses the first two classical (noncentral) moments and determines the two parameters by equating the sample moments to the theoretical moments of the distributions. A similar strategy is followed by the method of L-moments, which is regarded as more reliable than that of classical moments. However, Koutsoyiannis [2] raised the following two major questions on the logic of this strategy, particularly when the focus of the study is the occurrence of extremes:

1. *Why use the first and second moments and not, say, the second and third? One may easily justify the standard choice of using the lowest possible order of moments by the fact that higher moments are less accurately estimated. On the other hand, one may counter that, when we are interested in extremes, these are better reflected in higher-order moments. It is well known that a model can hardly be a perfect representation of reality. Thus, we cannot expect that a good*

model fitting on the first and second moment would be equally suitable for the distribution tail, i.e., the behaviour on extremes.

2. Why use two moments and not more? The standard answer, that two equations suffice to find two unknowns, may be adequate from a theoretical mathematical point of view but it is not from an empirical and engineering one. (As the saying goes, to draw a straight line, a mathematician needs two points but an engineer needs three). Certainly, an optimization framework (as in maximizing likelihood or in minimizing fitting error) is much preferable and superior to an equation solving method.

The K-moments and their advantages, which are particularly strong for an extreme-oriented modelling, enable a different fitting strategy, which uses as many moments as required for the particular application. For example, assuming that we have a sample of size n , we can estimate n upper K-moments and n lower K-moments for integer values of p . We can then use the former if we are interested in the body of the distribution and its upper tail, or the latter if we are interested in the body of the distribution and its lower tail, or we can use both sequences simultaneously. Alternatively, if the sample size is too large, we could choose a subset of them, e.g., 100 values of p (not necessarily integer) arranged in a geometric progression from 1 to n . The objective of the fitting algorithm is to minimize the average mean square error (possibly weighted) between the empirical and theoretical K-F relationships.

Koutsoyiannis [2] discusses a possible criticism of using high-order K-moments (up to order n), which is the fact that the higher-order moments are more uncertain than the lower ones, and notes:

This criticism would be valid if the true distribution function was known to be the one chosen as a model for the real-world process studied. But this is hardly the case. Let us assume that in the time series of flow observations we have three very high values and that we have chosen a certain model, e.g., a Lognormal distribution. How can we be sure that the model is correct? If we are not sure (which actually is always the case), and if we are to design a certain engineering construction, would we prefer a fitting of the chosen model that is consistent with theoretical considerations, e.g., based on the maximum likelihood method, even if this yields a departure for the three high values? Or would we feel safer if our fitting represents well the three high values?

The steps of the fitting algorithm, adapted from [2], are outlined below. These steps are based on the upper K-moments, K'_p , and are oriented toward a fitting that is good for the body of the distribution and the upper tail. If we are interested in extreme lows, we can substitute the lower K-moments \overline{K}'_p for the upper ones, and appropriately adapt the following steps. If we are equally interested in both tails, then we can use both upper and lower tails:

1. We choose a number $m + 1$ of moment orders, i.e., $p_i = n^{i/m}, i = 0, \dots, m$, with $p_0 = 1, p_m = n$. The rationale for this is that when dealing with samples of size n of the order of several thousands, the number m could be chosen much smaller, e.g., of the order of 100, to speed up calculations without compromising accuracy. The orders p_i need not be natural numbers.
2. We estimate the upper K-moments K'_{p_i} of orders $p_i, i = 0, \dots, m$ using Equations (15) and (18).
3. Assuming default values of the distribution parameters, represented as a vector λ , we determine Λ_1 from Equation (65), the theoretical relationships between parameters of the specified distribution and the mean $\mu = K'_1$ (such as those in Table 1 or Table 4) and the expression of the distribution function $F(x)$. Alternatively, but not preferably, we can estimate Λ_1 from the sample mean, along with Equation (65).
4. As the vector λ contains the tail index ζ , we determine Λ_∞ from the relationships of Table 5.
5. Given Λ_1 and Λ_∞ , for each p_i we estimate the empirical distribution function value $\hat{F}(K'_{p_i})$ from Equation (71).

6. Given the parameter values in vector λ , for each K-moment K'_{p_i} , we estimate the theoretical distribution function $F(\hat{K}'_{p_i})$ from the expression of the distribution function $F(x)$ setting $x = \hat{K}'_{p_i}$.
7. We form an expression for the total fitting error as the sum:

$$E(\lambda) := \sum_{i=0}^m w_i \left(\ln \left(\frac{\hat{F}(K'_{p_i})}{\hat{\bar{F}}(K'_{p_i})} \right) - \ln \left(\frac{F(\hat{K}'_{p_i})}{\bar{F}(\hat{K}'_{p_i})} \right) \right)^2 \tag{75}$$

where w_i denotes a weighting coefficient. The rationale of choosing the logarithmic deviations of the ratio $F(x)/\bar{F}(x)$, instead of $F(x)$, is that this better represents the distribution tails (see e.g., Figure 5, right). The error E is a function of the chosen parameters λ , and we evaluate it for the chosen parameter set.

