National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

11-19

The Multivariate Moments Problem: A Practical High Dimensional Density Estimation Method

Bradley Wakefield, Yan-Xia Lin, Rathin Sarathy and Krishnamurty Muralidhar

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522, Australia Phone +61 2 4221 5435, Fax +61 2 42214998.
Email: karink@uow.edu.au

The Multivariate Moments Problem:

A Practical High Dimensional Density Estimation Method

October 8, 2019

Bradley Wakefield

National Institute for Applied Statistics Research Australia

School of Mathematics and Applied Statistics

University of Wollongong, NSW 2522, Australia

Email: bnw722@uowmail.edu.au

Yan-Xia Lin

National Institute for Applied Statistics Research Australia

School of Mathematics and Applied Statistics

University of Wollongong, NSW 2522, Australia

Email: yanxia@uow.edu.au

Rathin Sarathy

Department of Management Science & Information Systems

Oklahoma State University, OK 74078, USA

Email: rathin.sarathy@okstate.edu

Krishnamurty Muralidhar

Department of Marketing & Supply Chain Management,

Research Director for the Center for Business of Healthcare

University of Oklahoma, OK 73019, USA

Email: krishm@ou.edu

Abstract

The multivariate moments problem and it's application to the estimation of joint density functions is often considered highly impracticable in modern day analysis. Although many results exist within the framework of this issue, it is often an undesirable method to be used in real world applications due to the computational complexity and intractability of higher order moments. This paper aims to take a new look at the application of a multivariate moment based method in estimating a joint density and demonstrate that sample moment based approximations can offer a legitimate rival to current density approximation methods.

Within this paper we use tools previously derived, such as the properties of hyper-geometric type eigenfunction polynomials, properties in the approximation of density functions and the completeness of orthogonal multivariate polynomials to obtain a sample based multivariate density approximation and associated properties of determinacy and convergence. Furthermore we attempt to overcome one of the most considerable issues in applying a moment based density approximation, the determination of the required parameters, by exploiting the analytical properties of orthogonal polynomials and applying various statistical estimation techniques.

It then follows that by deriving these various analytic results, this paper obtains the specific statistical tools required to take the multivariate moments problem from a highly specified mostly analytical idea to a practical high dimensional density estimation method.

Keywords: Density Estimation, Multivariate, Moments Problem, Orthogonal Polynomial, Sample Moments, Multivariate Density, Joint Density Estimation, Hyper-geometric Polynomial

1. INTRODUCTION

The construction of unique polynomial based series expansions for continuous functions is a common practice applied in a range of areas of mathematics, physics engineering and statistics (Munkhammar et al., 2017). As such, polynomial based expansions have also been applied in the study of probability density functions on both closed and infinite domains (Elderton and Johnson, 1969; Solomon and Stephens, 1978; Kendall et al., 1987; Provost, 2005) . Known more commonly as the moments problem, this area of study has been well studied and developed from as far back as the 19th century. However, practical use of such expansions are only more recently been applied due to the computational complexity required in the estimation becoming less significant with the advent of high processing computational technology. Moreover, the study of such an approach to estimating multivariate distributions is much more sporadic. As noted by Kleiber and Stoyanov (Kleiber and Stoyanov, 2013), the results for the multivariate case tend to be scattered throughout specialised journals and tend to be focused on a more analytical perspective.

Throughout the literature, researchers have tended to focus on the usefulness and effectiveness of a particular orthogonal polynomial basis in regards to estimating certain distributions, usually selecting classical polynomial families as the basis for the expansion (Gram, 1883; Edgeworth, 1905; Charlier, 1914; Szegő, 1939; Cramer, 1946; Alexits, 1961; Kendall et al., 1987; Provost, 2005). These classical polynomials are usually chosen based on the resemblance of their weight functions to the canonical distributions. Orthogonal polynomials with which a density approximation has been derived include Hermite (Edgeworth and Gram-Charlier Expansions), Legendre, Jacobi and Laguerre Polynomials (Gram, 1883; Edgeworth, 1905; Charlier, 1914; Szegő, 1939; Cramer, 1946; Alexits, 1961; Kendall et al., 1987; Provost, 2005) which have weight functions resembling that of the normal, uniform, beta and gamma distributions respectively. These polynomials can all be used to provide moment based expressions of certain density functions and are all eigenfunctions of a hypergeometric type second order differential equation (Sánchez-Ruiz and Dehesa, 1998).

Multivariate orthogonal polynomial expansions particularly with reference to Hermite, Jacobi and Laguerre polynomials have also been investigated (Xu, 1997; Griffiths and Spano, 2011). Joint density expressions for quasi-multivariate normal distributions have been developed by using a Hermite polynomial based expansion (Mehler, 1866; Withers and Nadarajah, 2010) and also Laguerre polynomial based expansions have been shown to be appropriate for a marginally gamma multivariate distribution (Kibble, 1940; Krishnamoorthy and Parthasarathy, 1951; Mustapha and Dimitrakopoulos, 2010). However, when considering Gaussian expansions using Hermite polynomials, such as the Edgeworth expansions, considerable issues regarding the behavior of the tail of the distribution are not well-defined (Kendall et al., 1987; Welling, 1999) . Although some Saddlepoint approximation methods have been proposed to overcome these issues (Daniels, 1983), these approximations tend to be restrictive and prone to failure (Provost, 2005). This is also of true of Laguerre polynomial based expansions especially when attempting to use sample moments in the approximation where the parameters can suffer from considerable sampling variance problems.

One of the key differences between the moment-based density expressions relates to the support of the varying orthogonal bases used in the approximation. It is imperative that the support of these bases are consistent with that of the sample space of the density function we are expressing in the expansion (Kleiber and Stoyanov, 2013). This inherently leads us to three distinct cases of problems: Hausdorff (density functions defined on a compact space), Stieltjes (density functions defined on a half bounded (lower or upper) space) (Stieltjes, 1894; Stoyanov and Tolmatz, 2005) and Hamburger (density functions defined on a completely unbounded space) (Shohat and Tamarkin, 1943; Akhiezer and Kemmer, 1965) and therefore there are varying cases we must consider in order to ensure the determinacy of the expansion. Determinacy refers to the existence of an unique moment based expansion of the density function and in terms of multivariate moments problem, this notion of determinacy is extensively reviewed in Kleiber and Stoyanov (Kleiber and Stoyanov, 2013). Interestingly, the paper investigates the relationship between the marginal determinacy and the determinacy of the multivariate distribution as a whole. It raises the Peterson result (Petersen, 1982), that marginal determinacy implies multivariate determinacy and conjectures that for absolutely continuous function, multivariate determinacy implies marginal. This issue of determinacy becomes fairly trivial for multivariate distribution functions defined on a compact spaces as the multivariate analogue of the Weierstrass Stone Approximation theorem (Szökefalvi-Nagy, 1965) confirms the uniqueness of the polynomial expansion.

In many practical situations, no additional information about the joint density is available beyond that attached to a random sample of observations from the distribution. It is therefore difficult to decided exactly which type of expansion would be more representative of the original distribution especially when attempting to apply only one sort of expansion to a multitude of varying distribution functions where domains may vary. When enforcing various assumptions about the underlying marginal distributions, we can also misrepresent the original statistical information contained in the random sample and hence fail to properly approximate the true joint density.

It is for this reason we may prefer a more uninformative approach by using a constant reference density on a compact space. This is the approach undertaken by a Legendre expansion (Provost, 2005) and allows for a more robust approach to density estimation by imputing less prior information into the construction of the density function, especially when trying to protect the original statistical information of the sample. Regardless, by taking a broader perspective towards the basis used, we can obtain results for all of these approaches and allow for easier comparison to select the more appropriate basis.

In this paper we use tools previously derived, such as the properties of hyper-geometric type eigenfunction polynomials (Yanez et al., 1994; Sánchez-Ruiz and Dehesa, 1998), properties in the approximation of density functions (Szökefalvi-Nagy, 1965; Provost, 2005) and the completeness of orthogonal multivariate polynomials (Xu, 2014) to obtain a sample based multivariate density approximation and associated properties of determinacy (Putinar and Schmüdgen, 2008; Kleiber and Stoyanov, 2013) and convergence. We focus on distributions with bounded marginal domains for each variable in order to ensure determinacy and makes no specific choice about the associated reference joint density (weight) used. However, this restriction to a bounded domain is certainly not necessary if moment determinacy can be assured and as such, all results presented in this paper can apply on unbounded domains in certain situations. Furthermore, as demonstrated in (Lin, 2014), which applies a univariate form of moment-based density estimation for the purposes of recovering information from perturbed data, provided the sample size is sufficiently large enough, assuming a compact domain can result in a good estimate for density functions which exist on an unbounded domain.

Part of the most considerable issues in applying a moment based density approximation is determining how to best select the required parameters in the estimation. This is an area unexplored in the previous literature and by deriving various analytic results, this paper derives the specific statistical tools required to take the multivariate moments problem from a highly specified mostly analytical idea to a practical high dimensional density estimation method comparable to that of other multivariate density approximation methods.

This paper is structured into seven main sections. The second section relates to the formula and results pertaining to using a generalised hyper-geometric type polynomial base for density estimation extensively studied in previous literature. The third section looks at applying these formula to obtain the multivariate density approximation with the fourth section then deriving properties and techniques for applying this approximation to a sample. The fifth section contains a simulation study using these methods to assess the ability of this methods to adequately approximate the joint density. The sixth section extends these results to the cumulative distribution function, the conditional cumulative distribution function and proposes a strategy for simulating observations from such an approximation. We will conclude this paper with remarks about the tools that were derived as well as a direction for further research.

2. BASIC FORMULA AND RESULTS

This paper investigates a method by which a continuous multivariate density function $f_{\mathbf{X}}$ where $X = (X_1, \ldots, X_N)$ is a vector of random variables, defined on a compact space $\mathbf{\Omega} = \prod_{k=1}^{N}[a_k, b_k]$ with defined finite moments, may be approximated by use of a reference distribution function $f_\nu = \prod_{k=1}^{N} f_{\nu_k}$ where f_{ν_k} is a probability density function defined for each $k = 1, \ldots, N$ on the closed interval $\Omega_k = [a_k, b_k]$ with corresponding orthonormal polynomials $\{P_{k,n_k}\}_{n_k \in \mathbb{N}_0}$, where n_k denotes the order of the polynomial P_{k,n_k} and N_0 refers to the set of all natural numbers including 0.

These polynomials are defined in a univariate sense for each kth reference distribution as particular polynomial solutions to the Sturm-Liouville differential equations of the form (Sánchez-Ruiz and Dehesa, 1998) (where λ_k is some constant and σ_k is a quadratic polynomial function):

$$\frac{d}{dx_k}\left[\sigma_k(x_k) f_{\nu_k}(x_k) P_{k,n_k}(x_k)\right] = \lambda_{n_k} f_{\nu_k}(x_k) P_{k,n_k}(x_k) \qquad \text{for} \;\; n_k = 0, 1, 2, ...$$

Which when complete, form an orthonormal basis which by the Weierstrass-Stone approximation Theorem (Szökefalvi-Nagy, 1965) and allows for the polynomial approximation of smooth functions.

It should be noted that the first two results of the following section follow from the work outlined in Sánchez-Ruiz and Dehesa (1998) but are given here for reading convenience. We first consider these polynomials P_{k,n_k} associated with one dimensional space and subsequently generalise to higher dimensions.

2.1 Univariate Orthonormal Polynomials

Let f_ν be the density function of the random variable ν with domain $\Omega = [a, b]$.

Definition 1. The Hyper-Geometric Type Differential Equation (Sánchez-Ruiz and Dehesa, 1998).

Given f_ν is such that there exists polynomial functions $\sigma(x)$ and $\tau(x)$ of degree no greater than

2 and 1 respectively such that:

$$\frac{d}{dx}\left[\sigma(x)f_\nu(x)\right] = \tau(x)f_\nu(x) \tag{1}$$

As well as having boundary conditions such that for $i = 0, 1, ...$

$$\lim_{x\to a} x^i\sigma(x)f_\nu(x) = \lim_{x\to b} x^i\sigma(x)f_\nu(x) = 0 \tag{2}$$

Then we define the following equation as the hyper-geometric-type differential equation:

$$\sigma(x)y''(x) + \tau(x)y'(x) - \lambda y(x) = 0$$

The Hyper-geometric type differential equation is a Sturm-Liouville problem which takes the self-adjoint form (Sánchez-Ruiz and Dehesa, 1998) :

$$\frac{d}{dx}\left[\sigma(x)y'(x)f_\nu(x)\right] = \lambda y(x)f_\nu(x) \tag{3}$$

This differential equation has infinitely many eigenfunction solutions with the eigenvalues $\lambda_n \in \mathbb{R}$ given by $\lambda_n = n\tau' + \frac{1}{2}n(n-1)\sigma''$ for $n = 0, 1, ...$ (noting that as $\sigma(x)$ and $\tau(x)$ are polynomials of degree no greater than 2 and 1 respectively then σ'' and τ' are constants) which have corresponding hypergeometric-type polynomial eigenfunctions $y_n(x)$ of degree n which are orthogonal on $[a, b]$ with respect to the weight function f_ν. We note that the solutions $y_n(x)$ are polynomial functions and hence will now be denoted as $P_n(x)$.

Furthermore we can obtain a Rodrigues' (type) Formula (Sánchez-Ruiz and Dehesa, 1998) for the nth eigenfunction of the Hyper-geometric type differential equation (Definition 1) which we will now denote as $P_n(x)$. These eigenfunctions are polynomials with Rodrigues' formula of the form (for the nth polynomial):

$$P_n(x) = \frac{B_n}{f_\nu(x)}\frac{d^n}{dx^n}\left[(\sigma(x))^n f_\nu(x)\right]$$

Where $B_n \in \mathbb{R}$ is the normalising constant in the Rodrigues' formula and can be chosen fairly arbitrarily in order to show consistency with classical orthogonal polynomial families. In this paper, we will choose B_n as a constant which ensures $P_n(x)$ are orthonormal.

We will now introduce some results related to the orthonormal polynomials $P_n(x)$. These properties will benefit the computation process proposed for estimating joint density functions.

Result 1: The expression of the normalising constant B_n.

The normalising constant from the Rodrigues' formula B_n which gives $P_n(x)$ as an orthonormal polynomial, such that $\int_a^b P_n(x)P_m(x)f_\nu(x)dx = \delta_{n,m}$ where $\delta_{n,m}$ denotes the Kronecker symbol, is given by:

$$B_n = \sqrt{\frac{(-1)^n \prod_{i=0}^{n-1}\left[\frac{1}{\tau' + \frac{1}{2}(n+i-1)\sigma''}\right]}{n! \int_a^b (\sigma(x))^n f_\nu(x)dx}} \qquad \text{for } n = 1, 2, \ldots \quad \text{and} \quad B_0 = 1$$

Proof. Sánchez-Ruiz and Dehesa (1998) gives the orthogonality constant $A_n = \mathbb{E}_\nu\left[P_n^2(\nu)\right]$ for a general B_n as

$$A_n = (-1)^n n! B_n^2 \prod_{i=0}^{n-1}\left[\tau' + \frac{1}{2}(n+i-1)\sigma''\right] \int_a^b (\sigma(x))^n f_\nu(x)dx$$

Hence as $\mathbb{E}_\nu\left[P_n^2(\nu)\right] > 0$ and by the definition of orthogonality $A_n \neq 0$ then it can be ascertained that the expression of B_n defined above ensures $A_n = 1$ for all $n = 0, 1, \ldots$. □

As the polynomials $P_n(x)$ are defined to be orthonormal in this paper we know B_n is given by the expression of B_n shown in Result 1.

Result 2: The completeness of the orthogonal base $\{P_n\}_{n\in\mathbb{N}_0}$

Consider the L_2 space $L_2(\Omega, \mu_\nu)$ where μ_ν is a measure in $(\Omega, \mathscr{F})$ given by $\mu_\nu(\omega) = \int_\omega f_\nu \, dx$ for $\omega \in \mathscr{F}$, then $P_n(x)$ are complete in $L_2(\Omega, \mu_\nu)$ space with $P_n \in L_2(\Omega, \mu_\nu)$ as $\int_\Omega P_n^2 d\mu_\nu = \int_\Omega P_n(x)^2 f_\nu(x)dx = 1 < \infty$ for all $n \in \mathbb{N}_0$. A complete proof can be found in (Szökefalvi-Nagy, 1965).

Result 3: The expression of the coefficients $a_{n,i}$ where $P_n(x) = \sum_{i=1}^n a_{n,i}x^i$.

