Probability in the Engineering and Informational Sciences (2024), **xx**:xx 1–25

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Hausdorff Moment Transforms and Their Performance

Xinyun Wang[1] and Martin Haenggi[1]

[1]Department of Electrical Engineering, University of Notre Dame, Notre Dame, 46556, Indiana, USA.

Corresponding author: Xinyun Wang; E-mail: xwang54@nd.edu

Keywords: Truncated Hausdorff moment problem; Numerical approximation; Orthogonal polynomial; Moment sequences

MSC 2020: 44A60; 33F05;33C45

Abstract

The problem of reconstructing a distribution with bounded support from its moments is practically relevant in many fields, such as chemical engineering, electrical engineering, and image analysis. The problem is closely related to a classical moment problem, called the truncated Hausdorff moment problem (THMP). We call a method that finds or approximates a solution to the THMP a Hausdorff moment transform (HMT). In practice, selecting the right HMT for specific objectives remains a challenge. This study introduces a systematic and comprehensive method for comparing HMTs based on accuracy, computational complexity, and precision requirements. To enable fair comparisons, we present approaches for generating representative moment sequences. The study also enhances existing HMTs by reducing their computational complexity. Our findings show that the performances of the approximations differ significantly in their convergence, accuracy, and numerical complexity and that the decay order of the moment sequence strongly affects the accuracy requirement.

1. Introduction

In the analysis of large-scale wireless network performance, a critical metric, known as the meta distribution (MD), has emerged as a cornerstone in revealing the performance of individual links [5, 7, 8]. MDs play a key role of dissecting the impact of different sources of randomness on network performance. Quantities that affect the quality of the service of the users, such as signal power, signal-to-interference ratio (SIR), instantaneous rate, or energy, are inherently random, and we call them performance random variables (PRVs) [28]. Because of the nature of wireless networks, the PRVs strongly depend on user locations and are also time-varying, in other words, different sources of randomness. In this context, the MD emerges as a powerful tool to disentangle the different sources of randomness by analyzing the distribution of conditional probabilities.

However, calculating MDs is far from trivial. In some simplified network models, closed-form expressions can be obtained [7], but in more complete models, directly deriving analytic expressions is most likely impossible [5, 26, 29]. Thankfully, with the tool of stochastic geometry, the moments of the underlying conditional probabilities can be calculated [5, 26, 29]. Thus, we take a detour to calculate the distribution from its integer moments, which is a classical moment problem, called the Hausdorff moment problem (HMP) [9, 14, 21, 23]. The HMP seeks to ascertain whether a particular sequence corresponds to the integer moments of a valid random variable [9, 21, 23], in other words, the existence of a solution (as a random variable) whose integer moments match a given sequence. The existence condition is that the given sequence is completely monotonic (c.m.) [9, 21, 23]. When this criterion is met, a solution exists, and it is unique [9, 21, 23]. The HMP is also relevant in many fields other than the MDs in wireless networks, such as the density of states in a harmonic solid (see, e.g., [16]), and particle size distributions (see, e.g., [11]). From a probability theory perspective, the existence of a solution is

the primary concern, however, in practical applications, we know a priori that solutions exist, since the sequences are derived from random variables. Therefore, the target is to find the solution–in other words, to invert the mapping from a probability distribution to its moments. Given any tolerance greater than 0, we can approximate the solution within this tolerance (as measured by a suitable distance metric) using finitely many integer moments. Related, in practical scenarios, it is possible that only a finite number of integer moments can be calculated. Both aspects lead us to consider the truncated HMP (THMP).

The THMP diverges from the HMP in its objective. The former seeks solutions or approximations when the given finite-length sequence is from a valid moment sequence [21, 23], while the latter focuses on the existence and uniqueness of the solution for a given infinite-length sequence [23]. Further complicating matters, the THMP usually yields infinitely many solutions when the given sequence is valid [21, 23]. More importantly, solutions or approximations to the THMP are approximations to the HMP. We call a method that finds a solution or approximates a solution to the THMP a *Hausdorff moment transform* (HMT).

In the context of practical applications, such as MDs, the density of states in a harmonic solid, and particle size distributions, the sequences applied to the THMP can theoretically be extended to any length, therefore, our primary focus centers on the ultimate behavior of the HMTs as the lengths of the sequences increases. Various HMTs to the THMP have been proposed in the literature [4, 14, 16, 18, 22, 23, 24], but there is a gap between developing the HMTs and applying them to solve practical problems. More specifically, the choice of the appropriate HMT is not clear. Different HMTs may exhibit varying effectiveness in different scenarios, giving rise to the need for a method of comparison that allows fair evaluation[1] and informed decision-making.

In this study, we bridge this gap by introducing a systematic and comprehensive method of comparison for HMTs, based on their accuracy, computational complexity, and precision requirements. This way, the most suitable HMT for given objectives can be chosen. The key point to make fair comparisons is to generate representative moment sequences. Generating random moment sequences is non-trivial since the probability that an unsorted sequence of length n with each element independent and identically distributed (i.i.d.) uniformly on $[0, 1]$ is a moment sequence is approximately 2^{-n^2} [1, Eq. 1.2] and the probability that a sorted sequence of length n with each element i.i.d. uniformly on $[0, 1]$ is a moment sequence is approximately $n!2^{-n^2}$. We provide three different ways to solve it: One is to generate random integer moment sequences that are uniformly distributed in the moment space; the other two are based on analytic expressions, either of moments or of the cdfs. Moreover, our contributions extend to the enhancement of existing HMTs on the implementation front, improving them with heightened efficiency and practical applicability. In particular, we write some of the HMTs as linear transforms, which significantly lower their computational complexity and enables carrying out the most computationally intensive tasks offline.

2. Preliminaries

2.1. The Hausdorff moment problem

Without loss of generality, we assume the support of the distributions is $[0, 1]$. The THMP can be formulated as follows. Let $[n] \triangleq \{1, 2, ..., n\}$ and $[n]_0 \triangleq \{0\} \cup [n]$. Given a sequence $\mathbf{m} \triangleq (m_k)_{k=0}^n$, $n \in \mathbb{N}$, find or approximate an F such that

$$\int_0^1 x^k \, dF(x) = m_k, \quad \forall k \in [n]_0, \tag{2.1}$$

where F is right-continuous and increasing with $F(0^-) = 0$ and $F(1) = 1$, i.e., F is a cumulative distribution function (cdf).

[1] [11] has compared some HMTs but merely based on specific chemical engineering problems, which may not be applicable to other problems.

Although in practical applications, we are not concerned with the existence and uniqueness of the solutions to the THMP, we still find it valuable to examine these conditions to account for scenarios where the moments may not be exact.

Let $\mathcal{F}_n$ be the family of F that solves (2.1). The existence and uniqueness of solutions can be verified by the Hankel determinants, defined by

$$\underline{H}_l \triangleq \begin{cases} \begin{vmatrix} m_0 & \dots & m_{l/2} \\ \vdots & \ddots & \vdots \\ m_{l/2} & \dots & m_l \end{vmatrix}, & \text{even } l, \\ \begin{vmatrix} m_1 & \dots & m_{(l+1)/2} \\ \vdots & \ddots & \vdots \\ m_{(l+1)/2} & \dots & m_l \end{vmatrix}, & \text{odd } l, \end{cases} \tag{2.2}$$

$$\overline{H}_l \triangleq \begin{cases} \begin{vmatrix} m_1 - m_2 & \dots & m_{l/2} - m_{l/2+1} \\ \vdots & \ddots & \vdots \\ m_{l/2} - m_{l/2+1} & \dots & m_{l-1} - m_l \end{vmatrix}, & \text{even } l, \\ \begin{vmatrix} m_0 - m_1 & \dots & m_{(l-1)/2} - m_{(l+1)/2} \\ \vdots & \ddots & \vdots \\ m_{(l-1)/2} - m_{(l+1)/2} & \dots & m_{l-1} - m_l \end{vmatrix}, & \text{odd } l. \end{cases} \tag{2.3}$$

A solution exists if $\underline{H}_l$ and $\overline{H}_l$ are non-negative for all $l \in [n]$ [25]. Furthermore, the solution is unique, i.e., $|\mathcal{F}_n| = 1$, if and only if there exists an $l \in [n]$ such that $\underline{H}_l = 0$ or $\overline{H}_l = 0$ [25]. If the solution is unique, $\underline{H}_k = \overline{H}_k = 0$ for all $k > l$, and the distribution is discrete and uniquely determined by the first l moments [25]. Example 2.1 shows an example with $n = 2$. A more interesting and relevant case is when the Hankel determinants $\underline{H}_k$ and $\overline{H}_k$ are positive for all $l \in [n]$, which is both a necessary and sufficient condition for the problem to have infinitely many solutions, i.e., $|\mathcal{F}_n| = \infty$. In the following, we focus on this case by assuming the positiveness of all the Hankel determinants.

Example 2.1 (Two sequences with $n = 2$)**.**

1. For $\mathbf{m} = (1, 1/2, 1/2)$, we have

$$\underline{H}_1 = m_1 = 1/2 > 0, \quad \overline{H}_1 = m_0 - m_1 = 1/2 > 0,$$
$$\underline{H}_2 = \begin{vmatrix} m_0 & m_1 \\ m_1 & m_2 \end{vmatrix} = 1/4 > 0, \quad \overline{H}_2 = m_1 - m_2 = 0.$$

   The uniquely determined discrete distribution has only two jumps, one at 0 and one at 1. Hence the probability density function (pdf) is $\frac{1}{2}\delta(x) + \frac{1}{2}\delta(x-1)$.
2. For $\mathbf{m} = (1, 1/2, 1/3)$, we have

$$\underline{H}_1 = m_1 = 1/2 > 0, \quad \overline{H}_1 = m_0 - m_1 = 1/2 > 0,$$
$$\underline{H}_2 = \begin{vmatrix} m_0 & m_1 \\ m_1 & m_2 \end{vmatrix} = 1/12 > 0, \quad \overline{H}_2 = m_1 - m_2 = 1/6 > 0.$$

There are infinitely many solutions since all Hankel determinants are positive, such as $f(x) = 1$ and $f(x) = \frac{1}{2}\delta\left(x - \left(3 + \sqrt{3}\right)/6\right) + \frac{1}{2}\delta\left(x - \left(3 - \sqrt{3}\right)/6\right)$.

2.2. Adjustment

Here, we introduce a mapping to restore properties of probability distributions.

Some approximations are essentially functional approximations which may not preserve the properties of cdfs, such as monotonicity and ranges restricted to $[0, 1]$. An example can be found in Figure 5. Besides, for practical purposes, the numerical evaluation of approximations are necessarily to be restricted to a finite set of $[0, 1]$. Here, we consider a finite set $\mathcal{U}_n$ with cardinality $(n + 1)$, and let $(u_0, u_1, u_2, ..., u_n)$ be the sequence of all elements in $\mathcal{U}_n$ in ascending order, where $u_0 = 0$ and $u_n = 1$. If there is no further specification, we assume $\mathcal{U}_n$ is the uniform set $\{\frac{i}{n}, i \in [n]_0\}$. Definition 2.2 introduces a tweaking mapping to adjust the values of the approximations so that they adhere to the properties of cdfs after adjustments.

