

Research Article

Frédéric Ouimet

On the boundary properties of Bernstein estimators on the simplex

<https://doi.org/10.1515/stat-2022-0111>

Received Jul 06, 2022; accepted Oct 25, 2022

Abstract: In this paper, we study the asymptotic properties (bias, variance, mean squared error) of Bernstein estimators for cumulative distribution functions and density functions near and on the boundary of the d -dimensional simplex. Our results generalize those found by Leblanc (2012b), who treated the case $d = 1$, and complement the results from Ouimet (2021a) in the interior of the simplex. Since the “edges” of the d -dimensional simplex have dimensions going from 0 (vertices) up to $d - 1$ (facets) and our kernel function is multinomial, the asymptotic expressions for the bias, variance and mean squared error are not straightforward extensions of one-dimensional asymptotics as they would be for product-type estimators studied by almost all past authors in the context of Bernstein estimators or asymmetric kernel estimators. This point makes the mathematical analysis much more interesting.

Keywords: asymptotic statistics, Bernstein estimators, simplex, multinomial distribution, nonparametric estimation, density estimation, cumulative distribution function estimation, boundary bias, variance, mean squared error.

2020 Mathematics Subject Classification: Primary: 62G05; Secondary: 60F99, 62G07, 62G20

1 Introduction

The d -dimensional (unit) simplex and its interior are defined by

$$\mathcal{S}_d := \{\mathbf{x} \in [0, 1]^d : \|\mathbf{x}\|_1 \leq 1\} \quad \text{and} \quad \text{Int}(\mathcal{S}_d) := \{\mathbf{x} \in (0, 1)^d : \|\mathbf{x}\|_1 < 1\},$$

where $\|\mathbf{x}\|_1 := \sum_{i=1}^d |x_i|$. For any cumulative distribution function F on \mathcal{S}_d (meaning that it takes the values 0 or 1 outside \mathcal{S}_d), define the Bernstein polynomial of order m for F by

$$F_m^*(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d} F(\mathbf{k}/m) P_{\mathbf{k},m}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}_d, \quad m \in \mathbb{N},$$

where the weights are the following probabilities from the Multinomial(m, \mathbf{x}) distribution:

$$P_{\mathbf{k},m}(\mathbf{x}) := \frac{m!}{(m - \|\mathbf{k}\|_1)! \prod_{i=1}^d k_i!} (1 - \|\mathbf{x}\|_1)^{m - \|\mathbf{k}\|_1} \prod_{i=1}^d x_i^{k_i}, \quad \mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d.$$

(Here, the set $m\mathcal{S}_d$ has the obvious meaning $\{\mathbf{y} \in \mathbb{R}^d : \mathbf{y} = m\mathbf{x} \text{ for some } \mathbf{x} \in \mathcal{S}_d\}$.) The Bernstein estimator of F , denoted by $F_{n,m}^*$, is the Bernstein polynomial of order m for the empirical cumulative distribution function

$$F_n(\mathbf{x}) := n^{-1} \sum_{i=1}^n \mathbb{1}_{(-\infty, \mathbf{x}]}(X_i),$$

Frédéric Ouimet: Centre de recherches mathématiques, Université de Montréal, Montreal, QC H3T 1J4, Canada. Department of Mathematics and Statistics, McGill University, Montreal, QC H3A 0B9, Canada. Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, United States of America

where the observations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are assumed to be independent and F distributed. Precisely, for $m, n \in \mathbb{N}$, let

$$F_{n,m}^*(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d} F_n(\mathbf{k}/m) P_{\mathbf{k},m}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}_d. \quad (1)$$

It should be noted that the c.d.f. estimator in (1) only makes sense here if the observations' support is contained in an hyperrectangle inside the unit simplex. If the observations have full support on the unit simplex, then the c.d.f. estimator would take values in $(0, 1)$ on the interior of the unit hypercube.