8. We repeat the calculations of steps 3–7 for different sets of parameter vectors λ until the fitting error becomes minimal. The repetitions are executed by a solver (available in every software environment) using as objective function (to be minimized) the error $E(\lambda)$.

The procedure is typically very fast, almost instant, as the only step with computational burden, that in step 2, is executed only once. A default value of the weight is $w_i = 1$. Two alternatives are:

$$w_i = \begin{cases} 1, & l \leq K'_{p_i} < u \\ 0, & \text{otherwise} \end{cases}, \quad w_i = \left(\frac{\hat{F}(K'_{p_i})}{\hat{\bar{F}}(K'_{p_i})} \right)^b \tag{76}$$

where the former focuses the fitting on a particular range of the variable, that between chosen lower and upper bounds l and u , respectively, while the latter gives more emphasis on fitting to high values of the distribution function by choosing a constant $b > 0$ (e.g., $b = 0.5$).

Remarkably, the above procedure does not need evaluation of the theoretical K-moments per se (except for the mean $\mu = K'_1$). While this evaluation is feasible for any distribution, it may be cumbersome as it may involve numerical integration. Notice in Equation (75) the quantity $\hat{F}(K'_{p_i})$ is estimated from Equation (71) as a function of p_i while $F(\hat{K}'_{p_i})$ is estimated from the data using Equations (15) and (18). This makes the procedure as simple as possible. In addition, it readily provides a visual comparison.

Several examples of distribution fitting using real world data are presented in [2]. Here we repeat, after adaptation, one of them for the sake of illustration. This is for hourly wind data of the MIT station in Boston, MA, USA (42.367° N, 71.033° W, 9.0 m), perhaps the longest worldwide. The data are available from the United States National Oceanic and Atmospheric Administration, but a great deal of effort was needed to convert them to hourly time series [23]. They cover, with minor gaps, the period of January 1945 to December 2014 (70 calendar years with a few missing values; a number of hourly values of 589,551, 1% of which are zero and were excluded from the analysis, which was made for the remaining $n = 583,465$ nonzero values). The model chosen is the ParetoBurr–Feller (PBF; [2]) distribution with four parameters (the tail indices ζ, ξ , the scale parameter λ and the lower bound x_L , whose characteristics that are needed for the fitting are:

$$F(x) = 1 - \left(1 + \zeta \xi \left(\frac{x - x_L}{\lambda} \right)^\zeta \right)^{-\frac{1}{\xi \zeta}},$$

$$A_1 = \left(1 + \left(\frac{B\left(\frac{1}{\xi \zeta}, \frac{1}{\xi}, \frac{1}{\zeta}\right)}{\xi} \right)^\zeta \right)^{\frac{1}{\xi \zeta}}, \quad \Lambda_\infty = \Gamma(1 - \xi)^{\frac{1}{\xi}}, \quad \Lambda_\infty = \Gamma\left(1 + \frac{1}{\xi}\right)^{-\xi} \tag{77}$$

The resulting fit is shown in Figure 7 in terms of a plot of the ratio $F(x)/\bar{F}(x)$, which enables visualization of the tails. By comparing the theoretical curve with the empirical one based on K-moments, we see that the fitting is excellent for the entire range that can be observed from the sample, spanning 11 orders of magnitude of the ratio $F(x)/\bar{F}(x)$.