Definition 2.2 (Tweaking mapping)**.** We define T that maps $F : \mathcal{U}_n \mapsto \mathbb{R}$ to $\hat{F} : \mathcal{U}_n \mapsto [0, 1]$ such that

1. $T(F)(u_i) = \min(1, \max(0, F(u_i))), \forall i \in [n]_0$.
2. $T(F)(0) = 0$ and $T(F)(1) = 1$.
3. $T(F)(u_{i+1}) = \max(F(u_i), F(u_{i+1})), \forall i \in [n-1]_0$.

The tweaking mapping in Definition 2.2 eliminates the out-of-range points and keeps the partial monotonicity. Let $F : \mathcal{U}_n \mapsto \mathbb{R}$ be an approximation of a solution to the truncated HMP. After tweaking, we obtain monotone and bounded approximations over a discrete set, denoted as $\hat{F} = T(F)$. Next, we interpolate the discrete points to obtain continuous functions. Specifically, we apply monotone cubic interpolation to $\hat{F}$. Monotone cubic interpolation preserves monotonicity. For convenience, we refer to the function after interpolation as $\hat{F}$ as well, and we call the original function F the *raw* function and the tweaked and interpolated function $\hat{F}$ the *polished* function. We use $\hat{F}|_{\mathcal{U}_n}$ to emphasize that the polished function $\hat{F}$ is originally sampled over $\mathcal{U}_n$. The polished function is continuous, monotone and $[0, 1] \mapsto [0, 1]$.

2.3. Distance metrics

In this part, we introduce two distance metrics between two functions $F_1, F_2 : [0, 1] \mapsto [0, 1]$ are.

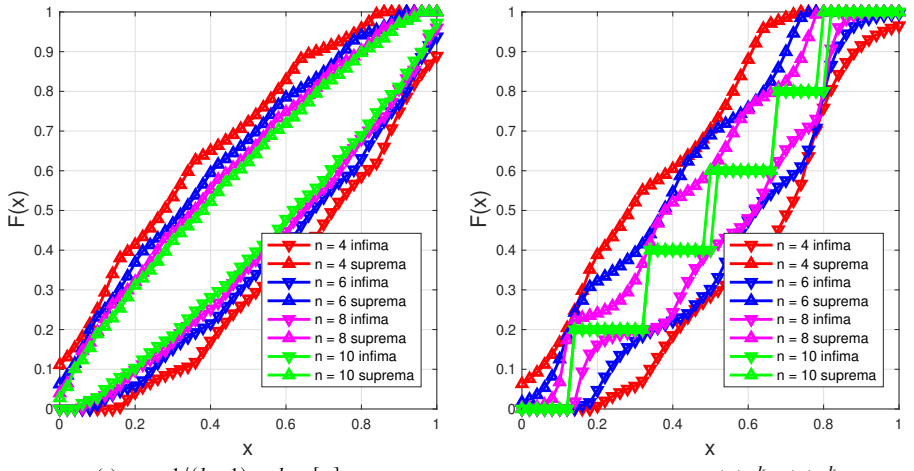
Definition 2.3 (Maximum and total distance)**.** The maximum distance between functions F_1 and F_2 is defined as the $\mathcal{L}_\infty$-norm of F_1 and F_2 over $[0, 1]$, i.e.,

$$\begin{aligned} d_{\mathrm{M}}(F_1, F_2) &\triangleq \|F_1 - F_2\|_{\mathcal{L}_\infty[0,1]}, \\ &= \max_{x \in [0,1]} |F_1(x) - F_2(x)|. \end{aligned}$$

The total distance between functions F_1 and F_2 is defined as the $\mathcal{L}_1$-norm of F_1 and F_2 over $[0, 1]$, i.e.,

$$\begin{aligned} d(F_1, F_2) &\triangleq \|F_1 - F_2\|_{\mathcal{L}_1[0,1]}, \\ &= \int_0^1 |F_1(x) - F_2(x)| dx. \end{aligned}$$

We mainly focus on the total distance and use the maximum distance as a secondary metric. As the maximum and total distance are designed to be applied to the polished functions, the exact value of the total distance and maximum distance of the linear/monotone cubic interpolation of the polished functions can be calculated.



(a) $m_k = 1/(k+1), \quad k \in [n]_0$.

(b) $m_k = \frac{1}{5}(8^{-k} + 3^{-k} + 2^{-k} + \left(\frac{3}{2}\right)^{-k} + \left(\frac{5}{4}\right)^{-k}), \quad k \in [n]_0$.

Figure 1. *The infima and suprema from the CM inequalities for* $n = 4, 6, 8, 10$. x *is discretized to* $\mathcal{U}_{50} = \{i/50, i \in [50]_0\}$..

3. Solutions and approximations

3.1. The Chebyshev-Markov (CM) inequalities

[23] provides a method for obtaining $\inf_{F \in \mathcal{F}_n} F(x_0)$ and $\sup_{F \in \mathcal{F}_n} F(x_0)$ for any $x_0 \in [0, 1]$. The most important step of the method is the construction of a discrete distribution in $\mathcal{F}_n$ where the maximum possible mass is concentrated at x_0. The details of the method can be found in [14, 23] and explicit expressions can be found in [27, 31]. The inequalities established by the infima and suprema in [27, Thm. 1] are also known as the CM inequalities. In contrast, the infima and suprema in [20] are for distributions where the support can be an arbitrary Borel subset of $\mathbb{R}$. As a result, they are loose for distributions with support $[0, 1]$.

Figure 1 shows two examples of the infima and suprema for different n. The convergence of the infima and suprema as n increases is apparent. The discontinuities of the derivatives in Figure 1 coincide with the roots of the orthogonal polynomials [27, Thm. 1]. For Figure 1b, $\underline{H}_9 = 0$ indicates a discrete distribution. The infima and suprema get drastically closer from $n = 8$ to $n = 10$ because $n = 10$ is sufficient to characterize a discrete distribution with 5 jumps while $n = 8$ is not. In fact, the infima and suprema match almost everywhere (except at the jumps) for $n = 10$, but Figure 1b only reveals the mismatch at 0.5 and 0.8 since only 0.5 and 0.8 are jumps that fall in the sampling set $\mathcal{U}_{50} = \{i/50, i \in [50]_0\}$.

While all solutions to the THMP lie in the band defined by the infima and suprema, the converse does not necessarily hold. Figure 2 gives an example, where F lies in the band but is not a solution to the THMP. Its corresponding m_1 and m_2 are $2/3$ and $1/2$, respectively, rather than $1/2$ and $1/3$ that define the band.

In the following subsections, we propose a method based on the Fourier-Chebyshev (FC) series and a CM method based on the average of the infima and suprema from the CM inequalities and compare them against the existing methods. We also improve and concretize some of the existing methods. In particular, we provide a detailed discussion of the existing Fourier-Jacobi (FJ) approximations and fully develop the one mentioned in [23]. Besides, among all the discussed methods, the maximum entropy (ME) method by design yields solutions to the THMP, while the other methods by design obtain approximations of a solution to the THMP.

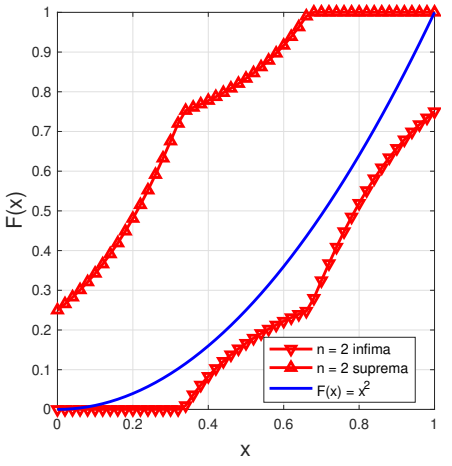


Figure 2. *The infima and suprema from the CM inequalities for* $m_n = 1/(n+1)$ *and* $F(x) = x^2$..

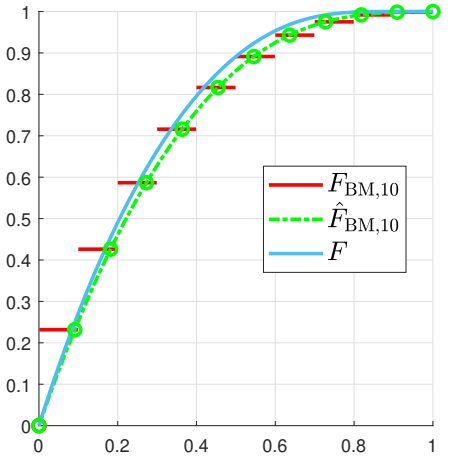


Figure 3. $F_{\mathrm{BM},10}$, $\hat{F}_{\mathrm{BM},10}$ *and* F, *where* $\bar{F} = \exp(-\frac{x}{1-x})(1-x)/(1-\log(1-x))$. *The* $\circ$ *denote* $F_{\mathrm{BM},10}|_{\mathcal{U}_{11}}$..

3.2. Well-established methods

3.2.1. Binomial mrixture (BM) method

A piecewise approximation of the cdf based on binomial mixtures[2] was proposed in [18]. For any positive integer n, the approximation[3] by the BM method is defined as follows.

Definition 3.1 (Approximation by the BM method)**.**

$$F_{\mathrm{BM},n}(x) \triangleq \begin{cases} \sum_{k=0}^{\lfloor nx \rfloor} h_k, & x \in (0,1], \\ 0, & x = 0, \end{cases} \tag{3.1}$$

where $h_k = \sum_{i=k}^{n} \binom{n}{i}\binom{i}{k}(-1)^{i-k} m_i$.

Note that $F_{\mathrm{BM},n} \to F$ as $n \to \infty$ for each x at which F is continuous [17, 18].

[2]Due to the continuous nature of the MDs, we do not consider other methods which find a solution to the THMP as a discrete distribution, such as Gauss quadrature.

[3]In this section, the term "approximations" is used when the outputs are not always solutions to the THMP.

Let $\hat{F}_{\mathrm{BM},n}$ denote the monotone cubic interpolation of $F_{\mathrm{BM},n}|_{\mathcal{U}_{n+1}}$ [6]. As shown in Figure 3, $F_{\mathrm{BM},n}$ is piecewise constant, and for a continuous F, $\hat{F}_{\mathrm{BM},n}$ provides a better approximation than $F_{\mathrm{BM},n}$. Note that $F_{\mathrm{BM},n}$ is a solution to the THMP only when $n = 1$ and $\hat{F}_{\mathrm{BM},n}$ is not necessarily a solution for any positive integer n.

3.2.2. ME method

The ME method [12, 16, 25] is also widely used to solve the THMP. For any positive integer n, the solution by the ME method is defined as follows.