$P_n(x)$ is a polynomial. Therefore $P_n(x)$ can be expressed as $P_n(x) = \sum_{i=1}^n a_{n,i}x^i$. We note that using the orthogonality property of the polynomials and integration by parts, Sánchez-Ruiz and Dehesa (1998) states that the leading term of the polynomial $P_n(x)$ (i.e. $a_{n,n}$ where $P_n(x) = \sum_{i=1}^n a_{n,i}x^i$) denoted κ_n for each $n = 1, 2, \ldots$ is given by:

$$\kappa_n = B_n \prod_{i=0}^{n-1}\left[\tau' + \frac{1}{2}(n+i-1)\sigma''\right]$$

We extend this result further to obtain a recurrence formula for the ith coefficient of the nth polynomial for $i < n$.

Lemma 2.1. *Consider the polynomial $P_n(x) = \sum_{i=0}^n a_{n,i}x^i$.*

The coefficient $a_{n,i}$ for fixed $n \in \mathbb{N}_0$ is given by:

$$a_{n,i} = \begin{cases} \frac{a_{n,i+1}(i+1)(i\sigma'(0)+\tau(0)) + a_{n,i+2}(i+2)(i+1)\sigma(0)}{\lambda_n - i(i-1)\frac{\sigma''}{2} - i\tau'} & i = 0, \ldots, n-2 \\ \kappa_n \frac{n(n-1)\sigma'(0) + n\tau(0)}{\tau' + (n-1)\sigma''} & i = n-1 \\ \kappa_n & i = n \end{cases} \tag{4}$$

Where $\kappa_n = B_n \prod_{i=0}^{n-1} \left[\tau' + \frac{1}{2}(n+i-1)\sigma''\right]$.

A proof of this result can be found in the supplementary materials under Appendix A.

Result 4: The expression of the hyper-geometric type polynomial integral.

We note that using properties of the second order differential equation we obtain the following lemma:

Lemma 2.2. *The integral of the hyper-geometric type polynomial* $P_n(x)$ *with respect to the measure function* $d\mu_\nu$ *is given by:*

$$\int P_n(x) d\mu_\nu = \int P_n(x) f_\nu(x) dx = \frac{\sigma(x)}{\lambda_n} P_n'(x) f_\nu(x) + c$$

Where $c \in \mathbb{R}$ *is some constant.*

Proof. From the self adjoint form (3) we note:

$$\lambda_n P_n(x) f_\nu(x) = \frac{d}{dx}\left[\sigma(x) P_n'(x) f_\nu(x)\right]$$

Hence it follows that

$$\int P_n(x) d\mu_\nu = \frac{1}{\lambda_n}\int \lambda_n P_n(x) f_\nu(x) dx = \frac{1}{\lambda_n}\int \frac{d}{dx}\left[\sigma(x) P_n'(x) f_\nu(x)\right] dx = \frac{\sigma(x)}{\lambda_n} P_n'(x) f_\nu(x) + c$$

Where $c \in \mathbb{R}$ is some constant. □

Remark From Lemma 2.2: $\int_a^t P_n(x) d\mu_\nu = \frac{\sigma(t)}{\lambda_n} P_n'(t) f_\nu(t)$ for $t \in [a, b]$ as by (2):

$$\lim_{t\to a} \sigma(t) P_n'(t) f_\nu(t) = \lim_{t\to a} \sigma(t) f_\nu(t) \sum_{i=0}^{n-1} (i+1) a_{n-1,i} t^i = \sum_{i=0}^{n-1} (i+1) a_{n-1,i} \left[\lim_{t\to a} t^i \sigma(t) f_\nu(t)\right] = 0$$

2.2 Multivariate Orthonormal Polynomials.

Consider now the N-dimensional probability space $(\boldsymbol{\Omega}, \mathscr{F}, P)$ defined on the compact space $\boldsymbol{\Omega} = \prod_{k=1}^N \Omega_k = \prod_{k=1}^N [a_k, b_k]$ and the random vector ν of N independent continuous random variables $\nu_k : \Omega_k \to \mathbb{R}$ with reference probability density functions $f_{\nu_k} : \Omega_k \to \mathbb{R}$ for $k = 1, \ldots, N$. We denote the joint probability density function of ν by $f_\nu : \boldsymbol{\Omega} \to \mathbb{R}$ and also consider the L_2 space $L_2(\boldsymbol{\Omega}, \mu_\nu)$ with measure function μ_ν defined by:

$$\mu_\nu(\omega) = \int_\omega f_\nu(\mathbf{x}) d\mathbf{x} = \int_\omega \left[\prod_{k=1}^N f_{\nu_k}(x_k)\right] d\mathbf{x} \qquad \omega \in \mathscr{F} \tag{5}$$

We now construct a set of multivariate polynomials by considering the formulation for some $n_1, \ldots, n_N \in \mathbb{N}_0$.

Definition 2. The Multivariate Hyper-Geometric Type Polynomial.

The multivariate hyper-geometric type polynomial is given as the product of N hyper-geometric type polynomials P_{k,n_k} of degree n_k for each $k = 1, \ldots, N$. We denote these polynomials as:

$$P_{n_1,\ldots,n_N}(\mathbf{x}) = \prod_{k=1}^{N} P_{k,n_k}(x_k) = \prod_{k=1}^{N} \frac{B_{k,n_k}}{f_{\nu_k}(x_k)} \frac{d^{n_k}}{dx^{n_k}} \left[(\sigma_k(x_k))^{n_k} f_{\nu_k}(x_k) \right] \tag{6}$$

Where σ_k and τ_k refer to the polynomials described in (1) of the kth hyper-geometric type polynomial and B_{k,n_k} refers to the kth normality constant described in Result 1.

The main purpose of constructing these multivariate polynomials is that they can then be used in the polynomial expansion of our multivariate density. By defining the polynomials in this way, it is possible to ensure that, regardless of other properties of the marginal distributions, the domain of each marginal distribution is consistent with whichever reference density (weight function) we decide to use. This inherently simplifies the problem to being a question of whether the our polynomials have the same domain as each marginal distribution. However it is also imperative to ensure that these polynomials are indeed orthonormal and complete.

Theorem 2.3. *The multivariate hyper-geometric type polynomials are orthonormal on the compact space $\mathbf{\Omega} = \prod_{k=1}^{N} \Omega_k$ and complete in the L_2 space $L_2(\mathbf{\Omega}, \mu_\nu)$.*

Proof. Firstly as the integrals on Ω_k are all finite, then the integral on the compact space $\mathbf{\Omega}$ can be expressed as:

$$\int_{\mathbf{\Omega}} P_{n_1,\ldots,n_N}(\mathbf{x}) P_{m_1,\ldots,m_N}(\mathbf{x}) d\mu_\nu(\mathbf{x}) = \prod_{k=1}^{N} \left[\int_{\Omega_k} P_{k,n_k}(x_k) P_{k,m_k}(x_k) f_{\nu_k} dx \right] = \prod_{k=1}^{N} \delta_{n_k,m_k}$$

Furthermore given that for each $k = 1, \ldots, N$ the set of orthonormal polynomials $\{P_{k,n_k}\}_{n_k \in \mathbb{N}_0}$ are complete in $L_2(\Omega_k, \mu_{\nu_k})$ then for any function $g(x_k)$ on Ω_k:

If $\int_{a_k}^{b_k} g(x_k) P_{k,n_k}(x_k) f_{\nu_k}(x_k) dx_k = 0$ for all $n_k \in \mathbb{N}_0$ then $g(x_k) = 0$ for all $x_k \in \Omega_k$ almost surely.

Therefore as $P_{n_1,\ldots,n_N}(\mathbf{x}) = \prod_{k=1}^{N} P_{k,n_k}(x_k)$ then for any function h on $\mathbf{\Omega} = \prod_{k=1}^{N} \Omega_k$ then for all $\mathbf{x} \in \mathbf{\Omega}$:

$$\text{If } \int_{\mathbf{\Omega}} h(\mathbf{x}) P_{n_1,\ldots,n_N}(\mathbf{x}) d\mu_\nu(\mathbf{x}) = 0 \text{ for all } n_1, \ldots, n_N \in \mathbb{N}_0$$

Then it follows that:

$$\int_{a_N}^{b_N} \ldots \int_{a_1}^{b_1} h(x_1, \ldots, x_N) \left(\prod_{k=1}^{N} P_{k,n_k}(x_k) f_{\nu_k}(x_k) \right) dx_1 \ldots dx_N = 0 \text{ for all } n_1, \ldots, n_N \in \mathbb{N}_0$$

and hence:

$$\int_{a_N}^{b_N}\left[\int_{a_{N-1}}^{b_{N-1}}\cdots\int_{a_1}^{b_1} h(x_1,\ldots,x_N)\left(\prod_{k=1}^{N-1} P_{k,n_k}(x_k)f_{\nu_k}(x_k)\right)dx_1\ldots dx_{N-1}\right]\times$$

$$P_{N,n_N}(x_N)f_{\nu_N}(x_N)dx_N = 0 \text{ for all } n_1,\ldots,n_N \in \mathbb{N}_0$$

Due to the fact $P_{N,n_N}(x_N)$ is complete, this infers that, for all $x_N \in \Omega_N$:

$$\int_{a_{N-1}}^{b_{N-1}}\cdots\int_{a_1}^{b_1} h(x_1,\ldots,x_N)\left(\prod_{k=1}^{N-1} P_{k,n_k}(x_k)f_{\nu_k}(x_k)\right)dx_1\ldots dx_{N-1} = 0$$

Applying the same reasoning we can see that for all $x_i \in \Omega_i (i = 1,\ldots,N)$ we have:

$$\int_{a_2}^{b_2}\left[\int_{a_1}^{b_1} h(x_1,\ldots,x_N)P_{1,n_1}(x_1)f_{\nu_1}(x_1)dx_1\right]P_{2,n_2}(x_2)f_{\nu_2}(x_2)dx_2 = 0 \text{ for all } n_1,n_2 \in \mathbb{N}_0$$

Which implies for all $\mathbf{x} \in \mathbf{\Omega}$:

$$\int_{a_1}^{b_1} h(x_1,\ldots,x_N)P_{1,n_1}(x_1)f_{\nu_1}(x_1)dx_1 = 0 \text{ for all } n_1 \in \mathbb{N}_0$$

Which finally implies that:

$$h(x_1,\ldots,x_N) = 0 \text{ for all } \mathbf{x} \in \mathbf{\Omega}$$

Therefore we can say that for any function h on $\mathbf{\Omega} = \prod_{k=1}^{N} \Omega_k$ then:

$$\text{If } \int_{\mathbf{\Omega}} h(\mathbf{x})P_{n_1,\ldots,n_N}(\mathbf{x})d\mu_\nu(\mathbf{x}) = 0 \text{ for all } n_1,\ldots,n_N \in \mathbb{N}_0 \text{ then } h(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \mathbf{\Omega}.$$

Therefore $\{P_{n_1,\ldots,n_N}(\mathbf{x})\}_{n_1,\ldots,n_N\in\mathbb{N}_0}$ are a complete set of multivariate orthonormal polynomials. □

3. MULTIVARIATE MOMENT-BASED DENSITY ESTIMATION

The multivariate polynomials given in Section 2.2 form a orthonormal base. In this section, we expand the joint density function $f_{\mathbf{X}}$ for the random vector $\mathbf{X} = (X_1,\ldots,X_N)$ on the orthonormal base $\{P_{n_1,\ldots,n_N}(\mathbf{x})\}_{n_1,\ldots,n_N\in\mathbb{N}_0}$.

Theorem 3.1. *The Multivariate Moment Based Density.*
Let $\mathbf{X} = (X_1,\ldots,X_N)$ be a random vector on a N-dimensional probability space $(\mathbf{\Omega}, \mathscr{F}, P)$ where $\mathbf{\Omega} = \prod_{k=1}^{N}\Omega_k = \prod_{k=1}^{N}[a_k,b_k]$. The entries of the random vector $\mathbf{X}$ are themselves random variables $X_k : \Omega_k \to \mathbb{R}$ for $k = 1,\ldots,N$. Denote the joint probability density function of $\mathbf{X}$ as $f_{\mathbf{X}} : \mathbb{R}^N \to \mathbb{R}$. Assume $\mathbf{X}$ has finite moments $\int_\Omega x_1^{i_1}\ldots x_N^{i_N} f_X(\mathbf{x})dx_1\ldots dx_N$ for all $i_1,\ldots,i_N \in \mathbb{N}_0$, and $\frac{f_X}{f_\nu} \in L_2(\mathbf{\Omega},\mu_\nu)$. Then $f_{\mathbf{X}}$ can be expressed as:

$$f_{\mathbf{X}}(\mathbf{x}) = f_\nu(\mathbf{x})\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N}P_{n_1,\ldots,n_N}(\mathbf{x}) \tag{7}$$

Where the coefficient terms $C_{n_1,\ldots,n_N}$ are given by:

$$C_{n_1,\ldots,n_N} = \mathbb{E}_{\mathbf{X}}\left[P_{n_1,\ldots,n_N}(X_1,\ldots,X_N)\right] = \int_{\Omega} P_{1,n_1}(x_1)\ldots P_{N,n_N}(x_N) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \tag{8}$$

Proof. By the Weierstrass-Stone Approximation (Szökefalvi-Nagy, 1965) we know there exists a unique polynomial approximation which converges uniformly to the continuous function $\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})}$ hence let:

$$\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})} = \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})$$

Given $\frac{f_{\mathbf{X}}}{f_{\nu}} \in L_2(\mathbf{\Omega},\mu_{\nu})$, then for fixed $m_1,\ldots,m_N \in \mathbb{N}_0$ by the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\int_{\mathbf{\Omega}}\left|\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})\right|d\mu_{\nu}(\mathbf{x})\\
&\qquad = \int_{\mathbf{\Omega}}\left|\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})}P_{m_1,\ldots,m_N}(\mathbf{x})\right|d\mu_{\nu}(\mathbf{x})\\
&\qquad \le \sqrt{\int_{\mathbf{\Omega}}\left|\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})}\right|^2 d\mu_{\nu}(\mathbf{x})\int_{\mathbf{\Omega}} P^2_{m_1,\ldots,m_N}(\mathbf{x})d\mu_{\nu}(\mathbf{x})}\\
&\qquad = \sqrt{\int_{\mathbf{\Omega}}\left|\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})}\right|^2 d\mu_{\nu}(\mathbf{x})} < \infty
\end{aligned}$$

Therefore by the Fubini-Tonelli theorem

$$\begin{aligned}
&\int_{\mathbf{\Omega}}\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_{\nu}(\mathbf{x})\\
&\qquad = \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N}\left[\int_{\mathbf{\Omega}} P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_{\nu}(\mathbf{x})\right]\\
&\qquad = \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N}\left[\prod_{k=1}^{N}\delta_{n_k,m_k}\right]\\
&\qquad = C_{m_1,\ldots,m_N}
\end{aligned}$$

Hence

$$C_{m_1,\ldots,m_N} = \int_{\mathbf{\Omega}}\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})}P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_{\nu}(\mathbf{x}) = \int_{\mathbf{\Omega}} P_{m_1,\ldots,m_N}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \mathbb{E}_{\mathbf{X}}\left[P_{m_1,\ldots,m_N}(\mathbf{X})\right]$$

Therefore by the uniqueness of the polynomial approximation then:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\nu}(\mathbf{x})\sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})$$

Where the coefficient terms $C_{n_1,\ldots,n_N}$ are given by:

$$C_{n_1,\ldots,n_N} = \mathbb{E}_{\mathbf{X}}\left[P_{n_1,\ldots,n_N}(X_1,\ldots,X_N)\right] = \int_{\Omega} P_{1,n_1}(x_1)\ldots P_{N,n_N}(x_N) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

Remark: It should be also noted that the existence of all moments of $\mathbf{X}$ ensures that $C_{n_1,\ldots,n_N} < \infty$ □

Corollary 3.1.1. *The coefficients $C_{n_1,\ldots,n_N}$ of the multivariate polynomial $P_{n_1,\ldots,n_N}(\mathbf{x})$ in the expansion of $f_{\mathbf{X}}(\mathbf{x})$ are given as a linear combination of joint moments:*

$$C_{n_1,\ldots,n_N} = \sum_{i_1=0}^{n_1} \cdots \sum_{i_N=0}^{n_N} a_{n_1,\ldots,n_N;i_1,\ldots,i_N} \mathbb{E}_{\mathbf{X}} \left[X_1^{i_1} \ldots X_N^{i_N} \right]$$

Where $a_{n_1,\ldots,n_N;i_1,\ldots,i_N} = \prod_{k=1}^{N} a_{n_k,i_k}$ where a_{n_k,i_k} is the i_kth coefficient of the n_kth polynomial $P_{k,n_k}(x_k)$.