For a density f supported on \mathcal{S}_d , we define the Bernstein density estimator of f by

$$\tilde{f}_{n,m}(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} m^d \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(\frac{\mathbf{k}}{m}, \frac{\mathbf{k}+1}{m}]}(\mathbf{X}_i) \right\} P_{\mathbf{k},m-1}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{S}_d, \quad (2)$$

where m^d is a scaling factor equal to the inverse of the volume of the hypercube $(\frac{\mathbf{k}}{m}, \frac{\mathbf{k}+1}{m}] := (\frac{k_1}{m}, \frac{k_1+1}{m}] \times \dots \times (\frac{k_d}{m}, \frac{k_d+1}{m}]$. As pointed out in Ouimet (2021a), the estimator $\tilde{f}_{n,m}$ is not a proper density but integrates to 1 asymptotically, and it can be written as a finite mixture of Dirichlet densities (in the one-dimensional case, finite beta mixtures are studied for example in McLachlan and Peel (2000)).

2 Review of the literature and motivation

Vitale (1975) was the first author to consider Bernstein density estimation on the compact interval $[0, 1]$, namely (2) with $d = 1$. The author computes the asymptotics of the bias, variance and mean squared error (MSE) at each point where the second derivative of f exists (also assuming that f is bounded everywhere). His proof rests on careful Taylor expansions for the density terms inside the bulk of the binomial distribution while concentration bounds are applied to show that the contributions coming from outside the bulk are negligible. The optimal rate, with respect to the MSE, is achieved when $m \asymp n^{2/5}$ and shown to be $\mathcal{O}_x(n^{-4/5})$ for $x \in (0, 1)$. For $x \in \{0, 1\}$ (at the boundary), he finds that the MSE is $\mathcal{O}_x(n^{-3/5})$ using $m \asymp n^{2/5}$, but this choice of m turns out to be suboptimal; the rate $\mathcal{O}_x(n^{-2/3})$ can be achieved with $m \asymp n^{1/3}$. These results were generalized by Gawronski and Stadtmüller (1980, 1981) to target densities with support of the form $[0, 1]$, $[0, \infty)$ and $(-\infty, \infty)$, where the binomial weights of the density estimator were replaced by more general convolutions.

Tenbusch (1994) was the first author to consider Bernstein estimation in the multivariate context. Assuming that f is twice continuously differentiable, he derived asymptotic expressions for the bias, variance and MSE of the density estimators on the two-dimensional unit simplex and the unit square $[0, 1]^2$, and also proved their uniform strong consistency and asymptotic normality. He showed that the optimal rate of the MSE is $\mathcal{O}_x(n^{-2/3})$ for interior points when $d = 2$, and it is achieved when $m \asymp n^{1/3}$. On the “edges” of dimension 1 (i.e., excluding the corners), he showed that the optimal rate is $\mathcal{O}_x(n^{-4/7})$ and it is achieved when $m \asymp n^{2/7}$. On the “edges” of dimension 0 (i.e., the three corners of the 2-dimensional simplex), he showed that the optimal rate is $\mathcal{O}_x(n^{-1/2})$ and it is achieved when $m \asymp n^{1/4}$. As we will see, this deterioration of the MSE near the boundary is solely due to an increase of the variance of the density estimator, while the bias remains asymptotically uniform and of order $\mathcal{O}(m^{-1})$. This is in contrast with traditional kernel estimators, where the opposite trade-off happens: the bias increases near the boundary while the variance remains asymptotically the same. For both classes of estimators, the rates of the MSE might be the same on the boundary but for different reasons.

It should be noted that this variance increase near and on the boundary was previously observed by other authors for product-type asymmetric kernel estimators (meaning when the kernel is a product of one-dimensional kernels) on the geometrically simpler domains $[0, 1]^d$ and $[0, \infty)^d$, see, e.g., (Bouezmarni and Rombouts, 2010, page 142 and (7a)), (Igarashi and Kakizawa, 2020, pages 50 and 52) and (Kakizawa, 2022, Corollary 1). Leblanc (2012b) also observed this phenomenon in one dimension and provided explicit asymptotic expressions for the bias, variance and MSE of the classical Bernstein density and c.d.f. estimators near

and on the boundary of $[0, 1]$. Our main goal in this paper is to extend Leblanc's results to the d -dimensional simplex domain and complete the theoretical analysis initiated by Ouimet (2021a) on the asymptotic properties of Bernstein density and c.d.f. estimators on the simplex (with multinomial weights).