Definition 3.2 (Solution by the ME method)**.**

$$f_{\mathrm{ME},n}(x) \triangleq \exp\left(-\sum_{k=0}^{n} \xi_k x^k\right), \tag{3.2}$$

where the coefficients $\xi_k, i \in [n]_0$, are obtained by solving the equations

$$\int_0^1 x^i \exp\left(-\sum_{k=0}^{n} \xi_k x^k\right) dx = m_i, \quad i \in [n]_0. \tag{3.3}$$

A unique solution of (3.3) exists if and only if the Hankel determinants of the moment sequence are all positive. It can be obtained by minimizing the convex function [25]

$$\Gamma(\xi_1, \ldots, \xi_n) \triangleq \sum_{k=1}^{n} \xi_k m_k + \log\left(\int_0^1 \exp\left(-\sum_{k=1}^{n} \xi_k x^k\right)\right) \tag{3.4}$$

and solving

$$\int_0^1 \exp\left(-\sum_{k=0}^{n} \xi_k x^k\right) dx = m_0 = 1 \tag{3.5}$$

for ξ_0. The corresponding cdf is given by

$$F_{\mathrm{ME},n}(x) = \int_0^x \exp\left(-\sum_{k=0}^{n} \xi_k v^k\right) dv. \tag{3.6}$$

Since (3.4) is convex, it can, in principle, be solved by standard techniques such as the interior point method. As the cdf given by (3.6) is a solution to the THMP, its error can be bounded by the infima and suprema from the CM inequalities, which, however, requires additional computation. $F_{\mathrm{ME},n} \to F$ in entropy and thus in $\mathcal{L}_1$ as $n \to \infty$ [25]. But due to ill-conditioning of the Hankel matrices [10], the ME method becomes impractical when the number of given moments is large, as solving (3.4) may not be feasible.

3.3. Improvements and concretization of existing methods

3.3.1. Gil-Pelaez (GP) method

The main part of this study focuses on the HMP and THMP, which are based on integer moments. However, for the purpose of performance evaluation in Section 4.5, here we also consider the reconstruction problem involving complex moments. By the Gil-Pelaez theorem [3], F at which F is continuous[4] can

[4]If F is not continuous at some point x_0, the result obtained by the Gil-Pelaez theorem is the average of $F(x_0)$ and $F(x_0^-)$.

be obtained from the imaginary moments by

$$F(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty r(s,x)ds, \quad x \in [0,1], \tag{3.7}$$

where $r(s,x) \triangleq \Im(e^{-js\log x}m_{js})/s$, $j \triangleq \sqrt{-1}$, and $\Im(z)$ is the imaginary part of the complex number z.

Since the integral in (3.7) cannot be analytically calculated, only numerical approximations can be obtained. An accurate evaluation is exacerbated by the facts that the integration interval is infinite and that the integrand may diverge at 0. We define the approximate value obtained through the trapezoidal rule as follows.

Definition 3.3 (Approximation by the GP method)**.**

$$F_{\mathrm{GP}(\Delta s,\upsilon)}(x) \triangleq \frac{1}{2} - \frac{\Delta s}{2\pi}\sum_{i=1}^{N}\left(r(s_i,x) + r(s_{i-1},x)\right), \quad x \in [0,1], \tag{3.8}$$

where $N \triangleq \lfloor \frac{\upsilon}{\Delta s} \rfloor$ and $s_i \triangleq i\Delta s$.

Note that as $\Delta s \to 0$ and $\upsilon \to \infty$, $F_{\mathrm{GP},\Delta s,\upsilon} \to F$ for each x at which F is continuous.

We evaluate $F_{\mathrm{GP}(\Delta s,\upsilon)}$ over the set $\mathcal{U}_{100}$ for different Δs and υ. Figure 4 shows the average of the total and maximum distance of F and polished $\hat{F}_{\mathrm{GP}(\Delta s,\upsilon)}$ as the step size Δs and upper integration limit υ change. As the step size Δs decreases to 0.1 and as the upper integration limit υ increases to 1000, the average of the total distance decreases to less than 10^{-3}, which indicates that the approximate cdf obtained by numerically evaluating the integral of Gil-Pelaez method is very close to the exact cdf and thus can be used as a reference. Both the integration upper limit and step size are critical in evaluating the integral. A general rule for selecting the upper integration limit is to examine the magnitude of m_{js}. For example, if $|m_{js}| \le 1/s$, the error resulting from fixing the upper integration limit at 1000 is $\int_{1000}^{\infty} r(s,x)ds < \int_{1000}^{\infty} 1/s^2 ds = 10^{-3}$. However, it is not always possible to determine the rate of decay of the magnitude of the imaginary moments with only numerical values. So it is difficult to come up with a universal rule for the integration range that is suitable for all moment sequences.

For the GP method, we propose a dynamic way to adjust the values of the upper limit and step size. We start at $F_{\mathrm{GP}(0.1,1000)}$. If the cdf determined by the previous selections exceeds $[0,1]$, we decrease the step size to $1/3$ of the previous one and increase the upper limit to 3 times the previous one, and then recalculate the cdf. We repeat this process until the range of the cdf is in $[0,1]$. In some cases, the updated step size and upper limit may cause memory problems. In the averaging process described in Section 4.5, we discard those cases if we detect that the required memory is larger than the available memory.

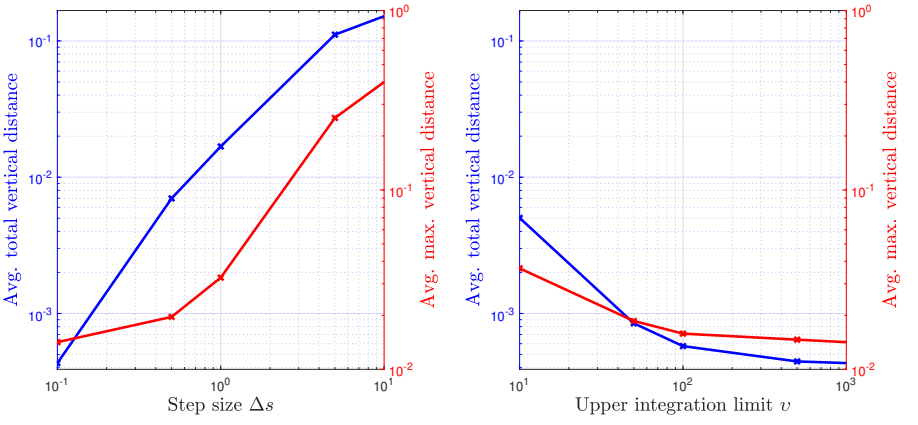
3.3.2. FJ method

Equipped with a complete orthogonal system on $[0,1]$, denoted as $(T_k(x))_{k=0}^{\infty}$, we can expand a function l defined on $[0,1]$ to a generalized Fourier series as [24]

$$l(x) = \sum_{k=0}^{\infty} c_k T_k(x). \tag{3.9}$$

The Jacobi polynomials $P_n^{(\alpha,\beta)}$ are such a class of orthogonal polynomials defined on $[-1,1]$. They are orthogonal w.r.t. the weight function $\omega^{(\alpha,\beta)}(x) \triangleq (1-x)^{\alpha}(x+1)^{\beta}$ [24], i.e.,

$$\int_{-1}^{1} \omega^{(\alpha,\beta)}(x) P_n^{(\alpha,\beta)}(x) P_m^{(\alpha,\beta)}(x)\,dx = \frac{2^{\alpha+\beta+1}\,(n+1)^{(\beta)}\,\delta_{mn}}{(2n+\alpha+\beta+1)\,(n+\alpha+1)^{(\beta)}}, \tag{3.10}$$



(a) $\upsilon = 1000$ (b) $\Delta s = 0.1$

Figure 4. *Average of the total and maximum distances between F and polished $\hat{F}_{\mathrm{GP}(\Delta s,\upsilon)}$ which are averaged over* 100 *randomly generated beta mixtures as described in Section 4.4. .*

where δ_{mn} is the Kronecker delta function, and $a^{(b)} \triangleq \frac{\Gamma(a+b)}{\Gamma(a)}$ is the Pochhammer function. For $\alpha = \beta = 0$, they are the Legendre polynomials; for $\alpha = \beta = \pm 1/2$, they are the Chebyshev polynomials. The shifted Jacobi polynomials

$$R_n^{(\alpha,\beta)}(x) \triangleq P_n^{(\alpha,\beta)}(2x-1) = \sum_{j=0}^{n} \binom{n+\alpha}{j}\binom{n+\beta}{n-j}(x-1)^{n-j}x^j, \tag{3.11}$$

are defined on $[0, 1]$. Letting $\eta_n^{(\alpha,\beta)} \triangleq \frac{(n+1)^{(\beta)}}{(2n+\alpha+\beta+1)(n+\alpha+1)^{(\beta)}}$, the corresponding orthogonality condition w.r.t. the weight function $w^{(\alpha,\beta)}(x) \triangleq (1-x)^\alpha x^\beta$ is

$$\int_0^1 w^{(\alpha,\beta)}(x) R_n^{(\alpha,\beta)}(x) R_m^{(\alpha,\beta)}(x)\, dx = \eta_n^{(\alpha,\beta)}\delta_{mn}. \tag{3.12}$$

When applying the FJ series to solve (2.1), the usual approach is to expand the target function (cdf or pdf) as

$$w^{(\alpha,\beta)}(x) \sum_{k=0}^{\infty} c_k R_k^{(\alpha,\beta)}(x), \tag{3.13}$$

so that the moments can be utilized in the orthogonality conditions. However, an approximation obtained by the n-th partial sum of the expansion is usually not a solution to the THMP, for two reasons. First, the remaining Fourier coefficients $c_m, m > n$, may not all be zero. Second, the approximations may not preserve the properties of probability distributions.

Convergence is a major concern when approximating a function by an FJ series. While no conditions are known that are both sufficient and necessary, there exist sufficient conditions and there exist necessary conditions, both of which depend on the values of α, β and the properties of the approximated functions [13, 19]. For the THMP, a basic question is whether α, β should depend on the moment sequence or not and whether to approximate pdfs or cdfs. In the following, we will discuss in detail three different ways which have been proposed in [4, 22, 23], respectively. In particular, [4] set α, β depending on the moment sequences and only approximate pdfs. In the other two ways, cdfs and pdfs are approximated, respectively, and both fix $\alpha = \beta = 0$ [22, 23].

In the FJ method based on the FJ series [4], α, β are chosen such that the first two Fourier coefficients c_1 and c_2 are zero, i.e., $\alpha + 1 = (m_1 - m_2)(1 - m_1)/(m_2 - m_1^2)$ and $\beta + 1 = (\alpha + 1)m_1/(1 - m_1)$. For any positive integer n, the approximation by the FJ method is defined as follows.

Definition 3.4 (Approximation by the FJ method [4])**.**

$$f_{\mathrm{FJ},n}(x) \triangleq w^{(\alpha,\beta)}(x) \sum_{k=0}^{n} c_k R_k^{(\alpha,\beta)}(x), \tag{3.14}$$

where, by the orthogonality condition,

$$c_m = \frac{1}{\eta_m^{(\alpha,\beta)}} \sum_{j=0}^{m} \binom{m+\alpha}{j} \binom{m+\beta}{m-j} \sum_{k=0}^{m-j} \binom{m-j}{k} (-1)^k m_{m-k}. \tag{3.15}$$

The corresponding cdf is

$$F_{\mathrm{FJ},n}(x) = \frac{(\alpha+1)^{(\beta+1)}}{\Gamma(\beta+1)} \int_0^x w^{(\alpha,\beta)}(t)dt - \sum_{k=1}^{n} \frac{c_k}{k} w^{(\alpha+1,\beta+1)}(x) R_{k-1}^{(\alpha+1,\beta+1)}(x). \tag{3.16}$$

$F_{\mathrm{FJ},1} = F_{\mathrm{FJ},2}$ correspond to the beta approximation [5], where m_1 and m_2 are matched with the corresponding moments of a beta distribution.