Corollary 3.1.2. *The sum of the squared coefficients is equal to the expected value of the ratio of the true density function and the reference density function which is finite.*

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N}^2 = \int_{\Omega} \frac{|f_{\mathbf{X}}(\mathbf{x})|^2}{f_{\nu}(\mathbf{x})} d\mathbf{x} = \mathbb{E}_{\mathbf{X}} \left[\frac{f_{\mathbf{X}}(X_1,\ldots,X_N)}{f_{\nu}(X_1,\ldots,X_N)} \right] < \infty$$

We note that $\sum_{n_1}^{\infty} \cdots \sum_{n_N}^{\infty} C_{n_1,\ldots,n_N}^2 < \infty$ is a direct result of $\frac{f_{\mathbf{X}}}{f_{\nu}} \in L_2(\mathbf{\Omega}, \mu_{\nu})$. This also infers that $C_{n_1,\ldots,n_N} \to 0$ as $n_1,\ldots,n_N \to \infty$.

It should also be noted that these polynomials may also be extended to unbounded spaces however moment determinacy may become a considerable issue in the moment based expansion of the joint density. Regardless, examples of reference densities (f_{k,ν_k}) which also satisfies these conditions and can be used in the expansion include normal densities (Hamburger Case) and gamma densities (Stieltjes Case) and have been previously studied. However, when looking at examples of these types of polynomials in the literature, it is often assumed that all marginal densities of the joint density being approximated have a consistent type of distribution. Here we make no such assumption, and only ensure that the domain of our multivariate polynomials is consistent with that of our joint density (provided that $\frac{f_X}{f_{\nu}} \in L_2(\mathbf{\Omega}, \mu_{\nu})$ and has finite moments).

A considerable issue with this type of expansion is that it requires an infinite sum of polynomials. In practice, it is fairly untenable to perform an infinite sum without requiring some kind of analytic solution. This therefore becomes a serious problem when hoping to use this formulation to create a sample estimate of the joint density function. Instead we will ultimately have to use the finite order multivariate moment-based density which is defined as follows:

Definition 3. The Finite Order Multivariate Moment Based Density.
The finite order multivariate moment-based density approximation of max order $\mathbf{K} = (K_1,\ldots,K_N)$ is defined as:

$$f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) = f_{\nu}(\mathbf{x}) \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})$$

Theorem 3.2. *The finite order multivariate moment-based density converges to the true density as all the elements of the max polynomial order vector* $\mathbf{K}$ *approaches infinity.*

$$\lim_{K_1\to\infty\ \dots\ K_N\to\infty} \int_{\mathbf{\Omega}} \left| \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X};\mathbf{K}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu = 0$$

Where $f_{\mathbf{X};\mathbf{K}}$ *is as defined in Definition 3.*

Proof.

$$\begin{aligned}
&\int_\Omega \left| \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X};\mathbf{K}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu \\
&= \int_\Omega \left| \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu - 2\int_\Omega \frac{f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{X};\mathbf{K}}(\mathbf{x})}{f_\nu^2(\mathbf{x})} d\mu_\nu + \int_\Omega \left| \frac{f_{\mathbf{X};\mathbf{K}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\dots,n_N} - 2\int_\Omega \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \sum_{m_1=0}^{K_1} \cdots \sum_{m_N=0}^{K_N} \Big[C_{n_1,\dots,n_N} C_{m_1,\dots,m_N} P_{n_1,\dots,n_N}(\mathbf{x}) \\
&\qquad\qquad P_{m_1,\dots,m_N}(\mathbf{x}) \Big] d\mu_\nu + \int_\Omega \left| \sum_{m_1=0}^{K_1} \cdots \sum_{m_N=0}^{K_N} C_{m_1,\dots,m_N} P_{m_1,\dots,m_N}(\mathbf{x}) \right|^2 d\mu_\nu \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\dots,n_N} - 2\sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C^2_{n_1,\dots,n_N} + \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C^2_{n_1,\dots,n_N} \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\dots,n_N} - \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C^2_{n_1,\dots,n_N}
\end{aligned}$$

Therefore as

$\sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\dots,n_N}$ is the limit of $\sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C^2_{n_1,\dots,n_N}$ as $K_i \to \infty$ for each $i = 1, \dots, N$

□

Example 3.1. Consider the Bivariate Beta Distribution discussed in Macomber and Myers (1983) which has density function:

$$f_{X_1,X_2}(x_1,x_2) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} (1-x_1-x_2)^{\alpha_3-1}$$

For $x_1, x_2 > 0$ and $x_1 + x_2 \leq 1$.

The moments of this bivariate density function are given by (Macomber and Myers, 1983):

$$E(X_1^m X_2^n) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3)\Gamma(\alpha_1+m)\Gamma(\alpha_2+n)}{\Gamma(\alpha_1+\alpha_2+\alpha_3+m+n)\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

Considering the case where $\alpha_1 = 3, \alpha_2 = 3$ and $\alpha_3 = 2$ and using a uniform reference density defined on $[0,1] \times [0,1]$ given by:

$$f_\nu(\nu_1,\nu_2) = 1 \qquad \text{where } \nu_1,\nu_2 \in [0,1]$$

We can obtain a polynomial approximation of the bivariate beta density function. It should be noted that this reference density does not impose the restriction that $x_1 + x_2 \leq 1$ and hence this property will have to be picked up in the moment information of the density function.

The first few terms are given by:

$$
\begin{aligned}
C_{0,0}P_{0,0}(x_1,x_2) &= 1 \times 1 = 1 \\
C_{0,1}P_{0,1}(x_1,x_2) &= \sqrt{3}\left(-1 + 2\frac{3}{8}\right) \times \sqrt{3}\left(-1 + 2x_2\right) = -\frac{3}{4}\left(-1 + 2x_2\right) \\
C_{0,2}P_{0,2}(x_1,x_2) &= \sqrt{\frac{5}{4}}\left(2 - 12\frac{3}{8} + 12\frac{1}{6}\right) \times \sqrt{\frac{5}{4}}\left(2 - 12x_2 + 12x_2^2\right) = -\frac{5}{4}\left(1 - 6x_2 + 6x_2^2\right) \\
C_{1,0}P_{1,0}(x_1,x_2) &= \sqrt{3}\left(-1 + 2\frac{3}{8}\right) \times \sqrt{3}\left(-1 + 2x_2\right) = -\frac{3}{4}\left(-1 + 2x_1\right) \\
C_{2,0}P_{2,0}(x_1,x_2) &= \sqrt{\frac{5}{4}}\left(2 - 12\frac{3}{8} + 12\frac{1}{6}\right) \times \sqrt{\frac{5}{4}}\left(2 - 12x_1 + 12x_1^2\right) = -\frac{5}{4}\left(1 - 6x_1 + 6x_1^2\right) \\
C_{1,1}P_{1,1}(x_1,x_2) &= 3\left(1 - 2\frac{3}{8} - 2\frac{3}{8} + 4\frac{1}{8}\right) \times 3\left(1 - 2x_1 - 2x_2 + 4x_1x_2\right) = 0 \\
&\vdots
\end{aligned}
$$

Applying these terms in the approximation for varying levels of maximum order of polynomial used, we obtain plots depicted in Figure 1 that show a convergence towards the true bivariate density function depicted in Figure 2

The approximations given in Example 3.1 require the true moment information for the joint density. However, in a practical situation the moment information is usually unavailable and the only available information is a random sample from the distribution.
Therefore we will now consider a sample moment based approximation of multivariate density functions.

4. MULTIVARIATE SAMPLE MOMENT-BASED DENSITY ESTIMATION

Definition 4. Multivariate Sample Moment-Based Density Approximation.
Given a random sample $\{\mathbf{X}_i\}_{i=1}^M = \{(X_{1,i}, X_{2,i}, \ldots, X_{N,i})\}_{i=1}^M$ of $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ respectively the sample moment-based estimator of $f_{\mathbf{X}}(\mathbf{x})$ with max polynomial order $\mathbf{K} = (K_1, \ldots, K_N)$ is given by:

$$\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M) = f_\nu(\mathbf{x}) \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} \hat{C}_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{x})$$

Where $\hat{C}_{n_1,\ldots,n_N} = \frac{1}{M}\sum_{j=0}^{M} P_{n_1,\ldots,n_N}(X_{1,j}, \ldots, X_{N,j})$ is a function of $\{\mathbf{X}_i\}_{i=1}^M$.

Theorem 4.1. *The sample moment-based density estimator of a fixed max polynomial order* $\mathbf{K}$ *converges almost surely to the finite order moment-based density uniformly with max order parameter* $\mathbf{K} = (K_1, \ldots, K_N)$ *as the sample size increases, that is:*

$$\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M) \xrightarrow{a.s.} f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) \text{ as } M \to \infty \text{ uniformly for each fixed } \mathbf{K}.$$

A complete proof of this theorem is given in the supplementary materials under Appendix B.

Applying this type of approximation to estimate the joint density function also requires some kind of metric to determine the quality of the estimate. This will inevitably entail some type of averaged measure of divergence from the true joint density function. For this reason we will consider the $\mathcal{N}_{\mathbf{K}}$-Norm defined below. This $\mathcal{N}_{\mathbf{K}}$-Norm denoted $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ is a metric by which goodness of fit of the approximation can be ascertained.

Definition 5. $\mathcal{N}_{\mathbf{K}}$-Norm of the Approximation.
We define the $\mathcal{N}_{\mathbf{K}}$-Norm of the sample $\{\mathbf{X}_i\}_{i=1}^M$ denoted $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ as the value of the $L_2(\mathbf{\Omega}, \mu_\nu)$ norm of the ratio of the difference between the sample approximation $\hat{f}_{\mathbf{X};\mathbf{K}}\left(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M\right)$ and the true joint density $f_{\mathbf{X}}(\mathbf{x})$, and the reference joint density $f_\nu(\mathbf{x})$. $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ can be expressed as:

$$\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right) = \int_\Omega \left| \frac{\hat{f}_{\mathbf{X};\mathbf{K}}\left(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M\right)}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu$$

The $\mathcal{N}_{\mathbf{K}}$-Norm is itself a function of the sample and is therefore a statistic. Hence we should also consider the expected value of the $\mathcal{N}_{\mathbf{K}}$-Norm $\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right]$.

Theorem 4.2. *The expected value of* $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ *is given by:*

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] = \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) + \mathcal{T}_{\mathbf{K}}$$

Where $\mathcal{T}_{\mathbf{K}} = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} C^2_{n_1,\ldots,n_N}$ *denotes the remainder term.*

A proof of this theorem can be found in the supplementary materials under Appendix C.

Lemma 4.3. *If the expected value of the* $\mathcal{N}_{\mathbf{K}}$*-Norm of the approximation is 0 then the approximation is almost surely equal to the true density function, that is:*

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] = 0 \Rightarrow \hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M) = f_{\mathbf{X}}(\mathbf{x}) \text{ almost surely } \mu_{\mathbf{X}}.$$

Proof. Given

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] = \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) + \mathcal{T}_{\mathbf{K}} = 0$$

Then it follows as

$$\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) \geq 0 \qquad \text{and} \qquad \mathcal{T}_{\mathbf{K}} \geq 0$$

That:

$$\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) = 0 \qquad \text{and} \qquad \mathcal{T}_{\mathbf{K}} = 0$$

Hence

$$\begin{aligned}
&\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) = 0 \\
&\Rightarrow Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) = 0 \text{ for all } n_1 \leq K_1, n_2 \leq K_2, \ldots, n_N \leq K_N \\
&\Rightarrow \hat{C}_{n_1,\ldots,n_N} = C_{n_1,\ldots,n_N} \text{ almost surely } \mu_{\mathbf{X}} \text{ for all } n_1 \leq K_1, n_2 \leq K_2, \ldots, n_N \leq K_N \\
&\Rightarrow \hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}) = f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) \text{ almost surely } \mu_{\mathbf{X}}.
\end{aligned}$$

And

$$\begin{aligned}
\mathcal{T}_{\mathbf{K}} = 0 &\Rightarrow \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} C^2_{n_1,\ldots,n_N} = 0 \\
&\Rightarrow C_{n_1,\ldots,n_N} = 0 \text{ for all } n_1 > K_1, n_2 > K_2, \ldots, n_N > K_N \\
&\Rightarrow f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x})
\end{aligned}$$

Therefore

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] = 0 \Rightarrow \hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}) \overset{a.s.\mu_{\mathbf{X}}}{=} f_{\mathbf{X};\mathbf{K}} = f_{\mathbf{X}} \text{ almost surely.}$$

□

Importantly, the expected value of $\mathcal{N}_{\mathbf{K}}$-Norm clearly demonstrates the dichotomy associated with applying the sample based approximation. Increasing the order of the approximation will undoubtedly lead to a reduction in the remainder term $\mathcal{T}_{\mathbf{K}}$. However, the increase in the order of the approximation also inevitably results in an increase of the number of parameters needed to be estimated and hence inflates the aggregated variance of the coefficient estimators. This tradeoff between the remainder term and the estimator variance is at the heart of determining the optimal strategy for approximating the joint density function. It is therefore important that we define an **optimal max polynomial order** to the approximation and develop methods to determine this value (or vector).

Definition 6. An Optimal Max Polynomial Order.

An optimal max polynomial order denoted $\mathbf{K}^* = (K_1^*, \ldots, K_N^*)$ is defined as a vector of max polynomial orders in the approximation such that:

$$\mathbb{E}\left[\mathcal{N}_{\mathbf{K}^*}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] \leq \mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right]$$

For any possible max polynomial order $\mathbf{K} = (K_1, \ldots, K_N)$.

Theorem 4.4. *Suppose there exists $\epsilon_0 > 0$ such that $Var_{\mathbf{X}}(\hat{C}_{n_1,\ldots,n_N}) \geq \epsilon_0$ for all $n_1, \ldots, n_N \in \mathbb{N}_0$. Then the expected $\mathcal{N}_{\mathbf{K}}$-Norm, denoted $\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right]$ (Definition 5), has a finite optimal max polynomial order.*

A proof of this theorem is given in the supplementary materials under Appendix D for reading convenience.

Therefore we know that for any examples where the variance of the coefficient estimators has a positive lower bound, then it follows that we can determine a finite optimal max polynomial order. However, in a practical situation the true joint density of a sample may not necessarily be known and hence it is untenable to calculate the $\mathcal{N}_{\mathbf{K}}$-Norm statistic. Although for the purposes of determining the optimum max polynomial order, it is also not necessarily critical that the value of the $\mathcal{N}_{\mathbf{K}}$-Norm needs to be known, just minimised.

Instead we will consider minimising the shifted $\mathcal{N}_{\mathbf{K}}$-Norm denoted $\mathcal{N}^*_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ as this term requires less information about the true density but has the same minimum as the $\mathcal{N}_{\mathbf{K}}$-Norm.

Definition 7. The Shifted L_2 Norm.

The shifted $\mathcal{N}_{\mathbf{K}}$-Norm denoted $\mathcal{N}^*_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ is defined as:

$$\mathcal{N}^*_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right) = \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} \hat{C}^2_{n_1,\ldots,n_N} - 2\mathbb{E}_{\mathbf{X}}\left[\frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{X}; \{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{X})}\right]$$

Lemma 4.5. *Given an approximation $\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^M)$ with vector of max polynomial order $\mathbf{K}$, of the joint density $f_{\mathbf{X}}$ then an optimum minimum max polynomial order $\mathbf{K}^*$ of $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ is also an optimum minimum max polynomial order of the shifted $\mathcal{N}_{\mathbf{K}}$-Norm $\mathcal{N}^*_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$.*

Proof. Let the optimum minimum max polynomial order of $\mathcal{N}^*_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ be denoted by $\mathbf{K}^*_1$.

By noting $\int_\Omega \frac{|f_{\mathbf{X}}(\mathbf{x})|^2}{f_\nu(\mathbf{x})} d\mathbf{x}$ is free from $\mathbf{K}$, we have:

$$\begin{aligned}
\mathbf{K}_1^* &= \text{argmin}_{\mathbf{K}} \left\{ \mathcal{N}_{\mathbf{K}}^*\left(\{\mathbf{X}_i\}_{i=1}^M\right) \right\} \\
&= \text{argmin}_{\mathbf{K}} \left\{ \int_\Omega \left| \frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{x})} \right|^2 f_\nu(\mathbf{x}) d\mathbf{x} - 2\int_\Omega \frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{x})} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right\} \\
&= \text{argmin}_{\mathbf{K}} \left\{ \int_\Omega \left[\left| \frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{x})} \right|^2 f_\nu(\mathbf{x}) - 2\frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{x})} f_{\mathbf{X}}(\mathbf{x}) + \frac{|f_{\mathbf{X}}(\mathbf{x})|^2}{f_\nu(\mathbf{x})} \right] d\mathbf{x} \right\} \\
&= \text{argmin}_{\mathbf{K}} \left\{ \int_\Omega \left| \frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 f_\nu(\mathbf{x}) d\mathbf{x} \right\} = \text{argmin}_{\mathbf{K}} \left\{ \mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^M\right) \right\} = \mathbf{K}^*
\end{aligned}$$

□

It should be noted that the shifted $\mathcal{N}_{\mathbf{K}}$-Norm $\mathcal{N}_{\mathbf{K}}^*\left(\{\mathbf{X}_i\}_{i=1}^M\right)$ is not necessarily greater than 0 but still has a lower bound. Also, the term still requires the true joint density in the evaluation of the expectation of the estimated density relative to the reference density. However, we do have a sample from the true density and hence as for a function $g(\mathbf{X})$, we can estimate the expectation with it's sample mean i.e. $\mathbb{E}\left[g(\mathbf{X})\right] \approx \frac{1}{M}\sum_{i=1}^M g(\mathbf{X}_i)$.