In Leblanc (2012b), Leblanc extended the results of Vitale (1975) by not only studying the asymptotics of the bias and variance for points that are on the boundary but also *near* the boundary (i.e., for $x = \lambda/m$ where $\lambda \geq 0$ remains fixed as $n, m \rightarrow \infty$). In particular, he showed that the leading term of the variance can be written using modified Bessel functions of the first kind. He also proved similar results for the Bernstein c.d.f. estimator. As mentioned above, our goal in this paper is to generalize the results of Leblanc (2012b) to all $d \geq 1$. The results will help us understand how the dimension of the “edge” we are closest to influences the variance and the MSE of the density estimator and the c.d.f. estimator.

Since the “edges” of the d -dimensional simplex have dimensions going from 0 (vertices) up to $d-1$ (facets) and our kernel function is multinomial, the asymptotic expressions for the bias, variance and MSE are not straightforward extensions of the one-dimensional expressions as they would be for product-type estimators such as the ones in (Bouezmarni and Rombouts, 2010, page 142 and (7a)), (Igarashi and Kakizawa, 2020, pages 50 and 52) and (Kakizawa, 2022, Corollary 1). This is why the results in the present paper are of mathematical interest. Apart from the present paper, the only article dealing with boundary results for multidimensional non product-type Bernstein estimators or asymmetric kernel estimators that we are aware of seems to be Corollary 2 of Kakizawa (2022), where the variance of a multivariate elliptical-based Birnbaum–Saunders kernel density estimator is computed near and on the boundary.

Other asymptotic results (not directly related to the boundary, which is our focus here) have been proved in the unidimensional case by Babu et al. (2002); Kakizawa (2004); Bouezmarni and Scaillet (2005); Bouezmarni and Rolin (2007); Leblanc (2009, 2010, 2012a); Igarashi and Kakizawa (2014); Lu (2015); Liu and Ghosh (2020); Lu and Wang (2021); Wang and Lu (2022), among other authors, and in the multidimensional case by Babu and Chaubey (2006); Belalia (2016); Ouimet (2021a). Most of these papers (and more) are reviewed in Section 2 of Ouimet (2021a). Similarly, for a detailed overview of the literature on the closely related Beta and Dirichlet kernel estimators, we refer to Section 2 of Ouimet and Tolosana-Delgado (2022).

3 Contribution, outline and notation

3.1 Contribution

Our theoretical contribution is to find asymptotic expressions for the bias, variance and MSE of the Bernstein c.d.f. and density estimators, defined respectively in (1) and (2), *near* and on the boundary of the d -dimensional simplex. We also deduce the asymptotically optimal choice of the bandwidth parameter m using the expressions for the MSE, which can be used in practice to implement a plug-in selection method. All these results generalize those found in Leblanc (2012b) for the unit interval and complement those found in Ouimet (2021a) in the interior of the simplex. Our rates of convergence for the MSE are in line with those recently found in Ouimet and Tolosana-Delgado (2022) for Dirichlet kernel estimator. In both cases, the general rule is that the variable smoothing that is built-in for Bernstein estimators and asymmetric kernel estimators yields an asymptotically smaller bias near the boundary at the cost of an increase in variance, compared to traditional multivariate kernel estimators.

Under the assumption that the target density is twice continuously differentiable, we find in particular that the variance is $\mathcal{O}_x(n^{-1}m^{d/2})$ in the interior of the simplex and it gets multiplied by a factor $m^{1/2}$ everytime we go near the boundary in one of the d dimensions. If we are near an edge of dimension $d - |\mathcal{J}|$ (see Section 4 for the definition of \mathcal{J}), then the variance is $\mathcal{O}_x(n^{-1}m^{(d+|\mathcal{J}|)/2})$. Additional smoothness conditions on the partial derivatives of the target density can improve those rates, see, e.g., Corollary 2 for more details on this point.

In contrast to other methods of boundary bias reduction (such as the reflection method or boundary kernels (see, e.g., (Scott, 2015, Chapter 6)), this property is built-in for Bernstein estimators, which makes them one of the easiest to use in the class of estimators that are asymptotically unbiased near (and on) the boundary. Bernstein estimators are also non-negative everywhere on their domain, which is definitely not the case of many estimators corrected for boundary bias. This is another reason for their desirability. Also, as an anonymous referee pointed out, the fact that the bandwidth parameter m is an integer makes the optimisation step easier to implement (for most bandwidth selection criteria like least-square cross validation, likelihood cross-validation, etc.) than for estimators where the bandwidth parameter h is a real number. Bandwidth selection methods and their consistency will be investigated thoroughly in upcoming work.