There are three drawbacks of the FJ method. First, as α, β vary significantly for different moment sequences, the convergence properties of the FJ method vary strongly as well. So it is impossible to give a conclusive answer to whether such approximations converge or not. It is a critical shortcoming of the FJ method that it only converges in some cases. Second, as α and β are generally not zero, the approximate pdf is always 0 or ∞ at $x = 0$ and $x = 1$, which holds only for a limited class of pdfs. Last but not least, it is possible that, as a functional approximation, the FJ method leads to non-monotonicity in the approximate cdf. We address this by applying the tweaking mapping in Definition 2.2 to $F_{\mathrm{FJ},n}|_{\mathcal{U}_n}$.

3.3.3. Fourier-Legendre (FL) method

[22, 23] suggest fixing $\alpha = \beta = 0$ in advance. The resulting polynomials are Legendre polynomials with weight function 1. To expand a function l by the FL series, a necessary condition for convergence is that the integral $\int_0^1 l(x)(1-x)^{-1/4}x^{-1/4}\,dx$ exists [24]. Furthermore, if $l \in \mathcal{L}^p$ with $p > 4/3$, the FL series converges pointwise [19]. There still remains the question whether we should approximate the pdf or the cdf. In the following, we will discuss the two cases. The two have been discussed separately in [22, 23], but the concrete expressions of the coefficients are missing. Besides, to the best of our knowledge, their convergence properties have not been compared.

We denote the corresponding pdf of F as f and $R_k^{(0,0)}$ as R_k.

Approximating the pdf [22]

The n-th partial sum of the FL expansion of the pdf is

$$\sum_{k=0}^{n} c_k R_k(x), \tag{3.17}$$

where, by the orthogonality condition,

$$c_m = (2m+1) \sum_{j=0}^{m} \binom{m}{j} \binom{m}{m-j} \sum_{k=0}^{m-j} \binom{m-j}{k} (-1)^k m_{m-k}. \tag{3.18}$$

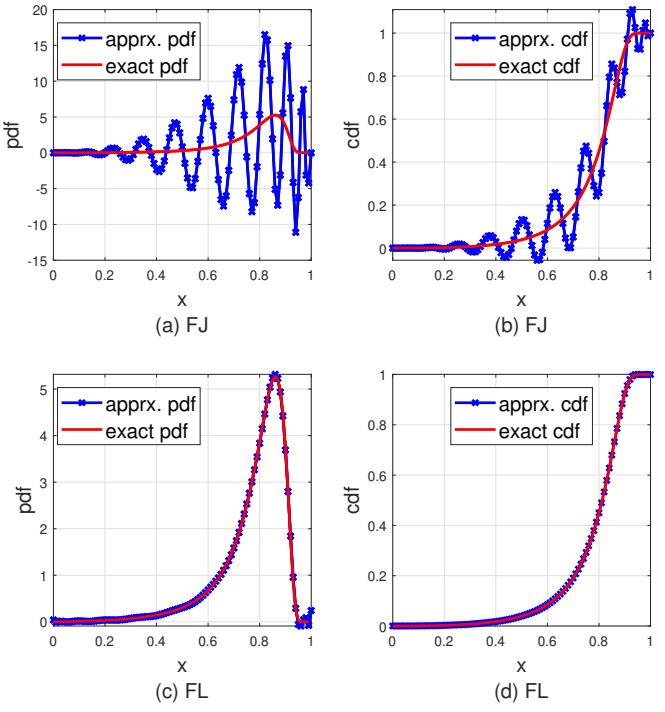


Figure 5. *The upper two plots are for the FJ method [4] with $n = 20$, and the lower two plots are for the FL series with $n = 20$. The cdf is given in (3.21).* .

It does not always hold that $f \in \mathcal{L}^p[0,1]$ with $p > 4/3$. Thus, the series does not converge for some f, for example, the pdf of Beta$(1,\beta)$ with $0 < \beta \le 1/4$ or Beta$(\alpha,1)$ with $0 < \alpha \le 1/4$. Worse yet, it is possible that the FL series diverges everywhere, e.g., the FL series of $f(x) = (1-x)^{-3/4}/4$.

So similar to the FJ method, approximating the pdf does not guarantee convergence.

Approximating the cdf [23]

The n-th partial sum of the FL expansion of the cdf is

$$\sum_{k=0}^{n} c_k R_k(x), \tag{3.19}$$

where, by the orthogonality condition,

$$c_m = (2m+1)\sum_{j=0}^{m}\binom{m}{j}\binom{m}{m-j}\sum_{k=0}^{m-j}\binom{m-j}{k}(-1)^k\frac{1-m_{m-k+1}}{m-k+1}. \tag{3.20}$$

Since F is always bounded and thus $F \in \mathcal{L}^p[0,1]$ for any $p \ge 1$, pointwise convergence holds [19].

Comparisons of the FJ method and approximating the pdf via the FL series

In Figures 5 and 6, where

$$F(x) = 1 - \frac{\exp\left(-\frac{x^2}{50(1-x)^2}\right)}{1+\frac{x^2}{50(1-x)^2}}, \tag{3.21}$$

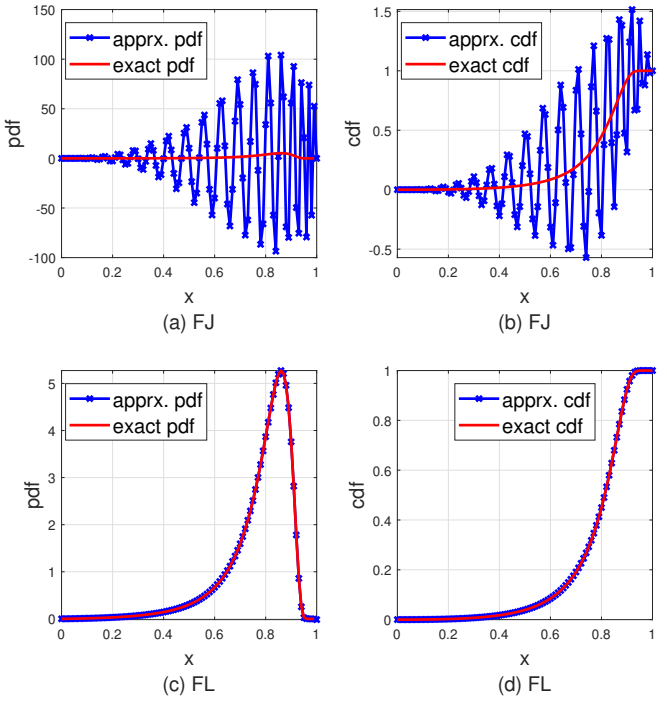


Figure 6. *The upper two plots are for FJ method [4] with* $n = 40$*, and the lower two plots are for the FL series with* $n = 40$*. The cdf is given in* (3.21)*. .*

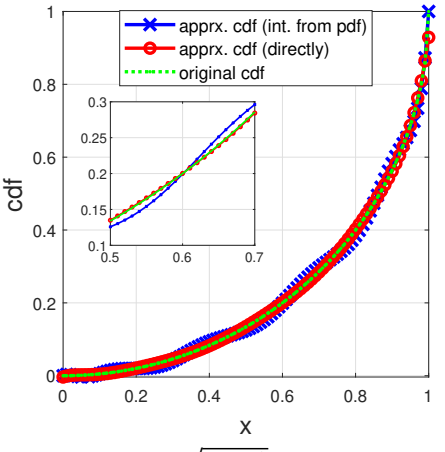


Figure 7. *Reconstructed cdfs of* $F(x) = 1 - \sqrt{1 - x^2}$ *based on* 10 *moments. The blue curve shows the reconstructed cdf from approximating the pdf, the red curve shows the reconstructed cdf from approximating the cdf, and the green curve shows* F*. .*

we approximate the pdf by the FJ method and by the FL series. We observe that in both cases, the FL approximations are much better than the FJ ones. Also, the FL approximations do better as n increases.

Comparisons of approximating the pdf and cdf by the FL series

Figure 7 shows the approximate cdfs by using the FL series to approximate the pdf and then integrate and to directly approximate the cdf, respectively. We observe that approximating cdfs leads to a more accurate result.

Considering that a pdf may not be in $\mathcal{L}^p[0,1]$ with $p > 4/3$ but a cdf always is, in the following, we use the FL series to approximate the cdf and call it the FL method. For any positive integer n, the approximation by the FL method is defined as follows.

Definition 3.5 (Approximation by the FL method)**.**

$$F_{\mathrm{FL},n}(x) \triangleq \sum_{m=0}^{n} c_m R_m(x), \tag{3.22}$$

where c_m, $m \in [n]_0$, are given by (3.20).

Since it is a functional approximation, we apply the tweaking mapping in Definition 2.2 to $F_{\mathrm{FL},n-1}|_{\mathcal{U}_n}$.

3.4. New methods

3.4.1. FC method

Similar to the idea in Section 3.3.3 that we fix α and β in advance, here we use the FC series for approximation, i.e., we set $\alpha = \beta = -1/2$. To guarantee convergence, similar to the FL method, we also focus on approximating the cdf and call it the FC method. For any positive integer n, the approximation by the FC method is defined as follows.

Definition 3.6 (Approximation by the FC method)**.**

$$F_{\mathrm{FC},n}(x) \triangleq \begin{cases} x^{-\frac{1}{2}}(1-x)^{-\frac{1}{2}} \sum_{m=0}^{n} c_m R_m^{\left(-\frac{1}{2},-\frac{1}{2}\right)}(x), & x \in (0,1), \\ x, & x \in \{0,1\}, \end{cases} \tag{3.23}$$

where, by the orthogonality condition,

$$c_m = c'_m \sum_{j=0}^{m} \binom{m-1/2}{j} \binom{m-1/2}{m-j} \sum_{k=0}^{m-j} \binom{m-j}{k} (-1)^k \frac{1-m_{m-k+1}}{m-k+1}, \tag{3.24}$$

$c'_0 = 1/\pi$ and $c'_m = 2\left(\frac{\Gamma(m+1)}{\Gamma(m+1/2)}\right)^2$, $m \in [n]$.

As F and $x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}$ are of bounded variation, the function $Fx^{\frac{1}{2}}(1-x)^{\frac{1}{2}}$ approximated by the FC series has bounded variation, thus uniform convergence holds [15]. To preserve the properties of probability distributions, we apply the tweaking mapping in Definition 2.2 to $F_{\mathrm{FC},n-1}|_{\mathcal{U}_n}$.

3.4.2. CM method

As the CM inequalities provide the infima and suprema at the points of interest among the family of solutions given a sequence of length n, it is important to analyze the distribution of the unique solution given the sequence of an infinite length. Here, we explore this distribution at point $x = 0.5$ by simulating the distribution of infima and suprema of a sequence of length 8 within the initial ranges define by the infima and suprema given the first three elements in the sequence. For example, let $\underline{F}_3$ and $\bar{F}_3$ denote the infimum and supremum defined by the CM inequalities at point 0.5 given a sequence (m_1, m_2, m_3), and let $\underline{F}_8$ and $\bar{F}_8$ denote the infimum and supremum defined by the CM inequalities at point 0.5 given a sequence $(m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8)$. We focus on the relative position of $\underline{F}_8$ and $\bar{F}_8$ within the range defined by $\underline{F}_3$ and $\bar{F}_3$, i.e., the normalized $x = (\underline{F}_8 - \underline{F}_3)/(\bar{F}_3 - \underline{F}_3)$ and $x = (\bar{F}_8 - \underline{F}_3)/(\bar{F}_3 - \underline{F}_3)$. The average is calculated as $(\bar{F}_8 + \underline{F}_8)/2$.