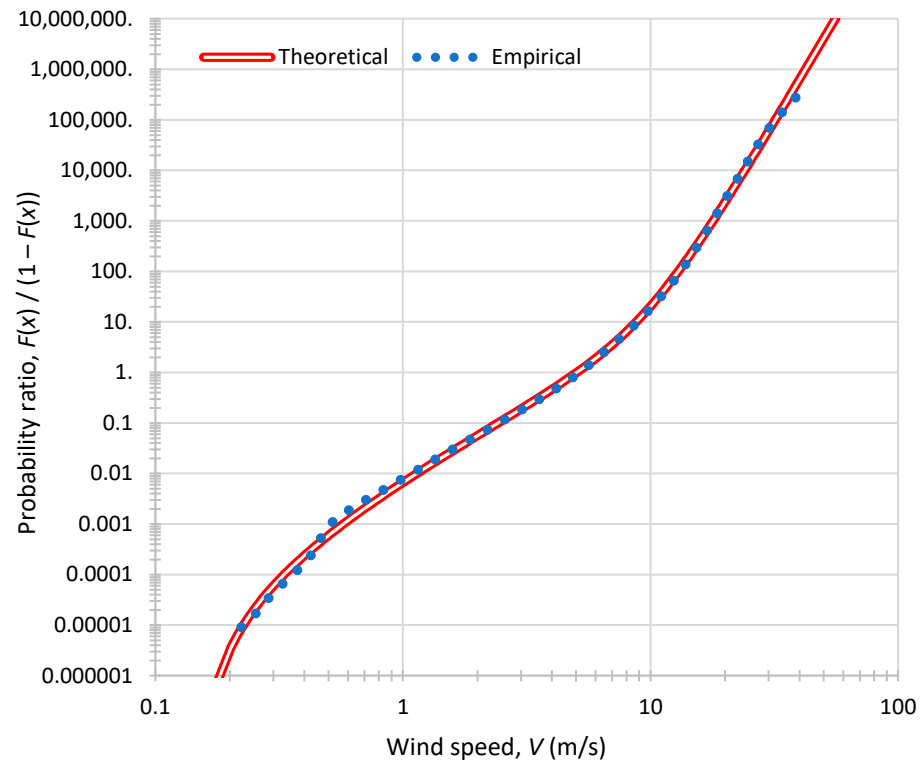


Figure 7. Comparison of empirical and theoretical PBF distribution fitted on the hourly wind speed time series at the MIT station in Boston, MA, USA, excluding the zero values. For visualization of the distribution tails, the distribution function is depicted in terms of the ratio $F(x)/\bar{F}(x)$ and the axes are logarithmic. The parameters are: $\zeta = 0.120, \zeta = 2.73, \lambda = 5.35 \text{ m/s}, x_L = 0.15 \text{ m/s}$ and were fitted on the entire domain of wind speed. The fitting was performed based on upper and lower K-moments using the approximation of Equation (71). The points for K-moments have been adapted for the bias effect of time dependence (see Section 3.6).

3.5. Estimation of Density Function

When assigning values of distribution function to the K-moments, the arrangement of point estimates thereof produces a smooth curve, which allows a direct estimate of the probability density. This is a strong advantage of the K-moments as it is the only method that can provide a detailed representation of the density function, replacing the rougher representation offered by the popular concept of the histogram [3].

Specifically, given the $2n - 1$ sample estimates of K-moments \hat{K}'_p and $\hat{\bar{K}}'_p$, namely the ordered values $\hat{\bar{K}}'_n \leq \hat{\bar{K}}'_{n-1} \leq \dots \leq \hat{\bar{K}}'_1 \equiv \hat{K}'_1 \leq \hat{K}'_2 \leq \dots \leq \hat{K}'_{n-1} \leq \hat{K}'_n$, and the estimates of the distribution function $\hat{F}(K'_p)$ and $\hat{F}(\bar{K}'_p)$ from Equation (71), it is then straightforward to approximate the density function $f(x) = dF(x)/dx$ by the discrete derivative:

$$\hat{f}(x) = \begin{cases} \frac{\hat{F}(K'_{p+1}) - \hat{F}(K'_p)}{\hat{K}'_{p+1} - \hat{K}'_p}, & \hat{K}'_p \leq x < \hat{K}'_{p+1} \\ \frac{\hat{F}(\bar{K}'_p) - \hat{F}(\bar{K}'_{p+1})}{\hat{\bar{K}}'_p - \hat{\bar{K}}'_{p+1}}, & \hat{\bar{K}}'_{p+1} \leq x < \hat{\bar{K}}'_p \end{cases} \quad (78)$$

This will result in $2n - 2$ different values of the density $\hat{f}(x)$, while it is possible to expand the number of estimation points by using non-integer orders p [3]. The method is illustrated in Figure 8 for the Generalized Pareto distribution, the same as in Figure 5, by means of Monte Carlo simulation with 100 realizations of samples of 100 items each. The estimated probability density, expressed in terms of the median of the simulations, harmonizes with the true shape of the probability density, and the uncertainty, depicted in terms of prediction limits, is low in the body of the distribution, but increases in the upper and lower tails.