Therefore given another independent random sample $\{\mathbf{X'}_i\}_{i=1}^{M_0} = \{(X'_{1,i},\ldots,X'_{N,i})\}_{i=1}^{M_0}$ of $X_1, X_2, \ldots, X_N$ with joint density function $f_{\mathbf{X}}(\mathbf{x})$ and the estimated joint density $\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M)$ we can obtain an estimate of the shifted $\mathcal{N}_{\mathbf{K}}$-Norm:

$$\mathcal{N}_{\mathbf{K}}^*\left(\{\mathbf{X}_i\}_{i=1}^M\right) = \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \hat{C}_{n_1,\ldots,n_N}^2 - 2\mathbb{E}_{\mathbf{X}}\left[\frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{X};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(\mathbf{X})}\right]$$

Given by:

$$\begin{aligned}
\hat{\mathcal{N}}_{\mathbf{K}}^*\left(\{\mathbf{X}_i\}_{i=1}^M;\{\mathbf{X'}_i\}_{i=1}^{M_0}\right) = \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \hat{C}_{n_1,\ldots,n_N}(\{\mathbf{X}_i\}_{i=1}^M)^2 \\
- \frac{2}{M_0}\sum_{j=1}^{M_0} \frac{\hat{f}_{\mathbf{X};\mathbf{K}}(X'_{1,j},\ldots,X'_{N,j};\{\mathbf{X}_i\}_{i=1}^M)}{f_\nu(X'_{1,j},\ldots,X'_{N,j})}
\end{aligned}$$

Of course it should be noted that there is an issue with using the same sample to estimate the coefficients and to estimate the shifted $\mathcal{N}_{\mathbf{K}}$-Norm. This is due to the fact the estimated coefficients are themselves a function of the sample and therefore taking an expectation with respect to the one sample should interact with both components.

One strategy to overcome this problem is to randomly partition the sample and use one partition of the observations in the estimation of the coefficients and the other in the estimation of the shifted $\mathcal{N}_{\mathbf{K}}$-Norm. By performing bootstrapping of this process we can also attempt to utilise more information from the sample whilst still preserving the independence assumption and obtaining standard error estimates as well.

We propose the following strategy for estimating the shifted $\mathcal{N}_{\mathbf{K}}$-Norm given a random sample $\{\mathbf{X}_i\}_{i=1}^{M}$.

1. Randomly partition observations of the random sample into two approximately equal partitions $\{\mathbf{X}_i\}_{i\in\mathcal{I}}$ and $\{\mathbf{X}_i\}_{i\in\overline{\mathcal{I}}}$.
2. Calculate the statistic:

$$\hat{\mathcal{N}}^{*}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i\in\mathcal{I}};\{\mathbf{X}_i\}_{i\in\overline{\mathcal{I}}}\right)=\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\hat{C}_{n_1,\ldots,n_N}(\{\mathbf{X}_i\}_{i\in\mathcal{I}})^2 - \frac{2}{||\overline{\mathcal{I}}||}\sum_{j\in\overline{\mathcal{I}}}\frac{\hat{f}_{\mathbf{X};\mathbf{K}}(X_{1,j},\ldots,X_{N,j};\{\mathbf{X}_i\}_{i\in\mathcal{I}})}{f_{\nu}(X_{1,j},\ldots,X_{N,j})}$$

3. Repeat 1 and 2 B times to obtain the set of bootstrapped shifted $\mathcal{N}_{\mathbf{K}}$-Norm statistics $\left\{\hat{\mathcal{N}}^{*}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i\in\mathcal{I}_b};\{\mathbf{X}_i\}_{i\in\overline{\mathcal{I}_b}}\right)\right\}_{b=1}^{B}$
4. Calculate $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)=\frac{1}{B}\sum_{b=1}^{B}\hat{\mathcal{N}}^{*}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i\in\mathcal{I}_b};\{\mathbf{X}_i\}_{i\in\overline{\mathcal{I}_b}}\right)$.

5. SIMULATION STUDY

Consider the truncated normal-gamma bivariate distribution defined by the density function:

$$f_{X,T}(x,t)\propto\frac{\beta^{\alpha}\sqrt{\lambda}}{\Gamma(\alpha)\sqrt{2\pi}}t^{\alpha-\frac{1}{2}}\exp\left(-\beta t-\frac{\lambda t(x-\mu)^2}{2}\right)$$

Where $\mu=2$, $\lambda=\frac{1}{5}, \alpha=6$ and $\beta=3$ defined on $x\in[-2.862,6.862]$ and $t\in(0,5.053]$ and is depicted in Figure 3. As in the previous example, we will again use a uniform reference density, however now we define the reference density on the domain $[-2.862,6.862]\times[0,5.053]$.

Therefore our reference density is given by:

$$f_{\nu}(\nu_1,\nu_2)=\frac{1}{6.862-(-2.862)}\times\frac{1}{5.053-0}\approx 0.02035$$

Where $\nu_1\in[-2.862,6.862]$ and $\nu_2\in[0,5.053]$.

By simulating a sample of size $M=1000$ we can obtain a sample moment-based density estimate of the normal-gamma density function. Figure 4 depicts this sample estimate where a maximum order of polynomial of 6 was used for each variable ($\mathbf{K}=(6,6)$). This figure includes both the marginal density values as well as the bivariate heat map. As can be seen from Figure 4 the estimate for a max polynomial order of $\mathbf{K}=(6,6)$ serves as a fairly good approximation for the true bivariate density shown in Figure 3.

We can see how this max polynomial order varies the quality of the approximation by performing a simulation study. We can also ascertain the ability of the bootstrapped-estimated shifted $\mathcal{N}_{\mathbf{K}}$-Norm $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)$ to adequately model the $\mathcal{N}_{\mathbf{K}}$-Norm. For simplicity in this simulation study, we will assume that the max polynomial order is consistent across both variables i.e. $\mathbf{K} = (K, K)$. This is certainly not necessary to do in general and usually we would not restrict this to a one dimensional optimisation problem. However correctly visualising bivariate optimisation problems in plots can be difficult to do and hence in order to properly demonstrate the nature of the optimisation problem in this simulation study we have reduced the dimension of the optimisation.

Consider 1000 samples of size 1000 from the normal-gamma density function. For each jth sample we calculate both the true $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right)$ and the bootstrapped-estimated shifted $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right)$. In order to compare these two statistics, it is also necessary to compute the constant term used in the shift.

This term is given by:

$$\int_{-2.862}^{6.862}\int_{0}^{5.053}\left|\frac{f_{X,T}(x,t)}{f_{\nu}(x,t)}\right|^2 dtdx \approx 3.1379$$

Figure 5 demonstrates plots of $\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right)$ (red) and $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right) + 3.1379$ (blue) over different max polynomial orders K. The corrected bootstrapped-estimated shifted $\mathcal{N}_{\mathbf{K}}$-Norm $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right) + 3.1379$ is calculated by randomly partitioning each sample in the simulation study into two distinct partitions of equal size 100 times ($B = 100$). For each pair of partitions we recalculate the coefficients using the observations in the first partition and then calculate the statistic using the observations in the second partition. We then average the norm values for each max polynomial order of the 100 bootstrapped samples to obtain the value of $\hat{\mathcal{N}}^{*}_{\mathbf{K};B}\left(\{\mathbf{X}_i^{(j)}\}_{i=1}^{M}\right) + 3.1379$ for each max polynomial order K and each jth sample of the simulation study.

These plots show the average value across the 1000 samples (solid line) and the 95% confidence intervals (region) bounded in dashed lines. Furthermore Figure 5a also shows a dotted line of the estimated $\mathcal{N}_{\mathbf{K}}$-Norm when the same sample is used to estimate the coefficients and the $\mathcal{N}_{\mathbf{K}}$-Norm.

In order to get an understanding of the quality of such a fit, we also apply a bivariate Kernel Density Estimation method to each sample. In doing so we would like to be sure that our moment-based approach can get similar, if not better norm values then that of the widely used kernel density estimators. As kernel density estimation is not the focus of this paper a detailed explanation of the method of calculation will not be given here. However, as we applied the **ks** package in R (Duong, 2018) using the Hpi plug-in bandwidth selector to obtain the kernel density estimates, complete explanation of the way in which these density estimates were calculated can be found in the literature for the package and references therein.

We define the $\mathcal{N}_{Kernal}$-Norm by:

$$\mathcal{N}_{Kernel}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right) = \int_{\Omega} \left| \frac{\hat{f}_{Kernel}\left(\mathbf{x}; \{\mathbf{X}_i\}_{i=1}^{M}\right)}{f_{\nu}(\mathbf{x})} - \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\nu}(\mathbf{x})} \right|^2 d\mu_{\nu}$$

Figure 5b depicts the average $\mathcal{N}_{Kernal}$-Norm $\mathcal{N}_{Kernel}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)$ (green) and the associated 95% confidence interval.

As can be seen from Figure 5, it is clear that there is a finite minimum max polynomial order which is the point just before the aggregated variance of the estimated coefficients becomes greater than that of the error associated with truncating the polynomial approximation.

Furthermore it shows that the greater the max polynomial order, the greater variation associated with both the true and estimated $\mathcal{N}_{\mathbf{K}}$-Norms. Figure 5a also shows the effect of using the complete sample in both the estimation of the coefficients and the $\mathcal{N}_{\mathbf{K}}$-Norm. This effect is shown by the bias of the dashed lines which appears to increase as the max polynomial order increases.

Regardless, bootstrapping the estimated $\mathcal{N}_{\mathbf{K}}$-Norm with equal partitions used to estimate the coefficients and the $\mathcal{N}_{\mathbf{K}}$-Norm results in our estimated optimal max polynomial order being determined as $\hat{\mathbf{K}}^* = (6, 6)$ (marked as the blue circle on Figure 5b), which is exactly the same as the true optimal max polynomial order $\mathbf{K}^* = (6, 6)$ (marked as the red circle on Figure 5b).

This method confirms that it is possible to obtain the optimal max polynomial order without any knowledge of the true density function. Of course one of the key problems with applying this strategy is the large run times required to perform the bootstrapping. However, the number of bootstrapped samples is within our control and the computation costs can be offset by allowing for greater variation of the estimates.

Furthermore we can see that the moment-based density estimation method, when compared to the kernel density estimate, obtains similar if not improved (although not significantly) norm values then that of the moment-based density estimate at the optimal max polynomial order. This suggests that moment-based density estimation is a reasonable alternative to kernel density estimation especially when we consider that the moment-based density estimate takes the form of a simple polynomial function whereas kernel density estimates, especially when a complex kernel function is used, can take a more complex form.

Regardless, we have demonstrated a method by which the optimal max polynomial order can be estimated. This, combined with other validation techniques such as evaluating the marginal histograms, bivariate histograms and goodness of fit tests such as the Kolmogorov-Smirnov test, can all be used to verify which max polynomial order would be best.

For instance, performing this process with respect to the first sample of the simulation study which was used in the approximation in Figure 4, then we would obtain the plot of the varying Kolmogorov-Smirnov (K-S) statistics for each max polynomial order shown in Figure 6. Figure 6 clearly shows that although the optimum max polynomial order value (based on the smallest K-S value) is different to the optimum max polynomial order obtained from both the average estimated and true $\mathcal{N}_{\mathbf{K}}$-Norm, there is not a considerable difference between the K-S statistic evaluated at each of the different statistics, especially if we were to also apply a penalty for the increased number of parameters required for larger max polynomial orders.

Given the focus of this paper we will leave both a more detailed discussion as to why a parameter penalty may be warrented on the Kolmogorov-Smirnov test statistic, and an analysis of the stability of this statistic, as an area for further research.

6. CUMULATIVE DISTRIBUTION FUNCTION AND SAMPLING

One of the key advantages to applying a moment-based estimation method to approximate multivariate density functions is the fact that we can obtain a generalised polynomial expression of the joint density function. These types of expressions are analytically fairly straightforward to manipulate by both integration and differentiation. This becomes especially useful when attempting to estimate the cumulative distribution function of a multivariate density function. Unlike other methods, we can restrict the computation of probabilities to only the probabilities required to be calculated and not requiring some type of numeric integration or interpolating frequency counts of binned observations.

In order to approximate the cumulative distribution function we must first consider how the cumulative distribution function can be expressed in terms of it's moments.

Lemma 6.1. *The Multivariate Moment-Based Cumulative Distribution Function.*
Given $f_{\mathbf{X}}(\mathbf{x})$ can be expressed in a polynomial form, it follows the cumulative distribution function $F_{\mathbf{X}}(\mathbf{x}) = \int_{a_1}^{x_1} \dots \int_{a_N}^{x_N} f_{\mathbf{X}}(\mathbf{t}) dt_1 \dots dt_N$ can be expressed as:

$$F_{\mathbf{X}}(\mathbf{x}) = f_{\nu}(\mathbf{x}) \left[\prod_{k=1}^{N} \sigma_k(x_k) \right] \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C_{n_1,\dots,n_N} \frac{\mathcal{D} P_{n_1,\dots,n_N}(\mathbf{x})}{\lambda_{1,n_1} \dots \lambda_{N,n_N}}$$

Where $\mathcal{D}$ refers to the differential operator $\frac{\partial^N}{\partial x_1 \dots \partial x_N}$

Proof. Applying the Remark under Lemma 2.2 it follows that:

$$\begin{aligned}
F_{\mathbf{X}}(\mathbf{x}) &= \int_{a_1}^{x_1} \cdots \int_{a_N}^{x_N} f_{\mathbf{X}}(\mathbf{t}) dt_1 \ldots dt_N \\
&= \int_{a_1}^{x_1} \cdots \int_{a_N}^{x_N} f_{\nu}(\mathbf{t}) \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} P_{n_1,\ldots,n_N}(\mathbf{t}) dt_1 \ldots dt_N \\
&= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} \left[\prod_{k=1}^{N} \frac{\sigma_k(x)}{\lambda_{k,n_k}} P'_{k,n_k}(x_k) f_{\nu_k}(x_k) \right] \\
&= f_{\nu}(\mathbf{x}) \left[\prod_{k=1}^{N} \sigma_k(x_k) \right] \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C_{n_1,\ldots,n_N} \frac{\mathcal{D} P_{n_1,\ldots,n_N}(\mathbf{x})}{\lambda_{1,n_1} \ldots \lambda_{N,n_N}}
\end{aligned}$$

Where $\mathcal{D}$ refers to the differential operator $\frac{\partial^N}{\partial x_1 \ldots \partial x_N}$. □

By applying the same reasoning used to derive our sample moment-based joint density approximation discussed in the previous section, when attempting to estimate the cumulative distribution function based on a sample we must also apply a finite max order on the polynomial terms and use the sample moments when estimating the coefficients. However, it also follows that the optimal max polynomial order of the density function is also the optimal max polynomial order of the cumulative distribution function. Furthermore, we can just apply the same reasoning use to integrate the true moment-based density expression to that of the sample moment-based estimate and obtain the following expression of the estimated cumulative distribution function.

$$\hat{F}_{\mathbf{X};\mathbf{K}}(\mathbf{x}) = f_{\nu}(\mathbf{x}) \left[\prod_{k=1}^{N} \sigma_k(x_k) \right] \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} \hat{C}_{n_1,\ldots,n_N} \frac{\mathcal{D} P_{n_1,\ldots,n_N}(\mathbf{x})}{\lambda_{1,n_1} \ldots \lambda_{N,n_N}}$$

Where $\mathcal{D}$ refers to the differential operator $\frac{\partial^N}{\partial x_1 \ldots \partial x_N}$.
This process also applies when considering the conditional cumulative distribution function of two variables.

Lemma 6.2. *Given random variables $X \in [a_X, b_X]$ and $Y \in [a_Y, b_Y]$ with joint density function $f_{X,Y}(x,y)$ which can be expressed in a polynomial form, it follows the conditional cumulative distribution function $F_{X|Y}(x|y) = \int_{a_X}^{x} \frac{f_{X,Y}(t,y)}{f_Y(y)} dt$ can be expressed as:*

$$F_{X|Y}(x|y) = \frac{\sigma_X(x) f_{\nu_X}(x)}{\sum_{n=0}^{\infty} C_{0,n} P_{Y,n}(y)} \sum_{n_X=0}^{\infty} \sum_{n_Y=0}^{\infty} \frac{C_{n_X,n_Y}}{\lambda_{X,n_X}} P'_{X,n_X}(x) P_{Y,n_Y}(y)$$

Where f_{ν_X} is the reference density function for X.