3.2 Outline

In Section 4 and Section 5, we state our results for the density estimator and the c.d.f. estimator, respectively. The proofs are given in Section 6 and Section 7. Some technical lemmas and tools are gathered in Appendix A.

3.3 Notation

Throughout the paper, the notation $u = \mathcal{O}(v)$ means that $\limsup |u/v| < C < \infty$ as m or n tend to infinity, depending on the context. The positive constant C can depend on the target c.d.f. F , the target density f or the dimension d , but no other variable unless explicitly written as a subscript. One common occurrence is a local dependence of the asymptotics with a given point \mathbf{x} on the simplex, in which case we would write $u = \mathcal{O}_{\mathbf{x}}(v)$. In a similar fashion, the notation $u = o(v)$ means that $\lim |u/v| = 0$ as m or n tend to infinity. The same rule applies for the subscript. The symbol \mathcal{D} over an arrow ' \longrightarrow ' will denote the convergence in law (or distribution). We will use the shorthand $[d] := \{1, 2, \dots, d\}$ in several places. The functions

$$I_\nu(z) := \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\nu}}{k!(k+\nu)!}, \quad \nu \in \{0, 1\}, \quad (3)$$

will denote the modified Bessel functions of the first kind of order 0 and 1, respectively. For any vector $\mathbf{v} \in \mathbb{R}^d$ and any subset of indices $A \subseteq [d]$, we write

$$\mathbf{v}_A := (v_i)_{i \in A} \quad \text{and} \quad \psi_A(\mathbf{x}) := \left[(4\pi)^{|A|} \left(1 - \sum_{i \in A} x_i \right) \prod_{i \in A} x_i \right]^{-1/2}, \quad (4)$$

with the conventions $\sum_{\emptyset} := 0$ and $\prod_{\emptyset} := 1$. Finally, the bandwidth parameter $m = m(n)$ is always implicitly a function of the number of observations, the only exceptions being in Lemmas 1, 3, 5, 6 and the related proofs.

4 Results for the density estimator $\tilde{f}_{n,m}$

For every result stated in this section, we will make the following assumption:

- $f \in C^2(\text{Int}(\mathcal{S}_d))$ and there exists an open set $\mathcal{U} \subseteq \mathbb{R}^d$ that contains \mathcal{S}_d and an extension $f_{\text{ext}}: \mathcal{U} \rightarrow \mathbb{R}$ such that $f_{\text{ext}} \equiv f$ on \mathcal{S}_d and $f_{\text{ext}} \in C^2(\mathcal{U})$. (5)

Remark 1. If $f \in C^2(\mathbb{R}^d)$, then we can just take $f_{\text{ext}} = f$. However, if f is discontinuous anywhere at some point on the boundary of \mathcal{S}_d , then, at that point, the partial derivatives of f will technically refer to the partial derivatives of f_{ext} .

In the first lemma, we obtain a general expression for the bias of the density estimator.

Lemma 1 (Bias of $\tilde{f}_{n,m}(\mathbf{x})$ on \mathcal{S}_d). Assume (5). Then, uniformly for $\mathbf{x} \in \mathcal{S}_d$, we have, as $m \rightarrow \infty$,

$$\mathbb{B}\text{ias}[\tilde{f}_{n,m}(\mathbf{x})] = m^{-1}\Delta_1(\mathbf{x}) + m^{-2}\Delta_2(\mathbf{x}) + \frac{1}{m^2} \sum_{i \in [d]} o\left(1 + \sqrt{(m-1)x_i(1-x_i)} + (m-1)x_i(1-x_i)\right),$$

where

$$\begin{aligned} \Delta_1(\mathbf{x}) &:= \sum_{i \in [d]} \left(\frac{1}{2} - x_i\right) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{2} \sum_{i,j \in [d]} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}), \\ \Delta_2(\mathbf{x}) &:= \sum_{i,j \in [d]} \left(\frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} - \frac{1}{2} x_i \mathbb{1}_{\{i=j\}} - \frac{1}{2} x_j + x_i x_j\right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}). \end{aligned} \quad (6)$$

By considering points $\mathbf{x} \in \mathcal{S}_d$ that are close to the boundary in some components (see the subset of indices $\mathcal{J} \subseteq [d]$ below), we get the bias of the density estimator near the boundary.