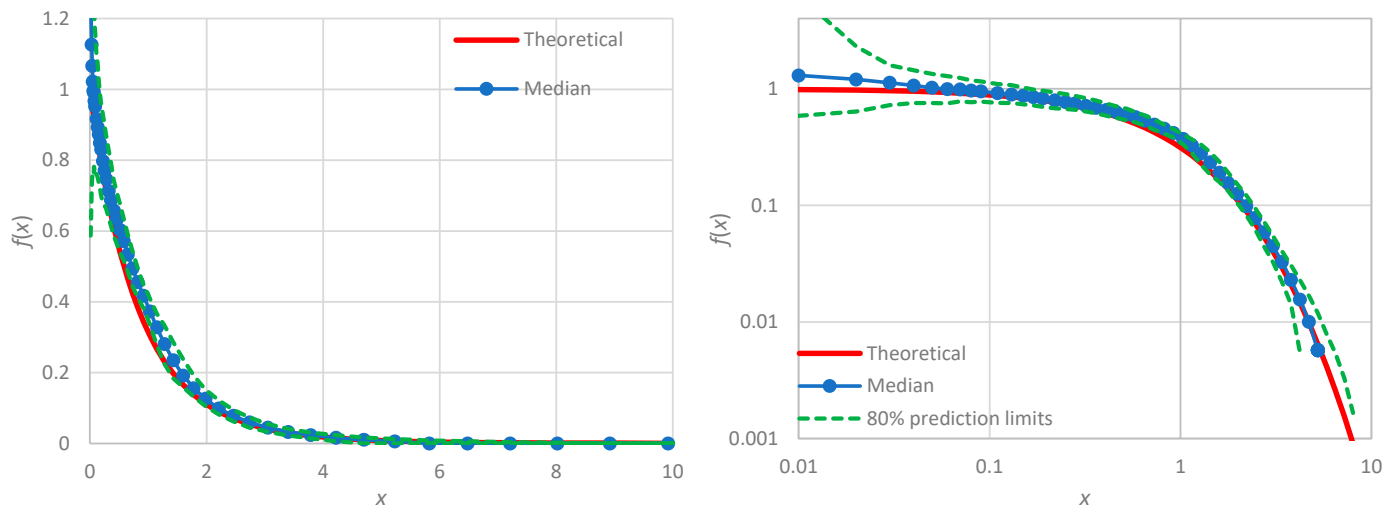


Figure 8. Illustration of the median estimate and uncertainty (in terms of prediction limits) of the probability density using the proposed method, plotted in Cartesian (left) and logarithmic (right) axes. The original results from the proposed method are interpolated at the points that are plotted in the graphs. The true distribution is Generalized Pareto (see Table 1) with $\mu = 1$ and $\zeta = 0.1$, from which 100 data series of $n = 100$ values each were generated and processed to calculate the median of the estimates and produce their uncertainty band.

3.6. Accounting for Estimation Bias in Time Series with Time Dependence

K-moments characterize the marginal (first order) distribution of a stochastic process and therefore are not affected by the dependence structure of the process. However, their estimators are affected. If we do not have a random sample of a stochastic variable, but rather a time series from a process with time dependence, then bias is induced to estimators of K-moments. Thus, unless the stochastic process is white noise, the estimator unbiasedness ceases to hold. It has been shown in [2] that the estimation bias is negligible in the case of processes with short-range dependence and those characterized by periodic oscillations.

However, in processes with long-range dependence (LRD), the bias can be substantial. LRD, also known as *Hurst–Kolmogorov* (HK) dynamics, is quantified by the Hurst parameter, H , which in turn is most conveniently defined through its climacogram as follows. Let us consider any stochastic process $\underline{x}(t)$ in continuous time t , and define the cumulative process as

$$\underline{X}(k) := \int_0^k \underline{x}(t) dt \tag{79}$$

Then $\underline{X}(k)/k$ is the time averaged process. The climacogram $\gamma(k)$ of the process $\underline{x}(t)$ is the variance of the time averaged process at time scale k [24], i.e.,

$$\gamma(k) := \text{var} \left[\frac{\underline{X}(k)}{k} \right] \tag{80}$$

The climacogram of the HK process is given as a power function of the time scale k , i.e.,

$$\gamma(k) = \lambda^2 \left(\frac{\alpha}{k}\right)^{2-2H} \tag{81}$$

where α and λ are scale parameters with units of time and $[x]$, respectively. In a purely random process, $H = 1/2$, while in most natural processes, $1/2 \leq H \leq 1$, as first observed by Hurst in 1951 [25]. A high value of H (approaching 1) indicates an enhanced presence of patterns such as grouping of similar events (e.g., floods, droughts) in time, enhanced change and enhanced uncertainty (e.g., in future predictions). A low value of H ($<1/2$) indicates enhanced fluctuation or antipersistence. Processes with more complex expressions of climacogram also exhibit LRD if

$$\lim_{k \rightarrow \infty} \gamma(k)k^{2-2H} = l \tag{82}$$

where l is finite.