Proof. Applying Lemma 2.2 it follows that:

$$\begin{aligned}
F_{X|Y}(x|y) &= \int_{a_X}^{x} \frac{f_{X,Y}(t,y)}{f_Y(y)} dt \\
&= \int_{a_X}^{x} \frac{f_{\nu_X}(t) f_{\nu_Y}(y)}{f_Y(y)} \sum_{n_X=0}^{\infty} \sum_{n_Y=0}^{\infty} C_{n_X,n_Y} P_{n_X,n_Y}(t,y) dt \\
&= \frac{f_{\nu_Y}(y)}{f_Y(y)} \sum_{n_X=0}^{\infty} \sum_{n_Y=0}^{\infty} C_{n_X,n_Y} P_{Y,n_Y}(y) \int_{a_X}^{x} P_{X,n_X}(t) f_{\nu_X}(t) dt \\
&= \frac{\sigma_X(x) f_{\nu_X}(x)}{\sum_{n=0}^{\infty} C_{0,n} P_{Y,n}(y)} \sum_{n_X=0}^{\infty} \sum_{n_Y=0}^{\infty} \frac{C_{n_X,n_Y}}{\lambda_{X,n_X}} P'_{X,n_X}(x) P_{Y,n_Y}(y)
\end{aligned}$$

□

This result can also be generalised for N dimensions however for simplicity we will not present the generalised form in this paper. Instead we will consider how this expression as well as the corresponding sample moment based estimate can be useful in considering how to simulate a sample from our density estimate.

It should be noted that the sample moment-based estimate of the bivariate conditional cumulative distribution function (given a sample $\{(X_i, Y_i)\}_{i=1}^{M}$ from $f_{X,Y}$) is derived by again applying the same reasoning to the sample moment-based joint density estimate and is given by:

$$\hat{F}_{X;K_X|Y;K_Y}(x|y; \{(X_i,Y_i)\}_{i=1}^{M}) = \frac{\sigma_X(x) f_{\nu_X}(x)}{\sum_{n=0}^{K_Y} \hat{C}_{0,n} P_{Y,n}(y)} \sum_{n_X=0}^{K_X} \sum_{n_Y=0}^{K_Y} \frac{\hat{C}_{n_X,n_Y}}{\lambda_{X,n_X}} P'_{X,n_X}(x) P_{Y,n_Y}(y)$$

Where $\hat{C}_{n_X,n_Y}$ are the estimated coefficients of the bivariate moment-based density estimate $\hat{f}_{X,Y;K_X,K_Y}(x, y; \{(X_i, Y_i)\}_{i=1}^{M})$.

Simulating data from the joint density estimate is an important tool in both applying cross validation techniques in assessing the goodness of fit of density estimates as well in the synthetic data generation for privacy protection purposes which is one of the applications of the methods developed in this paper.

To explain this process, we use a bivariate case as an example. Consider random variables $X \in [a_X, b_X]$ and $Y \in [a_Y, b_Y]$ with bivariate density function $f_{X,Y}(x, y)$ which can be expressed in a polynomial form. Given a moment-based estimate $\hat{f}_{X,Y;K_X,K_Y}(x, y; \{(X_i, Y_i)\}_{i=1}^{M})$ of the density function $f_{X,Y}(x, y)$, we would like to produce a random sample $\{X_i^*, Y_i^*\}_{i=1}^{M}$ from the estimated joint density function $\hat{f}_{X,Y;K_X,K_Y}(x, y; \{(X_i, Y_i)\}_{i=1}^{M})$.

Firstly, from Lemma 6.1, and given $\hat{f}_{X,Y;K_X,K_Y}(x, y; \{(X_i, Y_i)\}_{i=1}^{M})$ we can obtain an estimate for the marginal cumulative distribution function $F_Y(y)$ given by:

$$\hat{F}_{Y;K_Y}(y;\{Y_i\}_{i=1}^M) = f_{\nu_Y}(y)\sigma_Y(y)\sum_{n=0}^{K_Y}\frac{\hat{C}_{0,n}}{\lambda_{Y,n}}P'_{Y,n}(y)$$

Where $\hat{f}_{X,Y;K_X,K_Y}(x,y;\{(X_i,Y_i)\}_{i=1}^M) = f_{\nu_X}(x)f_{\nu_Y}(y)\sum_{n_X=0}^{K_X}\sum_{n_Y=0}^{K_Y}\hat{C}_{n_X,n_Y}P_{X,n}(x)P_{Y,n}(y)$.
We can then simulate a sample of size M from a uniform distribution $\{U_{Y,i}\}_{i=1}^M$ and apply the inverse transform sampling method to obtain a sample $\{Y_i^*\}_{i=1}^M$ by calculating $Y_i^* = \hat{F}_{Y;K_Y}^{-1}(U_i;\{Y_i\}_{i=1}^M)$ for each $i = 1,\dots M$.
It follows that each Y_i^* has the density function $\hat{f}_{Y;K_Y}(y;\{Y_i\}_{i=1}^M) = f_{\nu_Y}(y)\sum_{n=0}^{K_Y}\hat{C}_{0,n}P_{Y,n}(y)$.
Furthermore from Lemma 6.2 we can also obtain an estimate for the conditional cumulative distribution function $F_{X|Y}(x|y)$ given by:

$$\hat{F}_{X;K_X|Y;K_Y}\left(x|y;\{(X_i,Y_i)\}_{i=1}^M\right) = \frac{\sigma_X(x)f_{\nu_X}(x)}{\sum_{n=0}^{K_Y}\hat{C}_{0,n}P_{Y,n}(y)}\sum_{n_X=0}^{K_X}\sum_{n_Y=0}^{K_Y}\frac{\hat{C}_{n_X,n_Y}}{\lambda_{X,n_X}}P'_{X,n_X}(x)P_{Y,n_Y}(y)$$

Applying a similar process as before, we simulate another uniform sample of size M denoted by $\{U_{X,i}\}_{i=1}^M$ and for each observation of this sample we calculate
$X_i^* = \hat{F}_{X;K_X|Y;K_Y}^{-1}\left(U_{X,i}|Y = Y_i^*;\{(X_i,Y_i)\}_{i=1}^M\right)$.
Hence we have applied the inverse transform sampling method to $\hat{F}_{Y;K_Y}^{-1}(U_i;\{Y_i\}_{i=1}^M)$ and obtained the sample $\{Y_i^*\}_{i=1}^M$. Then for each Y_i^*, we have applied the inverse transform sampling method to $\hat{F}_{X;K_X|Y;K_Y}\left(x|y;\{(X_i,Y_i)\}_{i=1}^M\right)$ to obtain a sample of size 1 X_i^*. Hence as this is performed for each Y_i^* we obtain a bivariate sample $\{X_i^*,Y_i^*\}_{i=1}^M$ which has density function $\hat{f}_{X,Y;K_X,K_Y}(x,y;\{(X_i,Y_i)\}_{i=1}^M)$.
In the following we apply the sampling process to the bivariate example studied in the previous section. We compare the basic statistical information given by the samples from the actual joint density function and the estimated joint density function.

Example 6.1. Consider again the truncated bivariate normal-gamma distribution from the previous example with density function:

$$f_{X,T}(x,t) \propto \frac{\beta^\alpha\sqrt{\lambda}}{\Gamma(\alpha)\sqrt{2\pi}}t^{\alpha-\frac{1}{2}}\exp\left(-\beta t - \frac{\lambda t(x-\mu)^2}{2}\right)$$

Where $\mu = 2$, $\lambda = \frac{1}{5}, \alpha = 6$ and $\beta = 3$ defined on $x \in [-2.862, 6.862]$ and $t \in (0, 5.053]$. Using the first density estimate in the simulation study depicted in Figure 4, we can obtain an estimate of the cumulative distribution function by applying the results obtained in Lemma 6.2 using a maximum order of polynomial of 6 for both variables. A plot of the estimated cumulative distribution function as well as the estimated marginal cumulative distribution functions is given in Figure 7.

Moreover, by applying the process previously outlined using the estimated cumulative distribution function, we can simulate a sample (of the same size $M = 1000$) from the estimated density function. Comparisons between the original sample and simulated sample are shown in Figures 8 and 9. As can be seen in these figures, we can obtain consistent results between the original and sampled set of observations by applying this process.

7. CONCLUDING REMARKS

The multivariate moments problem and it's application to the estimation of joint density functions are often considered highly impracticable in modern day analysis. Although many results exist within the framework of this issue, it is often an undesirable method to be used in real world applications due to the computational complexity and intractability of higher order moments. Applying such an approach to sample moment based estimation is often rarely considered due to the fact that it requires a tradeoff between the improved accuracy of higher order approximations and the increased sampling variance these higher max polynomial orders incur.

This paper addresses this issue for a generalised form of polynomial expansion making the strategy applicable to not just one type of polynomial basis, but to any hyper-geometric type polynomial basis (provided determinacy is assured) despite the examples in this paper being focussed on Legendre-type polynomials. Furthermore we also extend this result to estimate both cumulative and conditional cumulative distribution functions which also allows us to simulate data from our density estimates without requiring grid based sampling techniques. This property has considerable ramifications when attempting to produce representative samples of the original data and to impute missing information.

Nevertheless, there is still considerable research that needs to be explored pertaining to establishing adequate conditions to ensure determinacy on the unbounded domain. Furthermore, if we are to assume our sample space is compact, then we must also investigate the effect of altering the assumed bounds on the space in the approximation.

Regardless, by deriving these various analytic results, this paper obtains the specific statistical tools required to take the multivariate moments problem from a highly specified mostly analytical idea to a much more practical high dimensional density estimation method.

Acknowledgement: This research has been conducted with the support of the Australian Government Research Training Program Scholarship.

REFERENCES

Akhiezer, N. I., and Kemmer, N. (1965), *The classical moment problem: and some related questions in analysis*, Vol. 5 Oliver & Boyd Edinburgh.

Alexits, G. (1961), *Convergence problems of orthogonal series*, International series of monographs in pure and applied mathematics Pergamon Press.

Charlier, C. V. L. (1914), "Frequency curves of type A in heterograde statistics," *Ark. Mat. Ast. Fysik*, 9, 1–17.

Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press, 85–89.

Daniels, H. E. (1983), "Saddlepoint approximations in statistics," *Biometrika*, 70, 89–96.

Duong, T. (2018), *ks: Kernel Smoothing.* R package version 1.11.3.
URL: *https://CRAN.R-project.org/package=ks*

Edgeworth, F. Y. (1905), "The law of error," *Proc. Cambridge Philos. Soc.*, 20, 36–65.

Elderton, W. P., and Johnson, N. L. (1969), *Systems of Frequency Curves,* Cambridge University Press.

Gram, J. (1883), "On the development of real functions in rows by the method of least squares," *J. Reine Angew Math*, 94, 41–73.

Griffiths, R. C., and Spano, D. (2011), "Multivariate Jacobi and Laguerre polynomials, infinite-dimensional extensions, and their probabilistic connections with multivariate Hahn and Meixner polynomials," *Bernoulli Society*, 17, 1095–1125.

Kendall, M. G., Stuart, A., and Ord, J. K., eds (1987), *Kendall's Advanced Theory of Statistics*, New York, NY, USA: Oxford University Press, Inc.

Kibble, W. F. (1940), "A two variate gamma type distribution," *Sankhya: The Indian Journal of Statistics*, 5(2), 137–150.

Kleiber, C., and Stoyanov, J. (2013), "Multivariate distributions and the moment problem," *Journal of Multivariate Analysis*, 113, 7–18.

Krishnamoorthy, A. S., and Parthasarathy, M. (1951), "A multivariate gamma-type distribution," *The Annals of Mathematical Statistics*, 22(4), 549–557.
URL: *http://www.jstor.org/stable/2236923*

Lin, Y.-X. (2014), "Density approximant based on noise multiplied data," in *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2014, Ibiza, Spain, September 17-19, 2014. Proceedings*, ed. J. Domingo-Ferrer Springer International Publishing, 89–104.
URL: *http://dx.doi.org/10.1007/978-3-319-11257-2_8*

Macomber, J. H., and Myers, B. L. (1983), "The bivariate beta distribution: comparison of Monte Carlo generators and evaluation of parameter estimates," *ACM SIGSIM Simulation Digest*, 14(1-4), 26–35.

Mehler, F. G. (1866), "G.Reihenentwicklungen nach Laplaceschen functionen hoherer Ordnung," *J. Reine Angew. Math.*, 66, 161–176.

Munkhammar, J., Mattsson, L., and Rydn, J. (2017), "Polynomial probability distribution estimation using the method of moments," *PLOS ONE*, 12(4), 1–14.
URL: *https://doi.org/10.1371/journal.pone.0174573*

Mustapha, H., and Dimitrakopoulos, R. (2010), "Generalized Laguerre expansions of multivariate probability densities with moments," *Computers & Mathematics with Applications*, 60(7), 2178 – 2189.
URL: *http://www.sciencedirect.com/science/article/pii/S0898122110005717*

Petersen, L. (1982), "On the relation between the multidimensional moment problem and the one-dimensional moment problem," *Mathematica Scandinavica*, 51, 361–366.

Provost, S. B. (2005), "Moment based density approximants," *The Mathematica Journal*, 9, 728–756.

Putinar, M., and Schmüdgen, K. (2008), "Multivariate determinateness," *Indiana University Mathematics Journal*, 2931–2968.

Sánchez-Ruiz, J., and Dehesa, J. S. (1998), "Expansions in series of orthogonal hypergeometric polynomials," *Journal of computational and applied mathematics*, 89(1), 155–170.

Shohat, J. A., and Tamarkin, J. D. (1943), *The problem of moments*, Vol. 1 American Mathematical Soc.

Solomon, H., and Stephens, M. A. (1978), "Approximations to density functions using Pearson curves," *Journal of the American Statistical Association*, 73(361), 153–160.
URL: *http://www.jstor.org/stable/2286537*

Stieltjes, T.-J. (1894), Recherches sur les fractions continues, in *Annales de la Faculté des sciences de Toulouse: Mathématiques*, Vol. 8, 1–122.

Stoyanov, J., and Tolmatz, L. (2005), "Method for constructing Stieltjes classes for M-indeterminate probability distributions," *Applied Mathematics and Computation*, 165(3), 669–685.

Szegő, G. (1939), *Orthogonal Polynomials*, Vol. 23 of *American Mathematical Society* American Mathematical Society.
URL: *https://books.google.com.au/books?id=WRoPAAAAIAAJ*

Szökefalvi-Nagy, B. (1965), *Introduction to Real Functions and Orthogonal Expansions,* Akadémiai Kiadó.

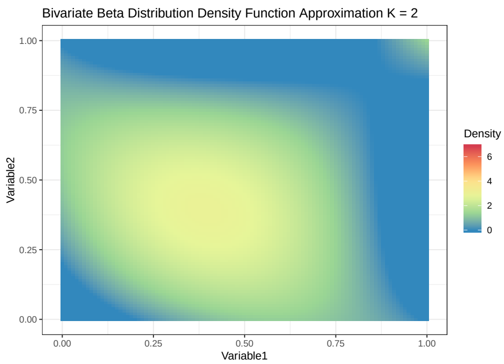
Welling, M. (1999), "Robust series expansions for probaility density estimation," *Technical Report, California Institute of Technology, Computational Vision Lab*, 1–22.

Withers, C. S., and Nadarajah, S. (2010), "Expansions for the multivariate normal," *Journal of Multivariate Analysis*, 101(5), 1311 – 1316.
URL: *http://www.sciencedirect.com/science/article/pii/S0047259X10000072*

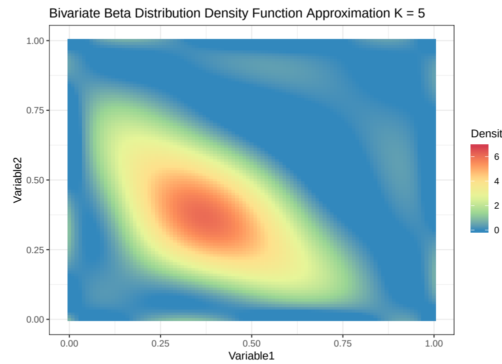
Xu, Y. (1997), "On orthogonal polynomials in several variables," *Special Functions, q-Series and Related Topics, The Fields Institute for Research in Mathematical Sciences, Communications Series*, 14, 247–270.