Theorem 1 (Bias of $\tilde{f}_{n,m}(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (5). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{B}\text{ias}(\tilde{f}_{n,m}(\mathbf{x})) &= \frac{b_{\mathcal{J}}(\mathbf{x})}{m} + \frac{1}{m^2} \left\{ - \sum_{i \in [d]} \lambda_i \frac{\partial}{\partial x_i} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} \right. \\ &\quad \left. + \sum_{i,j \in [d]} \left(\left(\frac{1}{6} + \lambda_i\right) \mathbb{1}_{\{i=j\}} + \left(\frac{1}{8} + \frac{\lambda_j}{2}\right) \mathbb{1}_{\{i \neq j\}} \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} \right\} \\ &\quad + o_{\lambda_{\mathcal{J}}}(m^{-2} + \mathbb{1}_{\{\mathcal{J} \neq [d]\}} m^{-1}), \end{aligned} \quad (7)$$

where $\mathbf{x}_{\mathcal{J}}$ and $\lambda_{\mathcal{J}}$ are defined in (4), and where

$$b_{\mathcal{J}}(\mathbf{x}) := \sum_{i \in [d]} \left(\frac{1}{2} - x_i \mathbb{1}_{\{i \in [d] \setminus \mathcal{J}\}}\right) \frac{\partial}{\partial x_i} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \sum_{i,j \in [d] \setminus \mathcal{J}} \frac{1}{2} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}}. \quad (8)$$

Next, we obtain a general expression for the variance of the density estimator.

Lemma 2 (Variance of $\tilde{f}_{n,m}(\mathbf{x})$ on \mathcal{S}_d). Assume (5). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \text{Var}(\tilde{f}_{n,m}(\mathbf{x})) &= \frac{m^d}{n} f(\mathbf{x}) \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k},m-1}^2(\mathbf{x}) \\ &\quad + \frac{m^{d-1}}{n} \sum_{i \in [d]} \mathcal{O} \left(1 + \sqrt{\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} |k_i - m x_i|^2 P_{\mathbf{k},m-1}(\mathbf{x})} \sqrt{\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k},m-1}^3(\mathbf{x})} \right). \end{aligned}$$

By combining Lemma 2 and the technical estimate in Lemma 5, we get the asymptotics of the variance of the density estimator near the boundary.

Theorem 2 (Variance of $\tilde{f}_{n,m}(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (5). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\text{Var}(\tilde{f}_{n,m}(\mathbf{x})) = n^{-1} m^{(d+|\mathcal{J}|)/2} \left\{ v_{\mathcal{J}}(\mathbf{x}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + o_{\mathbf{x}_{[d] \setminus \mathcal{J}}}(1) \mathbb{1}_{\{\mathcal{J} \neq [d]\}} \right\},$$

where $\mathbf{x}_{[d] \setminus \mathcal{J}}$ and $\lambda_{\mathcal{J}}$ are defined in (4), and where

$$v_{\mathcal{J}}(\mathbf{x}) := f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} \psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) \prod_{i \in \mathcal{J}} e^{-2\lambda_i} I_0(2\lambda_i). \quad (9)$$

By combining Theorem 1 and Theorem 2, we get the asymptotics of the mean squared error of the density estimator near and on the boundary. In particular, the optimal bandwidth parameter m will depend on the number of components of \mathbf{x} that are close to the boundary.

Corollary 1 (MSE of $\tilde{f}_{n,m}(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (5). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \text{MSE}(\tilde{f}_{n,m}(\mathbf{x})) &:= \mathbb{E}[(\tilde{f}_{n,m}(\mathbf{x}) - f(\mathbf{x}))^2] = \text{Var}(\tilde{f}_{n,m}(\mathbf{x})) + (\text{Bias}[\tilde{f}_{n,m}(\mathbf{x})])^2 \\ &= n^{-1} m^{(d+|\mathcal{J}|)/2} v_{\mathcal{J}}(\mathbf{x}) + m^{-2} b_{\mathcal{J}}^2(\mathbf{x}) + n^{-1} m^{(d+|\mathcal{J}|)/2} [\mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) \\ &\quad + \mathcal{O}_{\mathbf{x}_{[d] \setminus \mathcal{J}}}(1) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}] + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-3}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-2}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \end{aligned} \quad (10)$$