Koutsoyiannis [2] assumed that the estimate \hat{K}_p^d derived by the standard estimators (15)–(18), where the superscript ‘d’ stands for dependence, results in a bias that is proportional to K_p , so that the relative bias

$$\frac{\hat{K}_p^d - K_p}{K_p} = \Theta(n, H) \tag{83}$$

does not depend on the moment order p but rather only on the sample size n and the Hurst parameter H . Based on this assumption, he derived the relative bias for an HK process as

$$\Theta(n, H) \approx \frac{2H(1 - H)}{n - 1} - \frac{1}{2(n - 1)^{2-2H}} \tag{84}$$

Further, assuming that \hat{K}_p^d equals the true K-moment of an order $p' < p$, the adjusted moment order p' was derived as

$$p' \approx 2\Theta + (1 - 2\Theta)p^{(1+\Theta)^2}, \quad K_{p'} = K_p^d = (1 + \Theta)K_p \tag{85}$$

The above relationships enable the following simple algorithm to approximately estimate the true moments from a time series of length n , which is a realization of a process with LRD:

1. We construct the climacogram and estimate the Hurst parameter H .
2. From Equation (84), we estimate the relative bias $\Theta(n, H)$.
3. We choose a moment order p and estimate by the standard estimators (15)–(18) the K-moment K'_p (or K_p) as if we had an independent sample.
4. We adapt p by Equation (85) and infer that the quantity estimated in step 3 is an estimate of $K'_{p'}$ (or $K_{p'}$).

This is for moment order $q = 1$. A similar approach for higher q is discussed in [2].

An illustration is provided in Figure 9, which was made by stochastic simulation assuming an HK process with marginal distribution generalized Pareto. The generation of a simulated time series from the process is conducted using a simple method termed the *multiple time-scale fluctuation approach* [26]. This generates Gaussian stochastic variables \underline{u} with climacogram that closely approximates that of Equation (81), by adding three independent linear Markov processes (autoregressive of order 1), each having a different lag-one autocorrelation coefficient, calculated as a function of H , as described in [26]. The Gaussian series are subsequently transformed to have any desired distribution F (in our case, generalized Pareto) by $\underline{x} = F^{-1}(\Phi(\underline{u}))$, where Φ is the normal distribution function.