Figure 8 and Figure 9 show the normalized histograms with beta distribution fit. From both of them, we can observe that there is symmetry between the distribution of the infima and suprema, and their average is concentrated in the middle. In particular, in Figure 9, we observe that its average concentrates

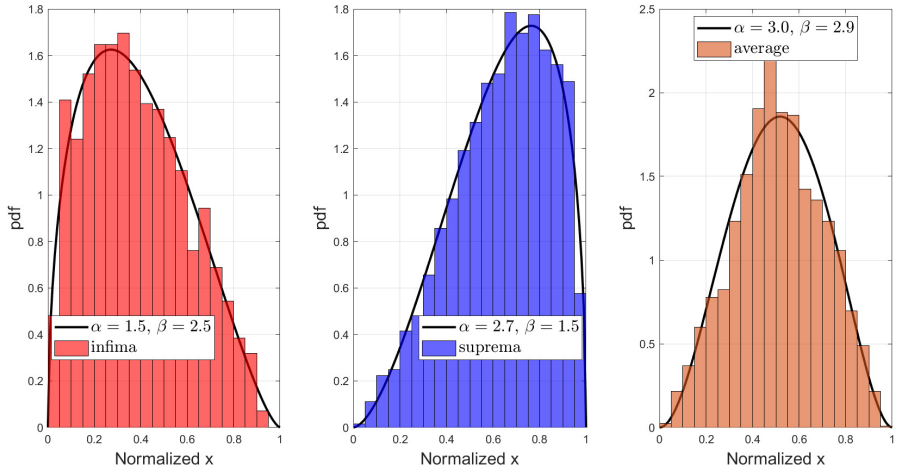


Figure 8. *The normalized histogram of infima, suprema, and average and the corresponding beta distribution fit for* 50 *realizations of moment sequences of length* 3 *generated by Algorithm 1, each of which has been further extended* 50 *times to length* 8 *by Algorithm 1. The two parameters of the beta distribution fit are based on the first two empirical moments. .*

more in the middle compared to Figure 8, since the first three moments come from the uniform distribution. This finding indicates that the unique solution of a moment sequence of infinite length is likely to be centered within the infima and suprema defined by its truncated sequence. Therefore, it is sensible to consider their average as an approximation to a solution to the THMP and we call it the CM method. For any positive integer n, the approximation by the CM method is defined as follows.

Definition 3.7 (Approximation by the CM method)**.**

$$F_{\mathrm{CM,n}}(x) \triangleq \frac{1}{2}\left(\sup_{F\in\mathcal{F}_n} F(x) + \inf_{F\in\mathcal{F}_n} F(x)\right), \tag{3.25}$$

where $\sup_{F\in\mathcal{F}_n} F(x)$ and $\inf_{F\in\mathcal{F}_n} F(x)$ are given by the CM inequalities.

From Figure 10, we observe that the average is very close to the original cdf, and it is slightly better than the beta approximation. It is worth noting that by averaging, the error can be bounded: For all $x \in [0,1]$ and for all $F \in \mathcal{F}_n$,

$$|F_{\mathrm{CM,n}}(x) - F(x)| \le \frac{1}{2}\left(\sup_{F\in\mathcal{F}_n} F(x) - \inf_{F\in\mathcal{F}_n} F(x)\right). \tag{3.26}$$

Although $F_{\mathrm{CM,n}}$ lies within the band defined by the infima and suprema from the CM inequalities, it is not necessarily one of the infinitely many solutions to the THMP. Moreover, it is almost never a solution. For example, $F_{\mathrm{CM},1}$ is a solution only when $m_1 = 1/2$.

3.5. The methods as linear transforms

The calculation of the coefficients in the BM method and methods based on the FJ series, such as the FL and FC method, where the coefficients α and β are fixed, can be written as linear transforms of

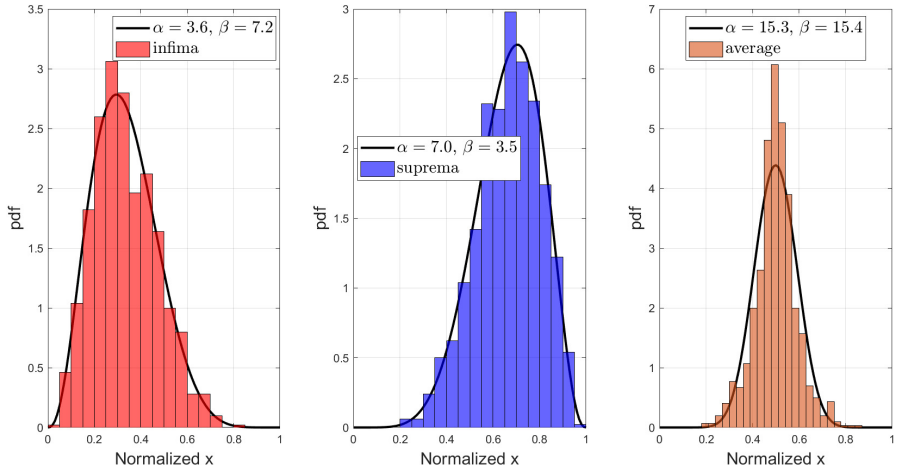
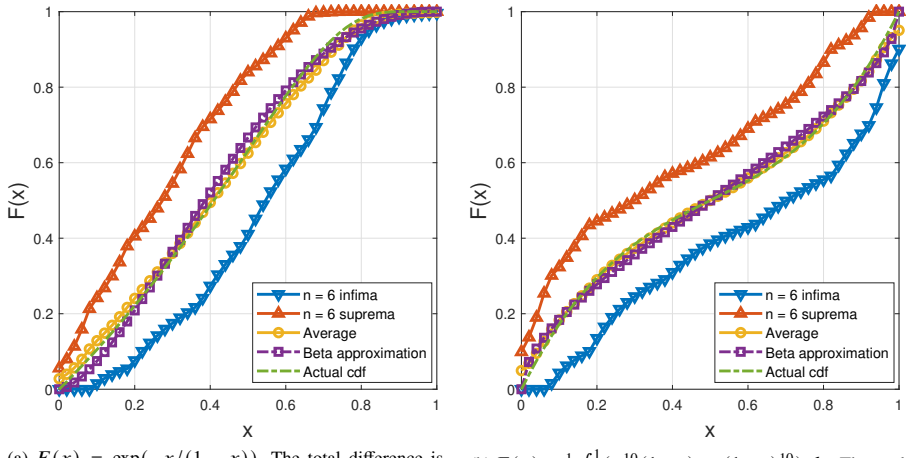


Figure 9. *The normalized histogram of infima, suprema, and average and the corresponding beta distribution fit for* 1000 *realizations of moment sequences of length* 8 *generated by Algorithm 1, each of which has the same first three elements as* $(1/2, 1/3, 1/4)$. *The two parameters of the beta distribution fit are based on the first two empirical moments. .*



(a) $F(x) = \exp(-x/(1-x))$. The total difference is 0.015 between $\hat{F}_{\mathrm{CM},6}$ and the actual cdf and 0.019 for the beta approximation.

(b) $F(x) = \frac{1}{2}\int_0^1 (x^{10}(1-x)+x(1-x)^{10})\,dx$. The total difference is 0.0199 between $\hat{F}_{\mathrm{CM},6}$ and the actual cdf and 0.0451 for the beta approximation.

Figure 10. *The infima and suprema from the CM inequalities, their average, and the beta approximation given* $(m_k)_{k=0}^6$. .

the given sequence, with the transform matrix being fixed for each method. This can be pre-computed offline, accelerating the calculation.[5] We discuss the BM and FL method in detail.

[5]The FJ method can also be written as a linear transform but doing so is meaningless due to the undetermined coefficients α and β.

As we recall, for the BM method [6], we can write $\mathbf{h} \triangleq (h_l)_{l=0}^n$ as a linear transform of $\mathbf{m} \triangleq (m_l)_{l=0}^n$, i.e.,

$$\mathbf{h} = \mathbf{A}\mathbf{m}, \tag{3.27}$$

where $\mathbf{m}$ and $\mathbf{h}$ are column vectors and the transform matrix $\mathbf{A} \in \mathbb{Z}^{(n+1)\times(n+1)}$ is given by

$$A_{ij} \triangleq \binom{n}{j}\binom{j}{i}(-1)^{j-i}\mathbb{1}(j \geq i), \quad i, j \in [n]_0, \tag{3.28}$$

where $\mathbb{1}$ is the indicator function. The transform matrix needs to be calculated only once for each n, which can be done offline. The matrix-vector multiplication requires $\frac{(n+1)(n+2)}{2}$ multiplications.

Similarly, for the FL method, by reorganizing (3.20), we can express $\mathbf{c} \triangleq (c_l)_{l=0}^n$ as a linear transformation of $\hat{\mathbf{m}} \triangleq (m_l)_{l=1}^{n+1}$, i.e.,

$$\mathbf{c} = \hat{\mathbf{A}}(\mathbf{1} - \hat{\mathbf{m}}), \tag{3.29}$$

where $\mathbf{c}$ and $\hat{\mathbf{m}}$ are understood as column vectors, $\mathbf{1}$ is the 1-vector of size $n+1$, and the transform matrix $\hat{\mathbf{A}} \in \mathbb{Z}^{(n+1)\times(n+1)}$ is given by

$$\hat{A}_{kl} \triangleq \frac{(-1)^{(k-l)}(2k+1)}{l+1}\sum_{j=0}^{l}\binom{k}{j}\binom{k}{k-j}\binom{k-j}{k-l}\mathbb{1}(l \leq k), \quad k, l \in [n]_0. \tag{3.30}$$

$\hat{\mathbf{A}}$ is a lower triangular matrix. As for the BM method, the transform matrix needs to be calculated only once for the desired level of accuracy, and the above matrix-vector multiplication requires $\frac{(n+1)(n+2)}{2}$ multiplications. Further details about the accuracy requirements are provided in Section 4.6.

4. Generation of moment sequences and performance evaluation

To compare the performance of different HMTs, moment sequences and corresponding functions serving as the ground truth are needed. Generating representative moment sequences is non-trivial since the probability that a sequence uniformly distributed on $[0, 1]^n$ is a moment sequence is about 2^{-n^2} [1, Eq. 1.2]. In this section, we provide three different ways to solve it: The first method generates uniformly distributed integer moment sequences, while the other two generate moments from either analytic expressions of moments or analytic expressions of cdfs. In these two cases, the parameters are randomized. We use the exact cdf as the reference function if it is given by analytic expressions. Otherwise, we use the GP approximations as the reference functions.

4.1. Moment sequences and their rates of decay

As discussed in Section 3.1, the CM method can fully recover discrete distributions with a finite number of jumps, while other methods, such as the ME method, cannot. This indicates that different HMTs may have different performance for different types of distributions, classified by the decay order of the sequence. We first provide some examples of moment sequences from common distributions.