Xu, Y. (2014), *Orthogonal polynomials of several variables, Chapter 2 in Multivariable Special Functions*, 2.
URL: *https://arxiv.org/abs/1701.02709*

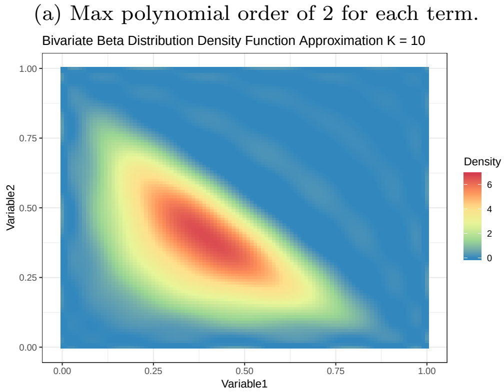
Yanez, R., Dehesa, J., and Nikiforov, A. (1994), "The three-term recurrence relation and the differentiation formulas for hypergeometric-type functions," *Journal of Mathematical Analysis and Applications*, 188(3), 855 – 866.
URL: *http://www.sciencedirect.com/science/article/pii/S0022247X84714673*



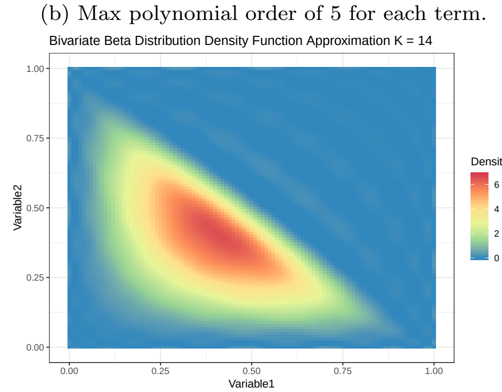
(a) Max polynomial order of 2 for each term.



(b) Max polynomial order of 5 for each term.



(c) Max polynomial order of 10 for each term.



(d) Max polynomial order of 14 for each term.

Figure 1: Plots of the moment-based bivariate beta density approximations for varying max polynomial orders.

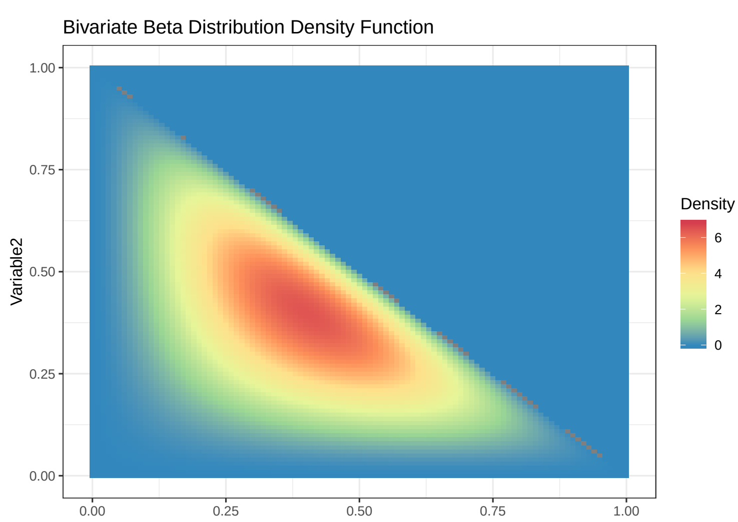


Figure 2: Plot of the true bivariate beta density function.

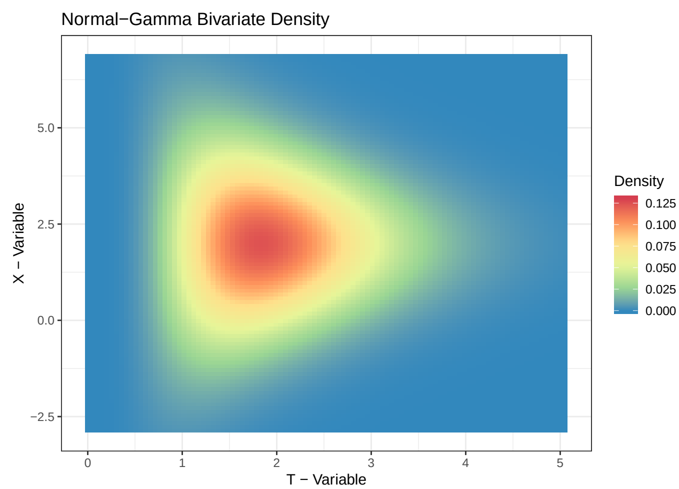
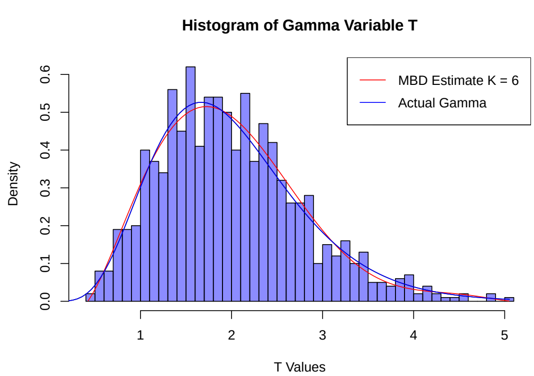
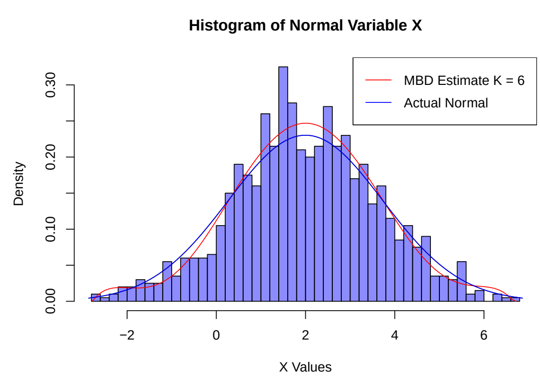


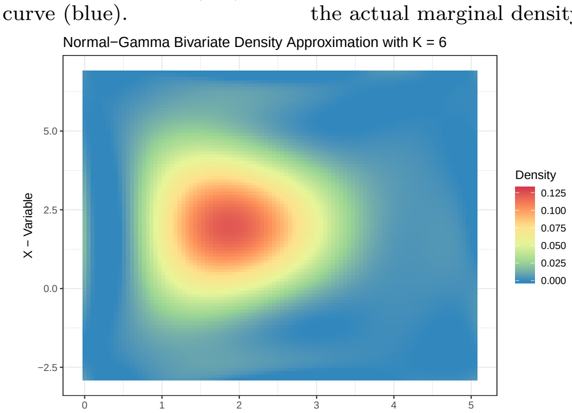
Figure 3: Plot of the true bivariate normal-gamma density function.



(a) Marginal histogram of the gamma random variable with both the estimated marginal density curve (red) and the actual marginal density curve (blue).

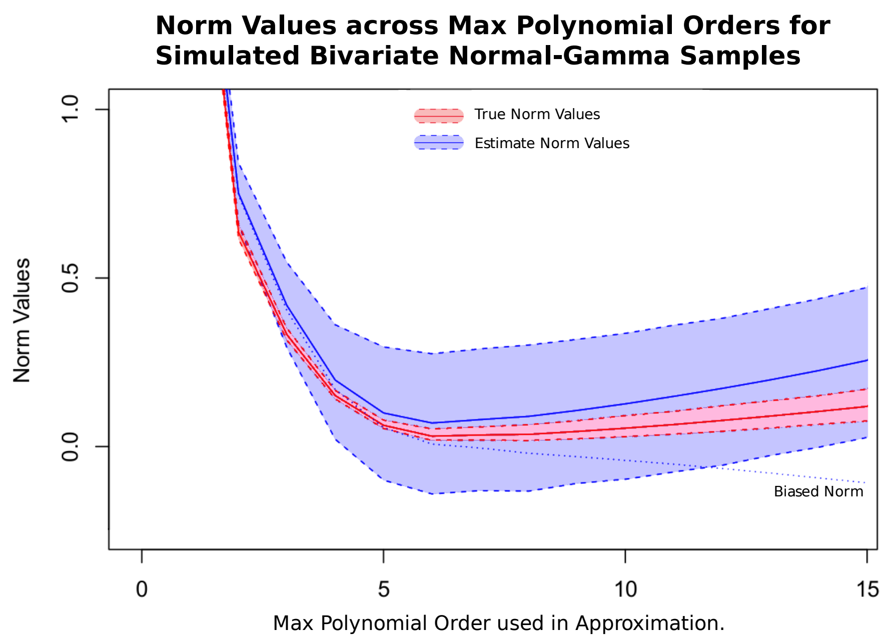


(b) Marginal histogram of the normal random variable with both the estimated marginal density curve (red) and the actual marginal density curve (blue).

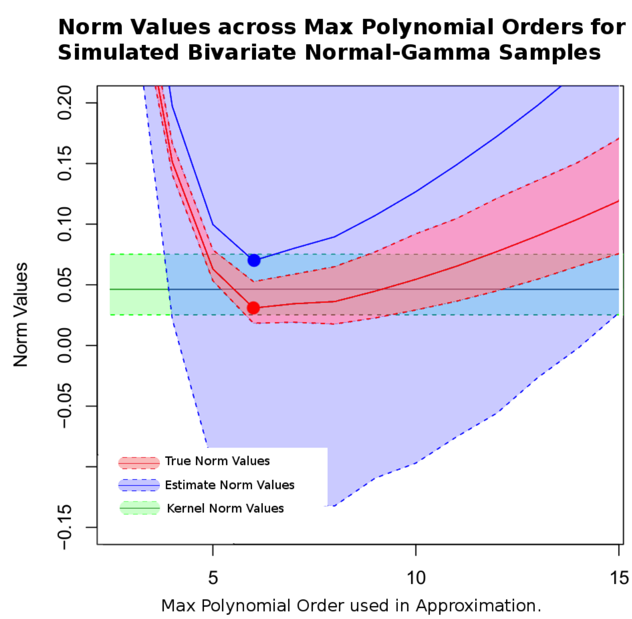


(c) Plot of the estimated bivariate normal-gamma density function.

Figure 4: Plots of the estimated bivariate normal-gamma density function based on a sample of size 1000 and using a maximum order of polynomial of 6 for each variable.



(a) Plot of $\mathcal{N}_{K,K}$ (red) and $\hat{\mathcal{N}}^*_{K,K} + 3.1379$ (blue) with 95% confidence intervals over varying levels of max polynomial order K.



(b) Plot of $\mathcal{N}_{K,K}$ (red) and $\hat{\mathcal{N}}^*_{K,K} + 3.1379$ (blue) with 95% confidence intervals over varying levels of max polynomial order K compared to the estimated kernel density norm (green) over varying max polynomial orders.

Figure 5: $\mathcal{N}_{\mathbf{K}}$-Norm simulation study for the normal-gamma bivariate density.

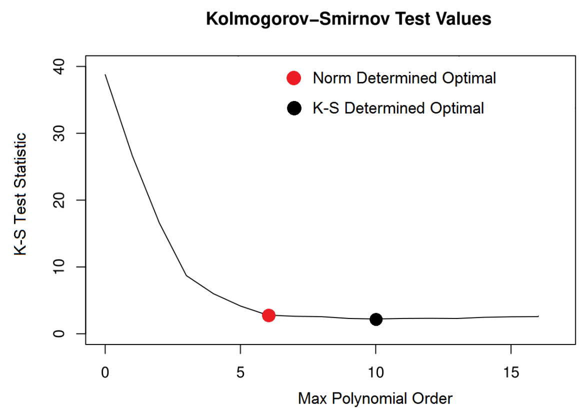
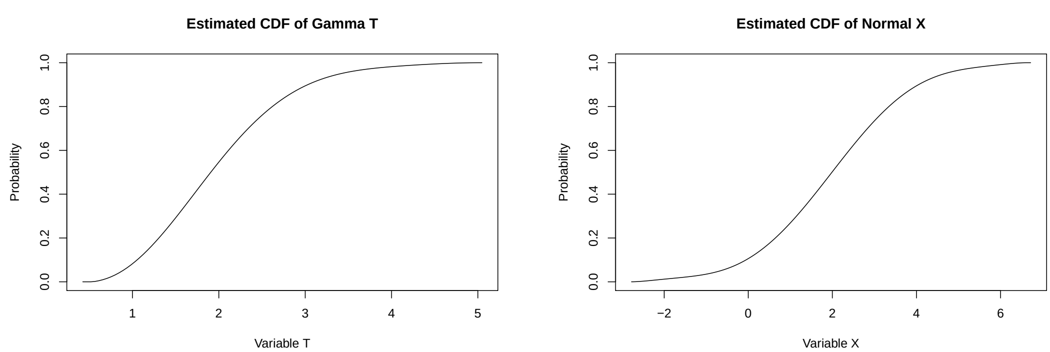
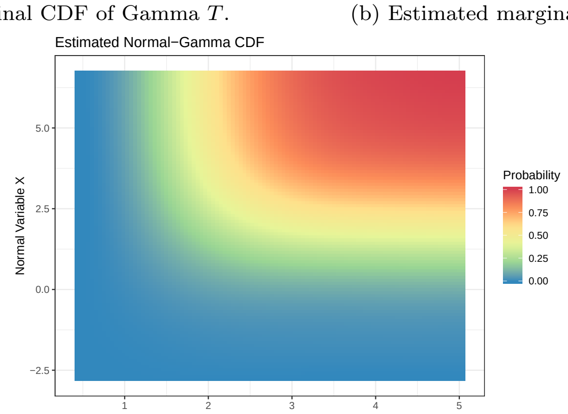


Figure 6: Plot of the Kolmogorov-Smirnov test statistic values for varying max polynomial orders given a sample from the bivariate normal-gamma distribution with the K-S values marked at each max polynomial order which gives the minimum value of the K-S (black) and the $\mathcal{N}_{\mathbf{K}}$-Norm (red) values.



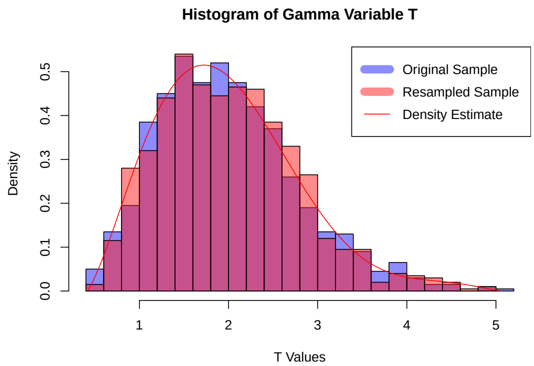
(a) Estimated marginal CDF of Gamma T.

(b) Estimated marginal CDF of Normal X.

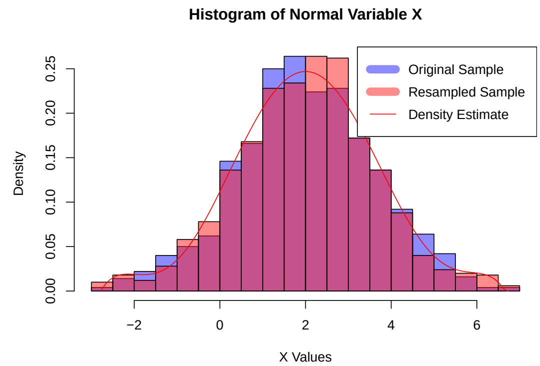


(c) Estimated Bivariate CDF of the Normal-Gamma Distribution.

Figure 7: Estimated moment-based cumulative distribution function of the truncated bivariate normal-gamma distribution where the maximum order of the polynomial approximation used was 6 for each variable.



(a) Histogram of the original and sampled Gamma T observations.



(b) Histogram of the original and sampled Normal X observations.

Figure 8: Histograms of the original and sampled observations using a max order of 6 for each variable.

	Gamma T		**Normal X**	
Statistics	Original	Sampled	Original	Sampled
Min	0.424	0.507	-2.778	-2.767
1st Quartile	1.427	1.438	0.992	0.979
Median	1.903	1.926	1.965	2.059
Mean	1.995	2.012	2.006	2.006
3rd Quartile	2.454	2.492	3.084	3.010
Max	5.047	4.880	6.716	6.652
Std Dev	0.783	0.777	1.616	1.616

	Bivariate	
Statistics	Original	Sampled
Correlation	-0.0085	-0.0050
Intercept $\hat{\beta}_0$	2.0032	2.0166
Std Error $\hat{\beta}_0$	0.0395	0.0392
Slope $\hat{\beta}_1$	-0.0042	-0.0024
Std Error $\hat{\beta}_1$	0.0153	0.0152

Note: Regression performed fits the model $T = \beta_0 + \beta_1 X + \epsilon$.

Figure 9: Table of summary statistics comparing the properties of the original observations and the sampled observations based on the moment-based density approximation.

Supplementary Materials

APPENDIX A.