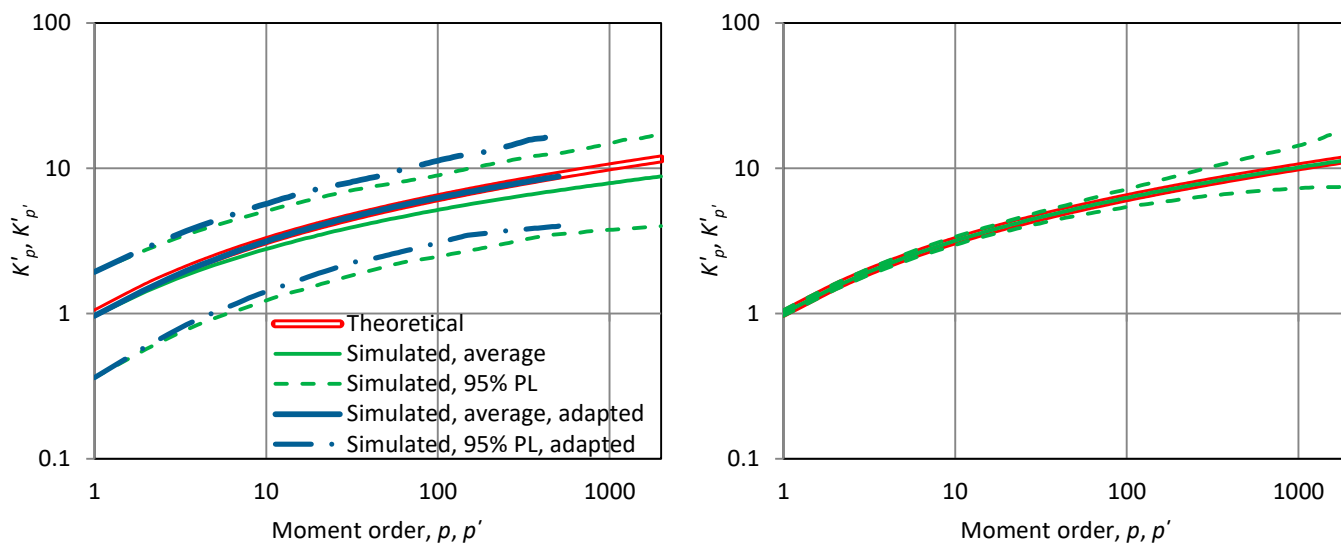


Figure 9. (Left) Illustration of the performance of the adaptation of K-moment estimation for an HK process with Hurst parameter $H = 0.9$ and generalized Pareto marginal distribution (with $\mu = 1$ and $\zeta = 0.1$). (Right) For comparison, a graph for independent sample ($H = 0.5$) is provided, which does not need adaptation. The estimates are averages of 200 simulations each with $n = 2000$ and are almost indistinguishable from the theoretical values (after the adaptation in the left graph). The 95% prediction limits (PL) are also shown.

The simulation results depicted in Figure 9 are for a high Hurst parameter, $H = 0.9$, to magnify the bias. Interestingly, what would be assigned as K-moment for $p = 2000$, without taking into account the effect of dependence, actually corresponds to the true K-moment of $p' \approx 500$. This has dramatic consequences in the assignment of probabilities of exceedance (or return periods). Without adaptation, we would assign a probability of exceedance four times lower than the true one (or a four-fold return period). Another important observation from comparing the two panels of the figure, which are for the same marginal distribution, is the dramatic broadening of the prediction intervals in the case of LRD. This illustrates how dependence amplifies uncertainty.

The above technique has already been applied to the wind data example in Figure 7, as the estimated Hurst parameter is high, $H = 0.9$, and its effect (bias) on distribution function estimation is substantial, with relative bias $\Theta = -0.06$. Removing the bias resulted in a reduction of the maximum order $p = n = 583,465$ to $p' = 140,117$ (i.e., $p' < p/4$).

4. Discussion

K-moments share some characteristics with the classical moments, L-moments [16] and probability-weighted moments (PWM; [27]), which are discussed in detail in [2] and in part in Sections 3.1 and 3.2 here. On the other hand, the differences are huge. Thus, the K-moments are knowable with unbiased estimators (from samples) for high orders, up to the sample size n , while in classical moments, the convergence bias is substantial and the estimation uncertainty is orders of magnitude higher. Furthermore, as discussed in Section 3.1, the K-moments can reveal what the estimators of classical moments actually estimate, which is different from the true values thereof.

As per the L-moments and PWMs, the logic and strategies that is followed for these concepts do not differ from those of classical moments. Up to this point, the K-moments are similar to them and produce similar summary statistics, as discussed in Section 3.2. However, the real power of K-moments is that their values can directly be assigned probabilities, through Λ -coefficients. This is similar to what happens with order statistics, but with some advantages over the latter, as the K-moments make optimal use of the entire dataset, instead of using just one data value to assign probability. Thus, the sequence of

K-moment values produces a smooth curve that can be used to empirically estimate the density function, a unique feature among all moment types.

The assignment of probabilities to numerous upper and lower K-moment values, for orders up to the sample size, enables a rich characterization of the distribution function, as well as a strategy for model fitting totally different from other methods of moments. This strategy involves the entire range of moment orders (rather than the first 1–4 moments), can focus on different parts of the distribution (body, upper tail, lower tail) and enables visualization of the fitting.

Another powerful characteristic of the K-moment framework, as presented above, is the simple and intuitive definition in terms of averages of maxima or minima of stochastic variables. This is different from other moment types and implies different behaviours (e.g., in discrete stochastic variables).

The fact that the K-moment framework assigns probabilities to moment values, which is a unique feature among all moment types, makes theoretical calculations thereof (i.e., by integration or numerical integration) unnecessary. Instead, the calculations are made in terms of the probabilities rather than values of the random variable. The simple approximation by Equation (71) is sufficient for most statistical tasks. When more accurate approximations are required, these can be found in [2].

Finally, the K-moment estimators, unlike other moment types, can explicitly (albeit approximately) take into account any existing dependence structure in case the data are not an independent sample but a time series (i.e., a realization of a stochastic process). This is also a unique feature among all types of moments.

5. Summary and Conclusions

Knowable moments or K-moments, defined as expectations of maxima or minima of a number of independent identically distributed stochastic variables, have some features that are not seen in other types of moments in stochastics and make them useful in relevant applications. Their name highlights the fact that, in contrast to classical moments, which are unknowable from samples for orders beyond 3–4, K-moments can be reliably and unbiasedly estimated from a sample for very high orders up to the sample size, even if this is several million or more. For such high orders, techniques to accelerate the calculations are proposed.

Not only can K-moments replace the information contained in classical moments and L-moments (e.g., summary statistics), but they can also provide information about what a classical moment estimator determines, which actually is not the true value of that classical moment.