Example 4.1 (Beta distributions)**.** For $X \sim \text{Beta}(\alpha, \beta)$,

$$m_n = \prod_{r=0}^{n-1}\frac{\alpha + r}{\alpha + \beta + r} = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)}. \tag{4.1}$$

By Stirling's formula, as $n \to \infty$,

$$m_n \sim e^{\beta} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \sqrt{\frac{\alpha+n-1}{\alpha+\beta+n-1}} \frac{(\alpha+n-1)^{\alpha+n-1}}{(\alpha+\beta+n-1)^{\alpha+\beta+n-1}}, \tag{4.2}$$

$$\sim e^{\beta} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{(\alpha+n-1)^{\alpha+n-1}}{(\alpha+\beta+n-1)^{\alpha+\beta+n-1}}, \tag{4.3}$$

$$\sim n^{-\beta}, \quad \text{as } \left(1+\frac{\beta}{x}\right)^{x} \to e^{\beta},\ x \to \infty. \tag{4.4}$$

Thus, $m_n = \Theta(n^{-\beta})$.

Example 4.2. Let the density function be $f(x) = 2x/s^2 \mathbf{1}_{[0,s]}(x)$ with $0 < s < 1$. Then $m_n = 2s^n/(n+2) = O(\exp(-\log(1/s)n)) = \Omega(\exp(-s_1 n))$ for all $s_1 > \log(1/s)$.

Given the examples shown above and that the performance of different HMTs is likely to be dependent on the rate of decay (or type of decay) of the moment sequences, we would like to develop a method to categorize different moment sequences based on their tail behavior. To this end, we introduce six types of moment sequences based on their rates of decay.

Definition 4.3 (Rate of decay)**.** 1) Sub-power-law decay: In this case, for all $s > 0$, $m_n = \Omega(n^{-s})$.
2) Power-law decay: In this case, there exists $s > 0$ such that $m_n = \Theta(n^{-s})$.
3) Soft power-law decay: In this case, $m_n \neq \Theta(n^{-s})$, $\forall s \geq 0$ but there exist $s_1 > s_2 \geq 0$ such that $m_n = \Omega(n^{-s_1})$ and $m_n = O(n^{-s_2})$.
4) Intermediate decay: In this case, for all $s_1 \geq 0$, $m_n = O(n^{-s_1})$ and for all $s_2 > 0$, $m_n = \Omega(\exp(-s_2 n))$.
5) Soft exponential decay: In this case, $m_n \neq \Theta(\exp(-sn))$, $\forall s > 0$ but there exists $s_1 > s_2 > 0$, $m_n = \Omega(\exp(-s_1 n))$ and $m_n = O(\exp(-s_2 n))$.
6) Exponential decay: In this case, there exists $s > 0$ such that $m_n = \Theta(\exp(-sn))$.

4.2. Uniformly random (integer) moment sequences

4.2.1. Generation

Let Λ denote the set of probability measures on $[0, 1]$. For $\lambda \in \Lambda$ we denote by

$$m_n(\lambda) \triangleq \int_0^1 x^n \, d\lambda, \quad n \in \mathbb{N}_0, \tag{4.5}$$

its n-th moment. Trivially, $m_0(\lambda) = 1, \forall \lambda \in \Lambda$. The set of all moment sequences is defined as

$$\mathcal{M} \triangleq \{(m_1(\lambda), m_2(\lambda), \ldots), \lambda \in \Lambda\}. \tag{4.6}$$

The set of the projection onto the first n coordinates of all elements in $\mathcal{M}$ is denoted by $\mathcal{M}_n$.

Now we recall how to establish a one-to-one mapping from the moment space int $\mathcal{M}$ onto the set $(0, 1)^{\mathbb{N}}$ [1], where int $\mathcal{M}$ denotes the interior of the set $\mathcal{M}$. For a given sequence $(m_k)_{k\in[n-1]} \in \mathcal{M}_{n-1}$, the achievable lower and upper bounds for m_n are defined as

$$m_n^- \triangleq \min\{m_n(\lambda), \lambda \in \Lambda \text{ with } m_k(\lambda) = m_k \text{ for } k \in [n-1]\}, \tag{4.7}$$

$$m_n^+ \triangleq \max\{m_n(\lambda), \lambda \in \Lambda \text{ with } m_k(\lambda) = m_k \text{ for } k \in [n-1]\}. \tag{4.8}$$

Note that $m_1^- = 0$ and $m_1^+ = 1$. m_n^- and m_n^+, $n \geq 2$, can be obtained by taking the infimum and supremum of the range obtained from solving the two linear inequalities $0 \leq \underline{H}_n$ and $0 \leq \overline{H}_n$.

We now consider the case where $(m_k)_{k\in[n-1]} \in \operatorname{int} \mathcal{M}_{n-1}$, i.e., $m_n^+ > m_n^-$. The canonical moments are defined as

$$p_n \triangleq \frac{m_n - m_n^-}{m_n^+ - m_n^-}. \tag{4.9}$$

$(m_k)_{k\in[n]} \in \operatorname{int} \mathcal{M}_n$ if and only if $p_n \in (0,1)$ [1, 2]. In this way, p_n represents the relative position of m_n among all the possible n-th moments given the sequence $(m_k)_{k\in[n-1]}$. The following lemma shows the relationship between the distribution on the moment space $\mathcal{M}_n$ and the distribution of p_k, $k \in [n]$.

Lemma 4.4. [1, Thm. 1.3] For a uniformly distributed random vector $(m_k)_{k\in[n]}$ on the moment space $\mathcal{M}_n$, the corresponding canonical moments p_k, $k \in [n]$, are independent and beta distributed as

$$p_k \sim \text{Beta}(n-k+1, n-k+1), \quad k \in [n].$$

Since $\mathcal{M}_n$ is a closed convex set, its boundary has Lebesgue measure zero. Therefore, although (4.9) only defines the mapping on $\operatorname{int} \mathcal{M}_n$, the p_k are still well defined.

Lemma 4.4 provides a method to generate uniformly random moment sequences. The details are summarized in the following algorithm.

Algorithm 1 Algorithm for random moment sequences by canonical moments

Input: The number of moments N
Output: $(m_n)_{n\in[N]}$
1: Set $m_1^- = 0$, $m_1^+ = 1$ and $k = 1$.
2: **for** $k \le N$ **do**
3: $p_k \sim \text{Beta}(n-k+1, n-k+1)$, $m_k = p_k(m_k^+ - m_k^-) + m_k^-$ and $k = k+1$.
4: Calculate m_k^- and m_k^+ based on $0 \le \underline{H}_k$ and $0 \le \overline{H}_k$.
5: **end for**

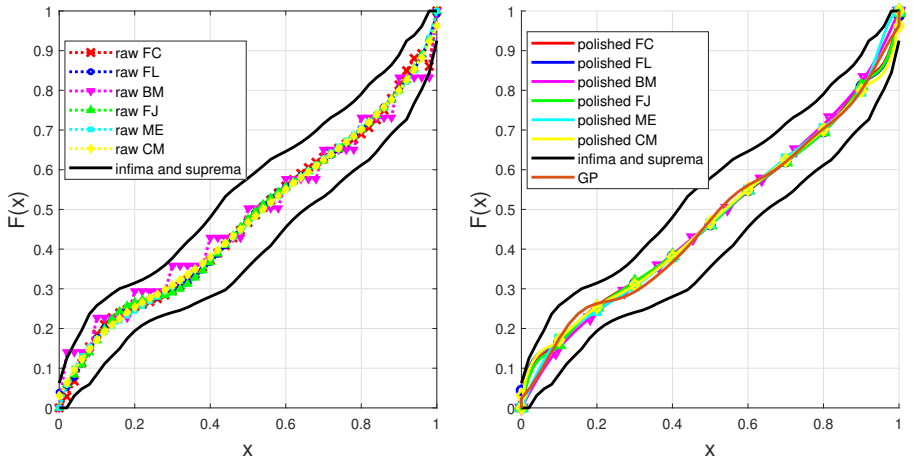
4.2.2. Usage

Figure 11a shows an example of the reconstructed cdfs (only raw versions) of each method where moments are generated by Algorithm 1 and 10 moments are used. But since a reference function serving as the ground truth is missing, we cannot compare the performance of each method. It is natural to look for a way to reconstruct a solution to the HMT via (infinitely many) integer moments. However, no known reconstruction methods exists for integer moments. To address this problem, we take a detour to approximate the characteristic function first, and then apply the Gil-Pelaez theorem to reconstruct the cdf [3]. Let $X \in [0,1]$ denote the random variable and $(m_n)_{n=0}^{\infty}$ denote its moments. The characteristic function of X is

$$\phi_X(s) = \mathbb{E}[\exp(jsX)], \quad s \in \mathbb{R}^+. \tag{4.10}$$

Written in terms of its Taylor expansion,

$$\phi_X(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(jsX)^k}{k!}\right]. \tag{4.11}$$



(a) Raw curves. (b) Polished curves.

Figure 11. *An example of the reconstructed cdf of each method. The* 10 *moments are randomly generated by the canonical moments and the moments are* 0.5256, 0.3893, 0.3188, 0.2747, 0.2442, 0.2215, 0.2038, 0.1895, 0.1775 *and* 0.1674. *The polished curves are obtained after the tweaking mapping in Definition 2.2 as well as the monotone cubic interpolation for* $\hat{F}_{\mathrm{BM},10}|_{\mathcal{U}_{11}}$, $\hat{F}_{\mathrm{ME},10}|_{\mathcal{U}_{10}}$, $\hat{F}_{\mathrm{FJ},10}|_{\mathcal{U}_{10}}$, $\hat{F}_{\mathrm{FL},9}|_{\mathcal{U}_{10}}$, $\hat{F}_{\mathrm{FC},9}|_{\mathcal{U}_{10}}$, *and* $\hat{F}_{\mathrm{CM},10}|_{\mathcal{U}_{10}}$..

By the dominated convergence theorem, we can change the order of expectation and summation, i.e., for any $s \in [0, \infty)$

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(jsX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{j^k s^k m_k}{k!}. \tag{4.12}$$

The imaginary moments of X, m_{js}, are the characteristic function of $\log(X)$. When applied to $X \sim \exp(1)$, the expansion in (4.12) yields the gamma function, which is consistent with the analytic continuation property of the moments.

The truncated sum

$$\hat{\phi}_X(s) = \sum_{k=0}^{N} \frac{j^k s^k m_k}{k!} \tag{4.13}$$

approximates the characteristic function $\phi_X(s)$, and, in turn, the integral

$$\hat{F}_X(x) \triangleq \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\Im(e^{-jsx} \hat{\phi}_X(s))}{s} ds, \quad x \in [0, 1], \tag{4.14}$$

is an approximation of the exact cdf

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\Im(e^{-jsx} \phi_X(s))}{s} ds, \quad x \in [0, 1]. \tag{4.15}$$

Therefore, by evaluating (approximating) $\hat{F}_X$, we obtain an approximation of the exact cdf F_X.