where $\mathbf{x}_{\mathcal{J}}$, $\mathbf{x}_{[d] \setminus \mathcal{J}}$ and $\lambda_{\mathcal{J}}$ are defined in (4), $b_{\mathcal{J}}(\mathbf{x})$ is defined in (8), and $v_{\mathcal{J}}(\mathbf{x})$ is defined in (9). If the quantity inside the big bracket is non-zero in (10), the asymptotically optimal choice of m , with respect to MSE, is

$$m_{\text{opt}} = n^{\frac{2}{d+|\mathcal{J}|+4}} \frac{\{b_{\mathcal{J}}(\mathbf{x})\}^{\frac{4}{d+|\mathcal{J}|+4}}}{\left\{\frac{d+|\mathcal{J}|}{4} v_{\mathcal{J}}(\mathbf{x})\right\}^{\frac{2}{d+|\mathcal{J}|+4}}},$$

($\psi_{[d] \setminus \mathcal{J}}$ is defined in (4)) in which case, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \text{MSE}(\tilde{f}_{n,m_{\text{opt}}}(\mathbf{x})) &= n^{-\frac{4}{d+|\mathcal{J}|+4}} \frac{\{b_{\mathcal{J}}(\mathbf{x})\}^{\frac{2(d+|\mathcal{J}|)}{d+|\mathcal{J}|+4}}}{\left[\left(\frac{\frac{4}{d+|\mathcal{J}|}}{\frac{4}{d+|\mathcal{J}|}+1}\right)^{\frac{4}{d+|\mathcal{J}|+4}}\right] \{v_{\mathcal{J}}(\mathbf{x})\}^{\frac{-4}{d+|\mathcal{J}|+4}}} \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(n^{-\frac{6}{d+|\mathcal{J}|+4}} + \mathbb{1}_{\{\mathcal{J} \neq [d]\}} n^{-\frac{5}{d+|\mathcal{J}|+4}}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(n^{-\frac{4}{d+|\mathcal{J}|+4}}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}. \end{aligned}$$

By imposing further conditions on the partial derivatives of f , we can remove terms from the bias in Theorem 1 and obtain another expression for the mean squared error of the density estimator near and on the boundary, and also the corresponding optimal bandwidth parameter m when $\mathcal{J} = [d]$.

Corollary 2 (MSE of $\tilde{f}_{n,m}(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (5) and also

$$\frac{\partial}{\partial x_i} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} = 0 \quad \forall i \in [d], \quad (\text{shoulder condition}) \quad \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} = 0 \quad \forall (i, j) \in ([d] \setminus \mathcal{J})^2, \quad (11)$$

(in particular, the first bracket in (7) is zero). Then, for any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \text{MSE}(\tilde{f}_{n,m}(\mathbf{x})) &= n^{-1} m^{(d+|\mathcal{J}|)/2} v_{\mathcal{J}}(\mathbf{x}) + m^{-4} L_{\mathcal{J}}^2(\mathbf{x}) \\ &\quad + n^{-1} m^{(d+|\mathcal{J}|)/2} [\mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + \mathcal{O}_{\mathbf{x}_{[d] \setminus \mathcal{J}}}(1) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}] \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-4} + \mathbb{1}_{\{\mathcal{J} \neq [d]\}} m^{-3}), \end{aligned} \quad (12)$$

where $\mathbf{x}_{\mathcal{J}}$, $\mathbf{x}_{[d] \setminus \mathcal{J}}$ and $\lambda_{\mathcal{J}}$ are defined in (4), $v_{\mathcal{J}}(\mathbf{x})$ is defined in (9), and where

$$L_{\mathcal{J}}(\mathbf{x}) := \sum_{(i,j) \in [d]^2 \setminus ([d] \setminus \mathcal{J})^2} \left(\left(\frac{1}{6} + \lambda_i \right) \mathbb{1}_{\{i=j\}} + \left(\frac{1}{8} + \frac{\lambda_j}{2} \right) \mathbb{1}_{\{i \neq j\}} \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}}.$$