Proof of Lemma 2.1

Firstly we note that the expression for the leading term κ_n of the nth polynomial was given in Sánchez-Ruiz and Dehesa (1998) and as such a proof of this expression is not given here. Applying the hyper-geometric type differential equation where $P_n(x) = \sum_{i=0}^{n} a_{n,i}x^i$ we have:

$$\sigma(x)P_n''(x)f_\nu(x) + \tau(x)P_n'(x)f_\nu(x) - \lambda_n P_n(x)f_\nu(x) = 0$$

$$\sigma(x)\left(\sum_{i=0}^{n} i(i-1)a_{n,i}x^{i-2}\right) + \tau(x)\left(\sum_{i=0}^{n} ia_{n,i}x^{i-1}\right) - \lambda_n\left(\sum_{i=0}^{n} a_{n,i}x^i\right) = 0$$

$$\left(\frac{\sigma''{}'}{2}x^2 + \sigma'(0)x + \sigma(0)\right)\left(\sum_{i=0}^{n} i(i-1)a_{n,i}x^{i-2}\right) +$$

$$(\tau' x + \tau(0))\left(\sum_{i=0}^{n} ia_{n,i}x^{i-1}\right) - \lambda_n\left(\sum_{i=0}^{n} a_{n,i}x^i\right) = 0$$

$$\sum_{i=0}^{n} i(i-1)a_{n,i}\sigma(0)x^{i-2} + \sum_{i=0}^{n}(i(i-1)\sigma'(0) + i\tau(0))a_{n,i}x^{i-1} +$$

$$\sum_{i=0}^{n}(i(i-1)\frac{\sigma''}{2} + i\tau' - \lambda_n)a_{n,i}x^i = 0$$

Therefore equating coefficients for each $i = 0, 1, \ldots, n-2$ we have:

$$(i+2)(i+1)\sigma(0)a_{n,i+2} + \Big(i(i+1)\sigma'(0) +$$

$$(i+1)\tau(0)\Big)a_{n,i+1} + \left(-\lambda_n + i(i-1)\frac{\sigma''}{2} + i\tau'\right)a_{n,i} = 0$$

Also we have:

$$\left(n(n-1)\sigma'(0) + n\tau(0)\right)\kappa_n + \left(-\lambda_n + (n-1)(n-2)\frac{\sigma''}{2} + (n-1)\tau'\right)a_{n,n-1} = 0$$

Therefore we obtain the following expressions

$$a_{n,i} = \frac{a_{n,i+1}(i+1)(i\sigma'(0) + \tau(0)) + a_{n,i+2}(i+2)(i+1)\sigma(0)}{\lambda_n - i(i-1)\frac{\sigma''}{2} - i\tau'}$$

$$a_{n,n-1} = \kappa_n\frac{n(n-1)\sigma'(0) + n\tau(0)}{\lambda_n - (n-1)(n-2)\frac{\sigma''}{2} - (n-1)\tau'}$$

Hence substituting $\lambda_n = n(n-1)\frac{\sigma''}{2} + n\tau'$ we obtain the required equations.

APPENDIX B.

Proof of Theorem 4.1

Firstly we note the following things:

- As f_ν is a continuous function on the compact space $\mathbf{\Omega}$ there exists a constant M_ν such that $\underset{\mathbf{x}\in\mathbf{\Omega}}{sup}|f_\nu(\mathbf{x})| = M_\nu$.
- As $P_{n_1,\ldots,n_N}$ is a continuous function on the compact space $\mathbf{\Omega}$ there exists a constant $M_{P_{n_1,\ldots,n_N}}$ for each $(n_1,\ldots,n_N) \leq (K_1,\ldots,K_N)$ element-wise such that $\underset{\mathbf{x}\in\mathbf{\Omega}}{sup}|P_{n_1,\ldots,n_N}(\mathbf{x})| = M_{P_{n_1,\ldots,n_N}}$.
  Furthermore let $M_P = \max\{M_{P_{n_1,\ldots,n_N}}$ for $(n_1,\ldots,n_N) \leq (K_1,\ldots,K_N)$ element-wise.$\}$
- From the strong law of large numbers the estimated coefficients $\hat{C}_{n_1,\ldots,n_N}$ converges almost surely as $M \to \infty$ to $C_{n_1,\ldots,n_N}$.

  Therefore it follows that for any $\epsilon > 0$:

$$P\left(\lim_{M\to\infty}\left|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\right| > \epsilon\right) = 0 \quad \text{for each } (n_1,\ldots,n_N) \in \mathbb{N}_0^N$$

- We denote the absolute sum of coefficients by $S_\mathbf{K}$. Hence

$$S_\mathbf{K} = \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}|C_{n_1,\ldots,n_N}| < \infty$$

  Furthermore $S_\mathbf{K} \geq 1$ as $C_{0,\ldots,0} = 1$.
- The number of coefficients is given by $||\mathbf{K}|| = \prod_{i=1}^N(K_i + 1) < \infty$

For a given $\epsilon_p > 0$ let $\epsilon_0 < \frac{\epsilon_p}{6}$ be a real number.

Functions $P_{n_1,\ldots,n_N}(\mathbf{x})$ for each $(n_1,\ldots,n_N) \leq (K_1,\ldots,K_N)$ element-wise and $f_{\mathbf{X};\mathbf{K}}(\mathbf{x})$ are continuous on the compact space $\mathbf{\Omega}$ therefore by the properties of continuous functions:

1. For the fixed $\frac{\epsilon_0}{M_\nu S_\mathbf{K}}$ there exists a $\delta_{n_1,\ldots,n_N} > 0$ for each $(n_1,\ldots,n_N) \leq (K_1,\ldots,K_N)$ such that for all $||\mathbf{x}-\mathbf{x}^*|| < \delta_{n_1,\ldots,n_N}$ then $|P_{n_1,\ldots,n_N}(\mathbf{x}) - P_{n_1,\ldots,n_N}(\mathbf{x}^*)| < \frac{\epsilon_0}{M_\nu S_\mathbf{K}}$ for each $(n_1,\ldots,n_N) \leq (K_1,\ldots,K_N)$ element-wise.
2. For the fixed $\frac{\epsilon_0}{M_\nu S_\mathbf{K}}$ there exists a $\delta_0 > 0$ such that for all $||\mathbf{x}-\mathbf{x}^*|| < \delta_0$ then $|f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}^*)| < \frac{\epsilon_0}{M_\nu S_\mathbf{K}}$.

Hence let $\delta = \min\{\delta_0, \delta_{0,\ldots,0}, \ldots, \delta_{K_1,\ldots,K_N}\}$.

There are a finite sequence of δ-neighbourhoods denoted $\{O_{\mathbf{x}_j^*}\}_{j=1}^{T(\delta)}$ such that:

1. $O_{\mathbf{x}_j^*} = \{\mathbf{x} \in \mathbf{\Omega} : ||\mathbf{x} - \mathbf{x}_j^*|| < \delta\}$.
2. $\cup_{j=1}^{T(\delta)} O_{\mathbf{x}_j^*}$ covers $\mathbf{\Omega}$ where $O_{\mathbf{x}_j^*}$ is centered on $\mathbf{x}_j^*$.

Therefore for all $\mathbf{x} \in O_{\mathbf{x}_j^*} \subseteq \mathbf{\Omega}$ for each $j = 1, \ldots, T(\delta)$ we have:

- $|P_{n_1,\ldots,n_N}(\mathbf{x}) - P_{n_1,\ldots,n_N}(\mathbf{x}^*)| < \frac{\epsilon_0}{M_\nu S_{\mathbf{K}}}$.
- $|f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}^*)| < \frac{\epsilon_0}{M_\nu S_{\mathbf{K}}}$.

Consider now for any $\mathbf{x} \in \mathbf{\Omega}$.

$$
\begin{aligned}
P\Bigg(\lim_{M\to\infty} \sup_{\mathbf{x}\in\mathbf{\Omega}} \Big|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x})\Big| > \epsilon_p\Bigg) \\
\leq \quad & \sum_{j=1}^{T(\delta)} P\Bigg(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \Big|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M) - \hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^M) + \hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^M) \\
& \qquad - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*) + f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x})\Big| > \epsilon_p\Bigg) \\
\leq \quad & \sum_{j=1}^{T(\delta)} P\left(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \Big|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^M) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^M)\Big| > \frac{\epsilon_p}{3}\right) \\
& + \sum_{j=1}^{T(\delta)} P\left(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \Big|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^M) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*)\Big| > \frac{\epsilon_p}{3}\right) \\
& + \sum_{j=1}^{T(\delta)} P\left(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \big|f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*)\big| > \frac{\epsilon_p}{3}\right)
\end{aligned}
$$

Considering each term for any $j = 1, \ldots, T(\delta)$:

$$
\begin{aligned}
& P\left(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \Big|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^M) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*)\Big| > \frac{\epsilon_p}{3}\right) \\
& \leq P\left(\lim_{M\to\infty} \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} |f_\nu(\mathbf{x})| \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} \Big|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\Big| \sup_{\mathbf{x}\in O_{\mathbf{x}_j^*}} \big|P_{n_1,\ldots,n_N}(\mathbf{x}_j^*)\big| > \frac{\epsilon_p}{3}\right) \\
& = P\left(\sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} \lim_{M\to\infty} \Big|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\Big| > \frac{\epsilon_p}{3M_\nu M_P}\right) \\
& \leq \sum_{n_1=0}^{K_1} \cdots \sum_{n_N=0}^{K_N} P\left(\lim_{M\to\infty} \Big|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\Big| > \frac{\epsilon_p}{3M_\nu M_P||\mathbf{K}||}\right) \\
& = 0 \qquad \text{As } \hat{C}_{n_1,\ldots,n_N} \text{ converges almost surely to } C_{n_1,\ldots,n_N}
\end{aligned}
$$

$$P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in O_{\mathbf{x}_j^*}}{sup} \left|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*;\{\mathbf{X}_i\}_{i=1}^{M})\right| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in O_{\mathbf{x}_j^*}}{sup} |f_\nu(\mathbf{x})| \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \left|\hat{C}_{n_1,\ldots,n_N}\right| \left|P_{n_1,\ldots,n_N}(\mathbf{x}) - P_{n_1,\ldots,n_N}(\mathbf{x}_j^*)\right| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\lim_{M\to\infty} \frac{\epsilon_0 M_\nu}{M_\nu S_{\mathbf{K}}} \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \left|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N} + C_{n_1,\ldots,n_N}\right| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\lim_{M\to\infty} \frac{\epsilon_0}{S_{\mathbf{K}}} \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \left|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\right| + \frac{\epsilon_0}{S_{\mathbf{K}}} \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} |C_{n_1,\ldots,n_N}| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} \left[\lim_{M\to\infty} \left|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\right|\right] > \frac{\epsilon_p S_{\mathbf{K}}}{6\epsilon_0}\right) + P\left(\epsilon_0 > \frac{\epsilon_p}{6}\right)$$
$$\leq \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} P\left(\lim_{M\to\infty} \left|\hat{C}_{n_1,\ldots,n_N} - C_{n_1,\ldots,n_N}\right| > \frac{\epsilon_p S_{\mathbf{K}}}{6\epsilon_0 ||\mathbf{K}||}\right) + 0$$
$$= 0 \qquad \text{As } \hat{C}_{n_1,\ldots,n_N} \text{ converges almost surely to } C_{n_1,\ldots,n_N}$$

$$P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in O_{\mathbf{x}_j^*}}{sup} \left|f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x}_j^*)\right| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in O_{\mathbf{x}_j^*}}{sup} |f_\nu(\mathbf{x})| \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} |C_{n_1,\ldots,n_N}| \left|P_{n_1,\ldots,n_N}(\mathbf{x}) - P_{n_1,\ldots,n_N}(\mathbf{x}_j^*)\right| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\lim_{M\to\infty} M_\nu \frac{\epsilon_0}{M_\nu S_{\mathbf{K}}} \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} |C_{n_1,\ldots,n_N}| > \frac{\epsilon_p}{3}\right)$$
$$\leq P\left(\epsilon_0 > \frac{\epsilon_p}{6}\right)$$
$$= 0$$

Hence as

$$0 \leq P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in\mathbf{\Omega}}{sup} \left|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x})\right| > \epsilon_p\right) \leq \sum_{j=1}^{T(\delta)} 0 + \sum_{j=1}^{T(\delta)} 0 + \sum_{j=1}^{T(\delta)} 0 = 0$$

Then it follows that:

$$P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in\mathbf{\Omega}}{sup} \left|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x})\right| > \epsilon_p\right) = 0$$

And hence for any $\epsilon > 0$ then it follows that

$$P\left(\lim_{M\to\infty} \underset{\mathbf{x}\in\mathbf{\Omega}}{sup} \left|\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M}) - f_{\mathbf{X};\mathbf{K}}(\mathbf{x})\right| < \epsilon\right) = 1$$

Therefore:

$$\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M}) \xrightarrow{a.s.} f_{\mathbf{X};\mathbf{K}}(\mathbf{x}) \text{ as } M \to \infty \text{ uniformly.}$$

APPENDIX C.

Proof of Theorem 4.2

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] \\
&\quad= \mathbb{E}_{\mathbf{X}}\left[\int_\Omega \left|\frac{\hat{f}_{\mathbf{X};\mathbf{K}}(\mathbf{x};\{\mathbf{X}_i\}_{i=1}^{M})}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})}\right|^2 d\mu_\nu\right] \\
&\quad= \mathbb{E}_{\mathbf{X}}\left[\int_\Omega \left|\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\hat{C}_{n_1,\ldots,n_N}P_{n_1,\ldots,n_N}(\mathbf{x}) - \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}C_{n_1,\ldots,n_N}P_{n_1,\ldots,n_N}(\mathbf{x})\right|^2 d\mu_\nu\right] \\
&\quad= \mathbb{E}_{\mathbf{X}}\Bigg\{\int_\Omega \Bigg[\sum_{n_1=0}^{K_1}\sum_{m_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\sum_{m_N=0}^{K_N}\hat{C}_{n_1,\ldots,n_N}\hat{C}_{m_1,\ldots,m_N}P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x}) \\
&\qquad\qquad - 2\sum_{n_1=0}^{K_1}\sum_{m_1=0}^{\infty}\cdots\sum_{n_N=0}^{K_N}\sum_{m_N=0}^{\infty}\hat{C}_{n_1,\ldots,n_N}C_{m_1,\ldots,m_N}P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x}) \\
&\qquad\qquad + \sum_{n_1=0}^{\infty}\sum_{m_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\sum_{m_N=0}^{\infty}C_{n_1,\ldots,n_N}C_{m_1,\ldots,m_N}P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})\Bigg] d\mu_\nu\Bigg\} \\
&\quad= \mathbb{E}_{\mathbf{X}}\Bigg\{\sum_{n_1=0}^{K_1}\sum_{m_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\sum_{m_N=0}^{K_N}\hat{C}_{n_1,\ldots,n_N}\hat{C}_{m_1,\ldots,m_N}\int_\Omega P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_\nu \\
&\qquad\qquad - 2\sum_{n_1=0}^{K_1}\sum_{m_1=0}^{\infty}\cdots\sum_{n_N=0}^{K_N}\sum_{m_N=0}^{\infty}\hat{C}_{n_1,\ldots,n_N}C_{m_1,\ldots,m_N}\int_\Omega P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_\nu \\
&\qquad\qquad + \sum_{n_1=0}^{\infty}\sum_{m_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}\sum_{m_N=0}^{\infty}C_{n_1,\ldots,n_N}C_{m_1,\ldots,m_N}\int_\Omega P_{n_1,\ldots,n_N}(\mathbf{x})P_{m_1,\ldots,m_N}(\mathbf{x})d\mu_\nu\Bigg\} \\
&\quad= \mathbb{E}_{\mathbf{X}}\Bigg\{\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\hat{C}^2_{n_1,\ldots,n_N} - 2\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\hat{C}_{n_1,\ldots,n_N}C_{n_1,\ldots,n_N} + \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}C^2_{n_1,\ldots,n_N}\Bigg\} \\
&\quad= \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\mathbb{E}_{\mathbf{X}}\left[\hat{C}^2_{n_1,\ldots,n_N}\right] - 2\sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}C_{n_1,\ldots,n_N}\mathbb{E}_{\mathbf{X}}\left[\hat{C}_{n_1,\ldots,n_N}\right] + \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}C^2_{n_1,\ldots,n_N} \\
&\quad= \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\mathbb{E}_{\mathbf{X}}\left[\hat{C}^2_{n_1,\ldots,n_N}\right] - \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}C^2_{n_1,\ldots,n_N} \\
&\qquad\qquad + \sum_{n_1=0}^{\infty}\cdots\sum_{n_N=0}^{\infty}C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}C^2_{n_1,\ldots,n_N} \\
&\quad= \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N}\left\{\mathbb{E}_{\mathbf{X}}\left[\hat{C}^2_{n_1,\ldots,n_N}\right] - C^2_{n_1,\ldots,n_N}\right\} + \mathcal{T}_{\mathbf{K}} \\
&\quad= \sum_{n_1=0}^{K_1}\cdots\sum_{n_N=0}^{K_N} Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) + \mathcal{T}_{\mathbf{K}}
\end{aligned}
$$

APPENDIX D.