Figure 11b shows the reconstructions by the above method and all the HMTs. We observe that all the HMTs approximate well. However, even though the series converges for $s \in [0, \infty)$, the evaluation of each element with alternating signs could cause serious numerical problems, as s increases. For each s, we need at least $N = \lceil se \rceil$ to approximate $\phi_X(s)$ so that the absolute error is within $\sqrt{2/(\pi N)}$.

Meanwhile, an accurate evaluation of $\hat{F}_X$ usually requires the calculation of the integrand up to $s = 1000$. This requires the calculation of $N = 2719$ moments, which is time-consuming and not practical for the performance evaluation purposes in Section 4.5.

4.3. Generating moments by analytic expressions

As we need to generate moment sequences of variable lengths, it is natural to directly start with the infinite sequences $(m_n)_{n=0}^{\infty}$ and truncate them to the desired length.

We first recall the definition of completely monotonic functions.

Definition 4.5 (Completely monotonic function [30])**.** A function f is said to be c.m. on an interval I if f is continuous on I and has derivatives of all orders on int I (the interior of I) and for all $n \in \mathbb{N}$,

$$(-1)^n f^{(n)}(x) \geq 0, \quad x \in \text{int}\, I. \tag{4.16}$$

The sequence $(f(k))_{k=0}^{\infty}$ is c.m. if f is c.m. on $[0, \infty)$ [30]. Thus, we can use c.m. functions to generate moment sequences.

Based on the six types of moment sequences discussed in Section 4.1, we define six classes of c.m. functions that decay at different rates.

Definition 4.6 (Six classes of c.m. functions)**.** Based on the definition of rate of decay in Definition 4.3, we define the following c.m. functions.

- Sub-power-law decay $\Omega(n^{-s})$: $F_1(n; a, s) \triangleq a^s(\log(n+1)+a)^{-s}$.
- Power-law decay $\Theta(n^{-s})$: $F_2(n; a, s) \triangleq a^s(n+a)^{-s}$.
- Soft power-law decay: $F_3(n; a, b, s) \triangleq a^s b(n+a)^{-s}(\log(n+1)+b)^{-1}$.
- Intermediate decay: $F_4(n; a, s) \triangleq \exp(s\sqrt{a} - s\sqrt{n+a})$.
- Soft exponential decay: $F_5(n; a, s) \triangleq \exp(-sn)a(n+a)^{-1}$.
- Exponential decay $\Theta(\exp(-sn))$: $F_6(n; s) \triangleq \exp(-sn)$.

$s \in \mathbb{R}^+$ is referred to as the rate, and $a \in \mathbb{R}^+$ and $b \in \mathbb{R}^+$ (if present) are referred to as the other parameters.

For moment sequences $(m_{i,n})_{n=0}^{\infty}$, $i = 1, ..., N$, the convex combination $m_n = \sum_{i=1}^{N} c_i m_{i,n}$, where $\sum_{i=1}^{N} c_i = 1$, $c_i > 0$, remains c.m. Hence, $(m_n)_{n=0}^{\infty}$ is also a moment sequence.

Now, we introduce our algorithm to generate $(m_n)_{n=0}^{k}$ with a desired rate of decay s and flexibility in the other parameters, i.e., a and/or b. Since the c.m. functions are analytic, the c.m. function of integer moments can be easily extended to complex moments by analytic continuation. To obtain complex moments m_z, we just need to replace n in steps 2 and 3 of Algorithm 2 with $z \in \mathbb{C}$.

Algorithm 2 Algorithm for moment sequences by c.m. functions

Input: A desired rate of decay (including the type of decay and the rate s), the number of mixed decays N, the other rates $s_i, 2 \leq i \leq N$, the other parameters a_i, b_i, $1 \leq i \leq N$, the weights c_i, $1 \leq i \leq N$, and the number of moments k.

Output: $(m_n)_{n=0}^{k}$

1: Based on the type of decay, set F_0 to be one of $F_1, F_2, F_3, F_4, F_5, F_6$ in Definition 4.6. Suppose F_0 has three parameters a_i, b_i, s. If it is not the case, we will ignore a_i and/or b_i in the following steps.
2: Set $m_{i,n} = F_0(n; a_i, b_i, s_i)$, $i \in [N], n \in [k]_0$.
3: Set $m_n = \sum_{i=1}^{N} c_i m_{i,n}, n \in [k]_0$.

To deterministically generate moments by Algorithm 2, we need to fix $s > 0, N \in \mathbb{N}, s_i > s, 2 \leq i \leq N$ for the exponential decay and $s_i = s$ for the others, $a_i, b_i, c_i > 0$ and $\sum_{i=1}^{N} c_i = 1$.

To randomly generate moments by Algorithm 2, we first choose $s \sim \text{Uniform}(0, 10)$ except for the exponential and soft exponential cases, where $s \sim -\log(\text{Uniform}(0, 1))$. We set $s_i = s$ except that $s_1 = s, s_i \sim -\log(\text{Uniform}(0, \exp(-s))), 2 \le i \le N$ for the exponential decay. Next, we randomly generate N from the set $\{1, 2, ..., 10\}$. Then we choose $a_i, b_i \sim \text{Uniform}(0, 10)$, $1 \le i \le N$. Finally we choose $c_i \sim \text{Uniform}(0, 10)$ and normalize them so that the sum is 1. In this way, we have at least $2N$ degrees of freedom.

Similar to Section 4.2.1, the exact cdf is unknown. We use the GP approximations as a reference when making comparisons.

4.4. Generating moments by known distributions

Beta distributions are continuous probability distributions supported on $[0, 1]$. Here, we use the beta mixture model to randomly generate distributions. The pdf of N beta mixtures is

$$f_{\text{beta}}(x, \mathbf{a}, \mathbf{b}, \mathbf{c}) \triangleq \sum_{i=1}^{N} c_i x^{a_i - 1}(1 - x)^{b_i - 1}, \tag{4.17}$$

where $\mathbf{a} \triangleq (a_i)_{i=1}^N, \mathbf{b} \triangleq (b_i)_{i=1}^N, \mathbf{c} \triangleq (c_i)_{i=1}^N, \mathbf{a}, \mathbf{b}, \mathbf{c} > \mathbf{0}$ and $\sum_{i=1}^N c_i = 1$. The n-th moment is

$$m_n = \sum_{i=1}^{N} c_i \prod_{r=0}^{n-1} \frac{a_i + r}{a_i + b_i + r}, \tag{4.18}$$

and the cdf is

$$F_{\text{beta}}(x, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \sum_{i=1}^{N} c_i \int_0^x t^{a_i - 1}(1 - t)^{b_i - 1}\, dt. \tag{4.19}$$

By randomly choosing $\mathbf{a}, \mathbf{b}, \mathbf{c}$, we are able to randomly generate the distribution and thus randomly generate the imaginary and integer moment sequences by definition. We first randomly generate N from the set $\{1, 2, ..., 10\}$.[6] Then we choose $a_i, b_i \sim \text{Uniform}(0, 10)$, $1 \le i \le N$. Finally we choose $c_i \sim \text{Uniform}(0, 10)$ and normalize them so that the sum is 1.

4.5. Performance evaluation

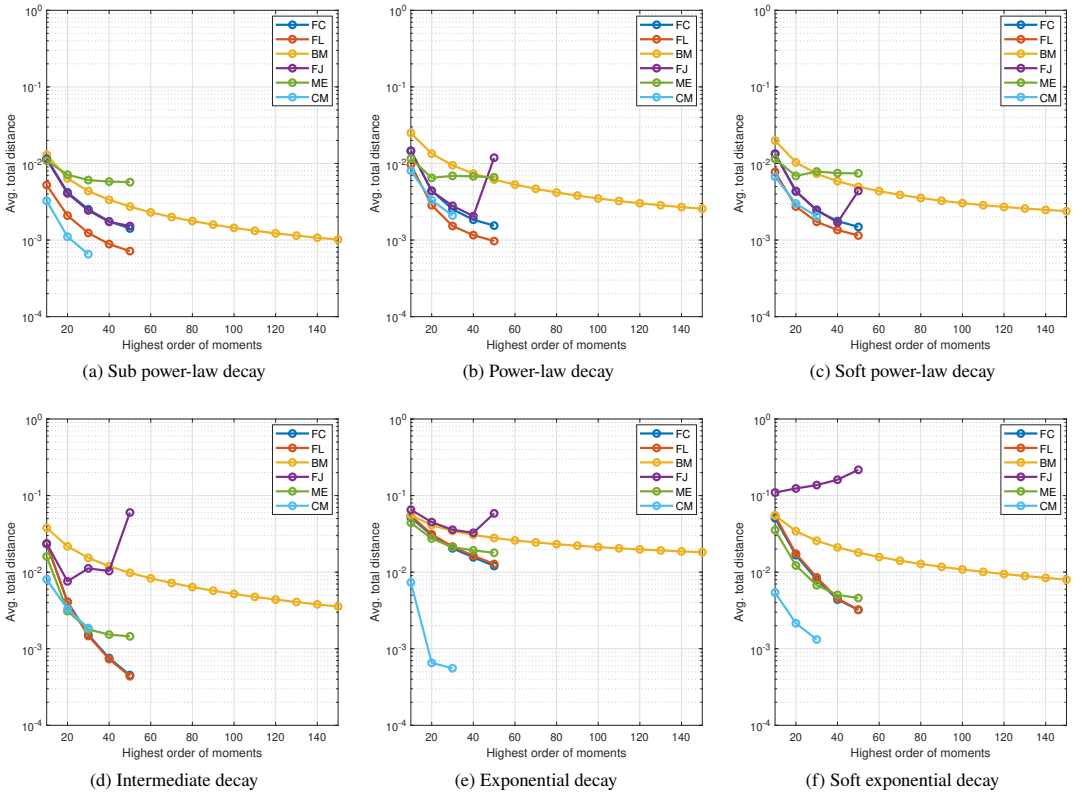
Using the moments randomly generated in Section 4.3, we compare the performance of the different Hausdorff moment transforms. Since the actual cdfs are not available in this case, we use the cdfs obtained by the GP method as the reference.[7]

For all methods except the GP, BM, FL, and FC methods, we consider $F_{\text{method,n}}|_{\mathcal{U}_n}$. For the GP method, we consider $F_{\text{GP}}|_{\mathcal{U}_{150}}$, for the BM method, we consider $F_{\text{BM,n}}|_{\mathcal{U}_{n+1}}$, and for the FL and FC methods, we consider $F_{\text{method,n}-1}|_{\mathcal{U}_n}$. We apply the tweaking mapping in Definition 2.2 followed by monotone cubic interpolation.

In Figure 12, we plot the average of the total distance versus the number of moments (m_0 is not counted) for different methods and different types of decays. For practical purposes regarding accuracy requirements in moments and basic operations and the difficulty in finding roots of polynomials, the number of moments of the FJ, FL, FC, and ME methods is limited to 50, the number of moments of the

[6] In order to increase the likelihood of generating moment sequences that can span the entire space, we randomly generate N, instead of a fixed value. Specifically, for a small number of realizations, such as 100, the span behavior is observed to vary with different values of N. Thus, randomizing N serves to increase the likelihood of achieving the desired outcome.