An important feature of K-moment values is their ability to be assigned values of the distribution function. This characteristic is shared by order statistics and not by any other type of moments. However, there are substantial advantages over order statistics, because K-moment estimators make optimal use of the entire dataset, rather than relying on a single data value to be assigned a probability. This makes the estimation of the probability density function from a sample possible, a feature not shared by any other type of moments.

Finally, K-moments offer the unique advantage of taking into account the estimation bias when the data are not an independent sample but a time series from a process with dependence, even of long range. Both for samples and time series, the K-moment concept offers a strategy of model fitting, including its visualization, that is not shared with any other method. This enables utilization of the highest possible moment orders, which are particularly useful in modelling extreme highs or lows that are closely associated with high-order moments.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/axioms12060590/s1>. A spreadsheet that illustrates the calculations to produce Figure 1.

Funding: This research received no external funding but was conducted for scientific curiosity.

Data Availability Statement: Not applicable.

Acknowledgments: I acknowledge the comments in three reviews, which triggered improvement of the study.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof of Equations (46) and (49)

In the case that q is odd we have

$$F_y(y) = P\{\underline{y} \leq y\} = P\{(x - c)^q \leq y\} = P\{x \leq c + y^{1/q}\} = F(c + y^{1/q}) \tag{A1}$$

and

$$f_y(y) = \frac{f(c + y^{1/q})}{qy^{1-1/q}} \tag{A2}$$

Hence the K-moment of order p of \underline{y} will be

$$K'_{p_y} = pE\left[F_y(\underline{y})^{p-1} \underline{y}\right] = p \int_{-\infty}^{\infty} F_y(y)^{p-1} y f_y(y) dy \tag{A3}$$

and by setting $x = c + y^{1/q}$

$$K'_{p_y} = p \int_{-\infty}^{\infty} F(x)^{p-1} (x - c)^q \frac{f(x)}{q(x - c)^{q-1}} q(x - c)^{q-1} dx \tag{A4}$$

from which we find Equation (46).

If q is even, then we have

$$\begin{aligned} F_y(y) &= P\{\underline{y} \leq y\} = P\{-y^{1/q} \leq x - c \leq y^{1/q}\} = P\{c - y^{1/q} \leq x \leq c + y^{1/q}\} \\ &= F(c + y^{1/q}) - F(c - y^{1/q}) \end{aligned} \tag{A5}$$

and

$$f_y(y) = \frac{f(c + y^{1/q}) + f(c - y^{1/q})}{qy^{1-1/q}} \tag{A6}$$

Hence the K-moment of order p of \underline{y} will be

$$K'_{p_y} = pE\left[F_y(\underline{y})^{p-1} \underline{y}\right] = p \int_0^{\infty} F_y(y)^{p-1} y f_y(y) dy \tag{A7}$$

and by setting $t = y^{1/q}$

$$\begin{aligned} K'_{p_y} &= p \int_0^{\infty} (F(c + t) - F(c - t))^{p-1} t^q \frac{f(c+t)+f(c-t)}{qt^{q-1}} qt^{q-1} dx \\ &= \int_0^{\infty} (F(c + t) - F(c - t))^{p-1} t^q (f(c + t) + f(c - t)) dt = \\ &= p \int_0^{\infty} (F(c + t) - F(c - t))^{p-1} t^q f(c + t) dt \\ &\quad + p \int_0^{\infty} (F(c + t) - F(c - t))^{p-1} t^q f(c - t) dt \end{aligned} \tag{A8}$$

By substituting $t \rightarrow -t$ in the last term we find that it can be written as

$$p \int_{-\infty}^0 (F(c+|t|) - F(c-|t|))^{p-1} t^q f(c+t) dt \quad (\text{A9})$$

Hence Equation (A8) yields

$$K'_{p,y} = p \int_{-\infty}^{\infty} (F(c+|t|) - F(c-|t|))^{p-1} t^q f(c+t) dt \quad (\text{A10})$$

and by setting $x = t + c$ we find Equation (49).