In order to prove Theorem 4.4 we also need to prove the following result:

Lemma D.1. *For a dimension* $j \in \{1, \ldots, N\}$ *as well as for any* $\epsilon > 0$ *and* $m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N \in \mathbb{N}_0$ *there exists* K_j *such that for all* $n_j > K_j$*:*

$$\sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}} \sum_{n_{j+1}=0}^{m_{j+1}} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} < \epsilon$$

Proof. We will only provide the detailed proof for 3 dimensions ($N = 3$) and note that the N dimension proof can be achieved by following the same process outlined for the 3 dimension proof.

From Theorem 3.2, for any $\epsilon > 0$ there exists $\mathbf{K}_0 = (K_{0,1}, \ldots, K_{0,N})$ such that:

$$\int_\Omega \left| \frac{f_{\mathbf{X}}(\mathbf{x})}{f_\nu(\mathbf{x})} - \frac{f_{\mathbf{X};\mathbf{K}_0}(\mathbf{x})}{f_\nu(\mathbf{x})} \right|^2 d\mu_\nu = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{K_{0,1}} \cdots \sum_{n_N=0}^{K_{0,N}} C^2_{n_1,\ldots,n_N} = \mathcal{T}_{\mathbf{K}_0} < \epsilon$$

Considering for $N = 3$ and for any $\epsilon_0 > 0$ we can fix $\epsilon = \frac{\epsilon_0}{2}$ and note that there therefore exists a $\mathbf{K}_0 = (K_{0,1}, K_{0,2}, K_{0,3})$ such that $\mathcal{T}_{\mathbf{K}_0} < \frac{\epsilon_0}{2}$ and hence:

$$\begin{aligned}
\mathcal{T}_{\mathbf{K}_0} &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=0}^{\infty} C^2_{n_1,n_2,n_3} - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} \\
&= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \left(\sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \right) - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} \\
&= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} \\
&= \sum_{n_1=0}^{\infty} \sum_{n_3=0}^{K_{0,3}} \left(\sum_{n_2=0}^{K_{0,2}} C^2_{n_1,n_2,n_3} + \sum_{n_2=K_{0,2}+1}^{\infty} C^2_{n_1,n_2,n_3} \right) + \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \\
&\qquad - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} \\
&= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \\
&- \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}
\end{aligned}$$

$$\mathcal{T}_{\mathbf{K}_0} = \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} \left(\sum_{n_1=0}^{K_{0,1}} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{\infty} C^2_{n_1,n_2,n_3} \right) + \sum_{n_1=0}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$+ \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$= \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$+ \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} - \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$= \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} < \frac{\epsilon_0}{2}$$

Therefore as every term is positive, we note that this implies that:

$$T_0 = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} < \frac{\epsilon_0}{2}$$

Furthermore, following the same process if we were to continue to partition all the infinite sums we would have:

$$\mathcal{T}_{\mathbf{K}_0} = \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$+ \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$
$$+ \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} < \frac{\epsilon_0}{2}$$

Labeling each summation term in the above result we have:

$$T_1 = \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$T_2 = \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$
$$T_3 = \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=0}^{K_{0,3}} C^2_{n_1,n_2,n_3}$$

$$T_4 = \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$

$$T_5 = \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$

$$T_6 = \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$

$$T_7 = \sum_{n_1=K_{0,1}+1}^{\infty} \sum_{n_2=K_{0,2}+1}^{\infty} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$

And hence $\sum_{i=1}^{7} T_i < \frac{\epsilon_0}{2}$.
Now consider for any m_1, m_2 the sum:

$$\sum_{n_1=0}^{m_1} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}$$

We have four distinct cases of this sum:
Case 1 $m_1 \leq K_{0,1}$ and $m_2 \leq K_{0,2}$
When $m_1 \leq K_{0,1}$ and $m_2 \leq K_{0,2}$ then it follows that:

$$\sum_{n_1=0}^{m_1} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \leq T_4 < \frac{\epsilon_0}{2} < \epsilon_0$$

Case 2 $m_1 \leq K_{0,1}$ and $m_2 > K_{0,2}$
When $m_1 \leq K_{0,1}$ and $m_2 > K_{0,2}$ then it follows that:

$$\begin{aligned}
\sum_{n_1=0}^{m_1} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} &= \sum_{n_1=0}^{m_1} \sum_{n_3=K_{0,3}+1}^{\infty} \left(\sum_{n_2=0}^{K_{0,2}} C^2_{n_1,n_2,n_3} + \sum_{n_2=K_{0,2}+1}^{m_2} C^2_{n_1,n_2,n_3} \right) \\
&= \sum_{n_1=0}^{m_1} \sum_{n_2=0}^{K_{0,2}} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{m_1} \sum_{n_2=K_{0,2}+1}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \\
&\leq T_4 + T_6 \\
&< \frac{\epsilon_0}{2} \\
&< \epsilon_0
\end{aligned}$$

Case 3 $m_1 > K_{0,1}$ and $m_2 \leq K_{0,2}$
When $m_1 > K_{0,1}$ and $m_2 \leq K_{0,2}$ then it follows that:

$$\begin{aligned}
\sum_{n_1=0}^{m_1} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} &= \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} \left(\sum_{n_1=0}^{K_{0,1}} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{m_1} C^2_{n_1,n_2,n_3} \right) \\
&= \sum_{n_1=0}^{K_{0,1}} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{m_1} \sum_{n_2=0}^{m_2} \sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} \\
&\leq T_4 + T_5 \\
&< \frac{\epsilon_0}{2} \\
&< \epsilon_0
\end{aligned}$$

Case 4 $m_1 > K_{0,1}$ and $m_2 > K_{0,2}$
When $m_1 > K_{0,1}$ and $m_2 > K_{0,2}$ then it follows that:

$$\begin{aligned}
\sum_{n_1=0}^{m_1}\sum_{n_2=0}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} &= \sum_{n_2=0}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty}\left(\sum_{n_1=0}^{K_{0,1}} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{m_1} C^2_{n_1,n_2,n_3}\right)\\
&= \sum_{n_1=0}^{K_{0,1}}\sum_{n_2=0}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{m_1}\sum_{n_2=0}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}\\
&= \sum_{n_1=0}^{K_{0,1}}\sum_{n_3=K_{0,3}+1}^{\infty}\left(\sum_{n_2=0}^{K_{0,2}} C^2_{n_1,n_2,n_3} + \sum_{n_2=K_{0,2}+1}^{m_2} C^2_{n_1,n_2,n_3}\right)\\
&\quad + \sum_{n_1=K_{0,1}+1}^{m_1}\sum_{n_3=K_{0,3}+1}^{\infty}\left(\sum_{n_2=0}^{K_{0,2}} C^2_{n_1,n_2,n_3} + \sum_{n_2=K_{0,2}+1}^{m_2} C^2_{n_1,n_2,n_3}\right)\\
&= \sum_{n_1=0}^{K_{0,1}}\sum_{n_2=0}^{K_{0,2}}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=0}^{K_{0,1}}\sum_{n_2=K_{0,2}+1}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}\\
&\quad + \sum_{n_1=K_{0,1}+1}^{m_1}\sum_{n_2=0}^{K_{0,2}}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} + \sum_{n_1=K_{0,1}+1}^{m_1}\sum_{n_2=K_{0,2}+1}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3}\\
&\leq T_4 + T_6 + T_5 + T_0\\
&< \frac{\epsilon_0}{2} + \frac{\epsilon_0}{2}\\
&= \epsilon_0
\end{aligned}$$

Therefore it follows for any m_1, m_2 we have

$$\sum_{n_1=0}^{m_1}\sum_{n_2=0}^{m_2}\sum_{n_3=K_{0,3}+1}^{\infty} C^2_{n_1,n_2,n_3} < \epsilon_0$$

Similarly, repeating this process we have:

For any m_1, m_3 $$\sum_{n_1=0}^{m_1}\sum_{n_2=K_{0,2}+1}^{\infty}\sum_{n_3=0}^{m_3} C^2_{n_1,n_2,n_3} < \epsilon_0$$

For any m_2, m_3 $$\sum_{n_1=K_{0,1}+1}^{\infty}\sum_{n_2=0}^{m_2}\sum_{n_3=0}^{m_3} C^2_{n_1,n_2,n_3} < \epsilon_0$$

□

This now leads us to the proof of Theorem 4.4

Proof of Theorem 4.4

Firstly for a fixed $j \in \{1, \ldots, N\}$ we define the notation for some vector $\mathbf{a} = (a_1, \ldots, a_N)$ that $\mathbf{a}_{(j)} = (a_1, \ldots, a_j + 1, \ldots, a_N)$.

We assume there exists $\epsilon_0 > 0$ such that $Var_{\mathbf{X}}(\hat{C}_{n_1,\ldots,n_N}) \geq \epsilon_0$ for all $n_1, \ldots, n_N \in \mathbb{N}_0$ and fix ϵ_0.

From Lemma D.1:

For a dimension $j \in \{1, \ldots, N\}$ as well as for any $\epsilon > 0$ and $m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N \in \mathbb{N}_0$ there exists K_j such that for all $n_j > K_j$:

$$\sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}} \sum_{n_{j+1}=0}^{m_{j+1}} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} < \epsilon$$

Therefore for $\epsilon = \epsilon_0$ and any $m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N$ there exists K_j such that for all $m_j > K_j$:

$$\begin{aligned}
\mathcal{T}_{\mathbf{m}} - \mathcal{T}_{\mathbf{m}_{(j)}} &= \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{m_1} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} C^2_{n_1,\ldots,n_N} \\
&\quad + \sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}} \sum_{n_j=0}^{m_j+1} \sum_{n_{j+1}=0}^{m_{j+1}} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} \\
&= \sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}} \sum_{n_j=0}^{m_j+1} \sum_{n_{j+1}=0}^{m_{j+1}} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} - \sum_{n_1=0}^{m_1} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,n_N} \\
&= \sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}} \sum_{n_{j+1}=0}^{m_{j+1}} \cdots \sum_{n_N=0}^{m_N} C^2_{n_1,\ldots,m_j+1,\ldots,n_N} \\
&< \epsilon_0
\end{aligned}$$

Therefore suppose that for any $m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_N$ and any m_j where $m_j > K_j$ and $\mathbf{m} = (m_1, \ldots, m_N)$:

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}_{(j)}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] \leq \mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right]$$

We note that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}_{(j)}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] \leq \mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}}\left(\{\mathbf{X}_i\}_{i=1}^M\right)\right] \\
&\Rightarrow \sum_{k_1=0}^{m_1} \cdots \sum_{k_N=0}^{m_N} Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_N}\right) + \sum_{k_1=0}^{m_1} \cdots \sum_{k_{j-1}=0}^{m_{j-1}} \sum_{k_{j+1}=0}^{m_{j+1}} \cdots \sum_{k_N=0}^{m_N} Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_{j-1},m_j+1,k_{j+1},\ldots,k_N}\right) \\
&\qquad + \mathcal{T}_{\mathbf{m}_{(j)}} - \sum_{k_1=0}^{m_1} \cdots \sum_{k_N=0}^{m_N} Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_N}\right) - \mathcal{T}_{\mathbf{m}} \leq 0 \\
&\Rightarrow \sum_{k_1=0}^{m_1} \cdots \sum_{k_{j-1}=0}^{m_{j-1}} \sum_{k_{j+1}=0}^{m_{j+1}} \cdots \sum_{k_N=0}^{m_N} Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_{j-1},m_j+1,k_{j+1},\ldots,k_N}\right) \leq \mathcal{T}_{\mathbf{m}} - \mathcal{T}_{\mathbf{m}_{(j)}} \\
&\Rightarrow \sum_{k_1=0}^{m_1} \cdots \sum_{k_{j-1}=0}^{m_{j-1}} \sum_{k_{j+1}=0}^{m_{j+1}} \cdots \sum_{k_N=0}^{m_N-1} Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_{j-1},m_j+1,k_{j+1},\ldots,k_N}\right) \\
&\qquad + Var_{\mathbf{X}}\left(\hat{C}_{m_1,\ldots,m_j+1,\ldots,m_N}\right) \leq \mathcal{T}_{\mathbf{m}} - \mathcal{T}_{\mathbf{m}_{(j)}} \\
&\Rightarrow Var_{\mathbf{X}}\left(\hat{C}_{m_1,\ldots,m_j+1,\ldots,m_N}\right) \leq \mathcal{T}_{\mathbf{m}} - \mathcal{T}_{\mathbf{m}_{(j)}} \quad \text{as } Var_{\mathbf{X}}\left(\hat{C}_{k_1,\ldots,k_N}\right) \geq 0 \text{ for all } k_1, \ldots, k_N \in \mathbb{N}_0
\end{aligned}$$

And hence as $m_j > K_j$:

$$\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}_{(j)}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] \leq \mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{m}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] \Rightarrow Var_{\mathbf{X}}\left(\hat{C}_{m_1,\ldots,m_j+1,\ldots,m_N}\right) \leq \mathcal{T}_{\mathbf{m}} - \mathcal{T}_{\mathbf{m}_{(j)}} < \epsilon_0$$

However as $Var_{\mathbf{X}}\left(\hat{C}_{n_1,\ldots,n_N}\right) \geq \epsilon_0$ for any $n_1,\ldots,n_N$.

$$Var_{\mathbf{X}}\left(\hat{C}_{m_1,\ldots,m_j+1,\ldots,m_N}\right) \geq \epsilon_0$$

Which is a contradiction.

Therefore we can conclude that if there exists $\epsilon_0 > 0$ such that $Var_{\mathbf{X}}(\hat{C}_{n_1,\ldots,n_N}) \geq \epsilon_0$ for all $\mathbf{n} = (n_1,\ldots,n_N)$ then for any $m_1,\ldots,m_{j-1},m_{j+1},\ldots,m_N$ there exists a K_j such that for all m_j where $m_j > K_j$ and $\mathbf{m} = (m_1,\ldots,m_N)$ then:

$$\mathbb{E}\left[\mathcal{N}_{\mathbf{m}_{(j)}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] > \mathbb{E}\left[\mathcal{N}_{\mathbf{m}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]$$

As this applies to any dimension $j \in \{1,\ldots,N\}$ then we can say that there exists a vector $\mathbf{K}_0$ whose elements are given by $K_{0,j} = K_j$ for each $j = 1,\ldots,N$ such that for any max polynomial order $\mathbf{K}$ where $K_i \geq K_{0,i}$ for at least a single $i = 1,\ldots,N$ then:

$$\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] > \mathbb{E}\left[\mathcal{N}_{\mathbf{K}_0}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]$$

Hence given that $\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] > 0$ for any $\mathbf{K} \in \mathbb{N}_0^N$ then an optimal max polynomial order (as defined in Definition 6) $\mathbf{K}^*$ given by: $\mathbf{K}^* = \text{argmin}_{\mathbf{K}\in\mathbb{N}_0^N}\{\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]\}$ can also be determined by considering the finite set $\mathcal{K} = \{\mathbf{K} = (K_1,\ldots,K_N) : K_i \in \mathbb{N}_0 \text{ and } K_i \leq K_{0,i} \text{ for each } i = 1,\ldots,N\}$ which is finite as $||\mathcal{K}||_c = \prod_{i=1}^{N}(K_{0,i}+1) < \infty$.
We note that for all $\mathbf{K} \in \mathbb{N}_0^N \backslash \mathcal{K}$

$$\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] > \mathbb{E}\left[\mathcal{N}_{\mathbf{K}_0}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]$$

Hence $\mathbb{E}\left[\mathcal{N}_{\mathbf{K}_0}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]$ is an infimum to the set $\{\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right] : \mathbf{K} \in \mathbb{N}_0^N \backslash \mathcal{K}\}$ and therefore we can say as $\mathbf{K}_0 \in \mathcal{K}$, that

$$\mathbf{K}^* = \text{argmin}_{\mathbf{K}\in\mathbb{N}_0^N}\{\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]\} = \text{argmin}_{\mathbf{K}\in\mathcal{K}}\{\mathbb{E}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]\}$$

Noting that all finite sets must have a minimum, we can then be sure that if there exists $\epsilon_0 > 0$ such that $Var_{\mathbf{X}}(\hat{C}_{n_1,\ldots,n_N}) \geq \epsilon_0$ for all $n_1,\ldots,n_N$, then the expected $\mathcal{N}_{\mathbf{K}}$-Norm $\mathbb{E}_{\mathbf{X}}\left[\mathcal{N}_{\mathbf{K}}\left(\{\mathbf{X}_i\}_{i=1}^{M}\right)\right]$ (Definition 5) has a finite optimal max polynomial order $\mathbf{K}^*$.