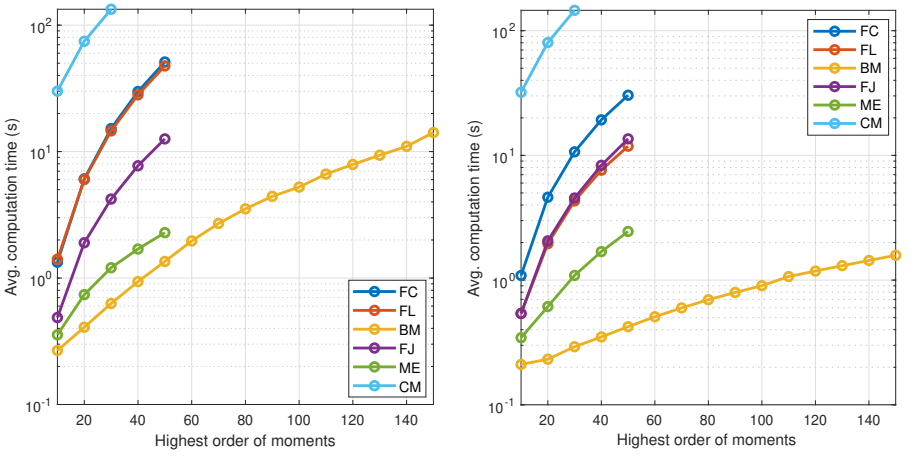
[7] As the GP method utilizes imaginary moments, it naturally processes the complete information to recover the exact distribution. However, the HMTs that produce solutions to the THMP, such as the ME method, may not be close to the exact one as they only utilize a finite number of integer moments. The gap does not necessarily indicate that an HMT performs poorly—the method may simply assume a different continuation of the moments sequence.



(a) Sub power-law decay (b) Power-law decay (c) Soft power-law decay

(d) Intermediate decay (e) Exponential decay (f) Soft exponential decay

Figure 12. *The average of the total distance versus the number of moments for different methods and different types of decays. Averaging is performed over* 100 *randomly generated moment sequences. .*



(a) Original implementation. (b) Linear transform.

Figure 13. *The average computation time versus the number of moments. Averaging is performed over* 100 *randomly generated moment sequences of power-law decay..*

CM method is limited to 30, and the number of moments of the BM method is limited to 150. We observe that the average of the total distance of all the methods except the FJ and the ME methods decreases as the number of moments increases. In most cases, the average of the total distance of the FJ method does not decrease since it is not guaranteed to converge. As for the ME method, although it converges in entropy, it suffers from the problem of finding the minimum in (3.4). Even if some of the curves in Figure 12 show a decreasing trend, we cannot rely on this method as long as we cannot make sure the results we get from solving (3.4) are the minima, i.e., the size of the gradient is less than the value of the optimality tolerance. In terms of their best possible performance, we also observe that the CM method is great for all types of decays, especially the exponential and soft exponential decays. For the exponential decay, the moments are from discrete distributions that have no more than 10 jumps. Thus, the raw CM method of order 20 is able to fully characterize the distributions by identifying jump locations and jump heights. Our comparisons are made based on the decay type. However, in practice, it is impossible to determine the decay type given the values of the moments. Thus, for practical purposes, it is preferable to utilize methods that work well overall. Among all the methods, the FL and FC methods have overall good performance and the FL method even outperforms the CM method in the power-law decay.

Figure 13 shows computation time versus the number of moments for the power-law decay and it is similar for the other types. In Figure 13a, the computation time of the FL, FC, and BM method is based on the original implementation, whereas in Figure 13b, it is based on the linear transform with the pre-calculated transform matrices. In Figure 13b, for all methods except the CM method, the computation time increases quadratically with the number of moments. This is consistent with the order of matrix-vector multiplication for the FL, FC and BM method. Furthermore, with the linear transform, the computation time for the FL, FC and BM method in Figure 13a is significantly reduced compared to that in Figure 13a. Even for small values of n, such as 10, 20, and 30, the CM method's computation time is at least ten times longer than that of the other methods for the same number of moments. As n increases, the computation time of the CM method is expected to grow even more, making it unsuitable for time-sensitive applications.

As shown in Example 4.1, the moment sequences generated by Section 4.4 exhibit power-law decay. Its results are similar to those in Figure 12b.

4.6. Accuracy requirements for HMTs

The accuracy requirements apply to both of the coefficients of the moments and the HMTs, such as the transform matrices for the BM and FL method. Here we just mention some typical requirements.

Accuracy requirements for transform matrices: For the BM method of order n, as $\max \|\mathbf{A}\| \sim \frac{\sqrt{27}}{2\pi}\frac{3^n}{n}$, $n \to \infty$, about $n/2 - \log_{10} n$ decimal digits are needed [6]. For the FL method of order n, as $\max \|\hat{\mathbf{A}}\| \leq \binom{2n}{n}\binom{n}{n/3} \sim \frac{3^{n+1}2^{4n/3}}{4\pi n}$, $n \to \infty$, about $0.9n$ decimal digits are sufficient.

Accuracy requirements for moments in Definition 4.6: For the power-law decay, about $s \log_{10} \frac{n+a}{a}$ decimal digits are needed. For exponential decay, about $sn/2$ decimal digits are needed.

The calculation can be carried out in the logarithm domain which may reduce the accuracy requirement. Moreover, the matrix multiplication of the FL method is pre-calculated, i.e., $\hat{\mathbf{A}}\mathbf{1}$ and $\hat{\mathbf{A}}\hat{\mathbf{m}}$ are done separately.

5. Conclusion

The truncated Hausdorff moment problem is both fundamental and practically relevant. While several theoretical methods for its solutions have been proposed, a systematic study and comparison of their performance and details in their numerical implementation has been missing. This work provides these important missing pieces by discussing advantages and drawbacks of different methods, comparing convergence properties, and proposing a tweaking method to ensure the resulting functions satisfy the properties of probability distributions. In addition, two new methods are introduced that compare

favorably with known ones. For a fair comparison, three methods for the random generation of moment sequences are discussed or proposed, so that the distances between approximated and actual distributions can be averaged over many realizations of moment sequences. It turns out that the decrease of the distances as a function of the number of moments depends strongly on the type of decay of the moment sequences. For this reason, we introduced a classification of the decay order in six classes. Our work also reveals the trade-off between performance and computational cost. If high accuracy is desired (at higher computational effort), the FC and FL methods are good choices, with the FC method having the advantage of uniform convergence; if computational resources are limited, the BM method is preferred.

Funding Statement. This work was supported in part by the US National Science Foundation through Grant 2007498.

Competing Interests. None

References

[1] Chang, F.-C., Kemperman, J., & Studden, W. (1993). A Normal Limit Theorem for Moment Sequences. *The Annals of Probability*, 21(3), 1295–1309.
[2] Dette, H. & Studden, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*. Wiley, New York.
[3] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38(3-4), 481–482.
[4] Guruacharya, S. & Hossain, E. (2018). Approximation of Meta Distribution and Its Moments for Poisson Cellular Networks. *IEEE Wireless Communications Letters*, 7(6), 1074–1077.
[5] Haenggi, M. (2016). The Meta Distribution of the SIR in Poisson Bipolar and Cellular Networks. *IEEE Transactions on Wireless Communications*, 15(4), 2577–2589.
[6] Haenggi, M. (2018). Efficient Calculation of Meta Distributions and the Performance of User Percentiles. *IEEE Wireless Communications Letters*, 7(6), 982–985.
[7] Haenggi, M. (2021a). Meta distributions—part 1: Definition and examples. *IEEE Communications Letters*, 25(7), 2089–2093.
[8] Haenggi, M. (2021b). Meta distributions—part 2: Properties and interpretations. *IEEE Communications Letters*, 25(7), 2094–2098.
[9] Hausdorff, F. (1921). Summationsmethoden und Momentfolgen. I. *Mathematische Zeitschrift*, 9(1-2), 74–109.
[10] Inverardi, P. N., Pontuale, G., Petri, A., & Tagliani, A. (2003). Hausdorff moment problem via fractional moments. *Applied Mathematics and Computation*, 144(1), 61–74.
[11] John, V., Angelov, I., Öncül, A., & Thévenin, D. (2007). Techniques for the reconstruction of a distribution from a finite number of its moments. *Chemical Engineering Science*, 62(11), 2890–2904.
[12] Kapur, J. N. (1989). *Maximum-entropy models in science and engineering*. John Wiley & Sons, Hoboken, NJ.
[13] Li, Z. (1995). Pointwise convergence of Fourier-Jacobi series. *Approximation Theory and its Applications*, 11(4), 58–77.
[14] Markov, A. (1884). Démonstration de certaines inégalités de M. Tchébychef. *Mathematische Annalen*, 24(2), 172–180.
[15] Mason, J. C. & Handscomb, D. C. (2002). *Chebyshev polynomials*. Chapman and Hall/CRC, Boca Raton, FL.
[16] Mead, L. R. & Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8), 2404–2417.
[17] Mnatsakanov, R. M. (2008). Hausdorff moment problem: reconstruction of distributions. *Statistics & Probability Letters*, 78(12), 1612–1618.
[18] Mnatsakanov, R. M. & Ruymgaart, F. (2003). Some properties of moment-empirical cdf's with application to some inverse estimation problems. *Mathematical Methods of Statistics*, 12(4), 478–495.
[19] Pollard, H. (1972). The convergence almost everywhere of Legendre series. *Proceedings of the American Mathematical Society*, 35(2), 442–444.
[20] Rácz, S., Tari, Á., & Telek, M. (2006). A moments based distribution bounding method. *Mathematical and Computer Modelling*, 43, 1367–1382.
[21] Schmüdgen, K. (2017). *The moment problem*. Springer, Berlin.
[22] Schoenberg, I. (1973). Remarks concerning a numerical inversion of the Laplace transform due to Bellman, Kalaba, and Lockett. *Journal of Mathematical Analysis and Applications*, 43(3), 823–828.
[23] Shohat, J. & Tamarkin, J. (1943). *The Problem of Moments*. American Mathematical Society, Providence, RI.
[24] Szegö, G. (1939). *Orthogonal polynomials*. American Mathematical Society, Providence, RI.
[25] Tagliani, A. (1999). Hausdorff moment problem and maximum entropy: a unified approach. *Applied Mathematics and Computation*, 105(2-3), 291–305.
[26] Tang, X., Xu, X., & Haenggi, M. (2020). Meta Distribution of the SIR in Moving Networks. *IEEE Transactions on Communications*, 68(6), 3614–3626.
[27] Wang, X. & Haenggi, M. (2022). The Chebyshev-Markov inequalities. Technical report, Department of Electrical Engineering, University of Notre Dame, Notre Dame.

[28] Wang, X. & Haenggi, M. (2023). Fast hausdorff moment transforms for meta distributions in wireless networks. *IEEE Transactions on Wireless Communications*, pages 1–1.

[29] Wang, Y., Cui, Q., Haenggi, M., & Tan, Z. (2018). On the SIR Meta Distribution for Poisson Networks With Interference Cancellation. *IEEE Wireless Communications Letters*, 7(1), 26–29.

[30] Widder, D. (1931). Necessary and sufficient conditions for the representation of a function as a Laplace integral. *Transactions of the American Mathematical Society*, 33(4), 851–892.

[31] Zelen, M. (1954). Bounds on a distribution function that are functions of moments to order four. *Journal of Research of the National Bureau of Standards*, 53(6), 377–81.