References

1. Koutsoyiannis, D. Knowable moments for high-order stochastic characterization and modelling of hydrological processes. *Hydrol. Sci. J.* **2019**, *64*, 19–33. [CrossRef]
2. Koutsoyiannis, D. *Stochastics of Hydroclimatic Extremes—A Cool Look at Risk*, 2nd ed.; Kallipos Open Academic Editions: Athens, Greece, 2022; 346p, ISBN 978-618-85370-0-2. [CrossRef]
3. Koutsoyiannis, D. Replacing histogram with smooth empirical probability density function estimated by K-moments. *Sci* **2022**, *4*, 50. [CrossRef]
4. Koutsoyiannis, D. Advances in stochastics of hydroclimatic extremes. *L'Acqua* **2021**, 23–32.
5. Koutsoyiannis, D.; Montanari, A. Bluecat: A local uncertainty estimator for deterministic simulations and predictions. *Water Resour. Res.* **2022**, *58*, e2021WR031215. [CrossRef]
6. Koutsoyiannis, D.; Iliopoulou, T. Ombrian curves advanced to stochastic modeling of rainfall intensity. In *Rainfall Modeling, Measurement and Applications*; Chapter 9; Elsevier: Amsterdam, The Netherlands, 2022; pp. 261–283.
7. Iliopoulou, T.; Malamos, N.; Koutsoyiannis, D. Regional ombrian curves: Design rainfall estimation for a spatially diverse rainfall regime. *Hydrology* **2022**, *9*, 67. [CrossRef]
8. Iliopoulou, T.; Koutsoyiannis, D. A parsimonious approach for regional design rainfall estimation: The case study of Athens. In *"Innovative Water Management in a Changing Climate"*, *Proceedings of the 7th IAHR Europe Congress, Athens, Greece, 7–9 September 2022*; International Association for Hydro-Environment Engineering and Research (IAHR): Beijing, China, 2022.
9. Koutsoyiannis, D.; Iliopoulou, T.; Koukouvinos, A.; Malamos, N.; Mamassis, N.; Dimitriadis, P.; Tepetidis, N.; Markantonis, D. Τεχνική Έκθεση [Technical Report; in Greek]. In *Production of Maps with Updated Parameters of the Ombrian Curves at Country Level (Implementation of the EU Directive 2007/60/EC in Greece)*; Department of Water Resources and Environmental Engineering—National Technical University of Athens: Athens, Greece, 2023.
10. Regional Ombrian Curves: Design Rainfall Estimation for a Spatially Diverse Rainfall Regime—Prehistory, Rejection by Frontiers. Available online: <https://www.itia.ntua.gr/en/getfile/2188/2/documents/PrehistoryRejectionFromFrontiers.pdf> (accessed on 8 May 2023).
11. Koutsoyiannis, D. An open letter to the Editor of Frontiers, ResearchGate. 2021. Available online: <https://doi.org/10.13140/RG.2.2.34248.39689> (accessed on 8 May 2023). [CrossRef]
12. Hemelrijk, J. Underlining random variables. *Stat. Neerl.* **1966**, *20*, 1–7. [CrossRef]
13. Oberhettinger, F. *Tables of Mellin Transforms*; Springer: Berlin/Heidelberg, Germany, 1974; 275p.
14. Koutsoyiannis, D.; Dimitriadis, P. Towards generic simulation for demanding stochastic processes. *Sci* **2021**, *3*, 34. [CrossRef]
15. Fornberg, B. Generation of finite difference formulas on arbitrarily spaced grids. *Math. Comput.* **1988**, *51*, 699–706. [CrossRef]
16. Hosking, J.R. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. B* **1990**, *52*, 105–124. [CrossRef]
17. Weibull, W. A statistical theory of strength of materials. *Ing. Vetensk. Akad. Handl.* **1939**, *151*, 1–45.
18. Hazen, A. The storage to be provided in impounding reservoirs for municipal water supply. *Trans. Am. Soc. Civ. Eng.* **1914**, *77*, 1539–1669. [CrossRef]
19. Blom, G. *Statistical Estimates and Transformed Beta Variables*; Wiley: New York, NY, USA, 1958.
20. Royston, J.P. Expected normal order statistics (exact and approximate). *J. R. Stat. Soc. Ser. C* **1982**, *31*, 161–165.
21. Anderson, C.W. Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *J. Appl. Probab.* **1970**, *7*, 99–113. [CrossRef]
22. Hitz, A.; Davis, R.; Samorodnitsky, G. Discrete extremes. *arXiv* **2017**, arXiv:1707.05033.
23. Dimitriadis, P.; Koutsoyiannis, D. Stochastic synthesis approximating any process dependence and distribution. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1493–1515. [CrossRef]
24. Koutsoyiannis, D. A random walk on water. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 585–601. [CrossRef]
25. Hurst, H.E. Long term storage capacities of reservoirs. *Trans. Am. Soc. Civil Eng.* **1951**, *116*, 776–808. [CrossRef]

26. Koutsoyiannis, D. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* **2002**, *47*, 573–595. [[CrossRef](#)]
27. Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C.; Wallis, J.R. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* **1979**, *15*, 1049–1054. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.