Note that the last error term in (12) is bigger than the main term except when $\mathcal{J} = [d]$. Therefore, if $\mathcal{J} = [d]$ and we assume that the quantity inside the big bracket is non-zero in (12), the asymptotically optimal choice of m , with respect to MSE, is

$$m_{\text{opt}} = n^{\frac{1}{d+4}} \frac{\{L_{[d]}(\mathbf{0})\}^{\frac{2}{d+4}}}{\left\{\frac{d}{4} v_{[d]}(\mathbf{0})\right\}^{\frac{1}{d+4}}},$$

so that

$$\text{MSE}(\tilde{f}_{n,m_{\text{opt}}}(\mathbf{x})) = n^{-\frac{4}{d+4}} \frac{\{L_{[d]}(\mathbf{0})\}^{\frac{2d}{d+4}}}{\left[\left(\frac{\frac{4}{d}}{\frac{4}{d}+1}\right)^{\frac{4}{d+4}}\right] \{v_{[d]}(\mathbf{0})\}^{\frac{-4}{d+4}}} + \mathcal{O}_{\lambda}(n^{-\frac{4}{d+4}}), \quad \text{as } n \rightarrow \infty.$$

Remark 2. The last part shows that, when we assume the smoothness conditions (11), the optimal rate of the MSE near any vertex of the simplex is of the same order as in the interior of simplex (Corollary 1 with $\mathcal{J} = \emptyset$) without the smoothness conditions.

Remark 3. In order to optimize m when $\mathcal{J} \neq [d]$ in Corollary 2, we would need a more precise expression for the bias in Theorem 7 by assuming more regularity conditions on f than we did in (5).

5 Results for the c.d.f. estimator $F_{n,m}^*$

For every result stated in this section, we will make the following assumption:

- F is three-times continuously differentiable on \mathcal{S}_d . (13)

Below, we obtain a general expression for the bias of the c.d.f. estimator on the simplex, and then near the boundary.

Lemma 3 (Bias of $F_{n,m}^*(\mathbf{x})$ on \mathcal{S}_d). Assume (13). Then, uniformly for $\mathbf{x} \in \mathcal{S}_d$, we have, as $m \rightarrow \infty$,

$$\begin{aligned} \mathbb{E}[F_{n,m}^*(\mathbf{x})] &= F(\mathbf{x}) + \frac{1}{2m} \sum_{i,j \in [d]} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \\ &\quad + \frac{1}{6m^2} \sum_{i,j,\ell \in [d]} (2x_i x_j x_\ell - \mathbb{1}_{\{i=j\}} x_i x_\ell - \mathbb{1}_{\{j=\ell\}} x_i x_j - \mathbb{1}_{\{i=\ell\}} x_j x_\ell + \mathbb{1}_{\{i=j=\ell\}} x_i) \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \\ &\quad + \frac{1}{m^3} \sum_{i,j,\ell \in [d]} o\left(\left(\mathbb{E}[|\xi_i - m x_i|^2]\right)^{1/2} \left(\mathbb{E}[|\xi_j - m x_j|^4]\right)^{1/4} \left(\mathbb{E}[|\xi_\ell - m x_\ell|^4]\right)^{1/4}\right), \end{aligned}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_d) \sim \text{Multinomial}(m, \mathbf{x})$.

Theorem 3 (Bias of $F_{n,m}^*(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (13). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i > 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\mathbb{B}\text{ias}(F_{n,m}^*(\mathbf{x})) = m^{-1} B_{\mathcal{J}}(\mathbf{x}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-3}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-3/2}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}},$$

where

$$B_{\mathcal{J}}(\mathbf{x}) := \frac{1}{2} \sum_{i,j \in [d] \setminus \mathcal{J}} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \frac{1}{2m} \sum_{i \in [d]} \lambda_i \frac{\partial^2}{\partial x_i^2} F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{0}}. \quad (14)$$

In the case where $x_i = 0$ for some $i \in \mathcal{J} \neq \emptyset$, notice that $\mathbb{B}\text{ias}(F_{n,m}^*(\mathbf{x})) = 0$ because $F_{n,m}^*(\mathbf{x}) = 0$ a.s.

Next, we obtain a general expression for the variance of the c.d.f. estimator on the simplex.

Lemma 4 (Variance of $F_{n,m}^*(\mathbf{x})$ on \mathcal{S}_d). Assume (13). Then, uniformly for $\mathbf{x} \in \mathcal{S}_d$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) &= n^{-1} F(\mathbf{x})(1 - F(\mathbf{x})) - n^{-1} \sum_{i \in [d]} \frac{\partial}{\partial x_i} F(\mathbf{x}) \sum_{\mathbf{k}, \ell \in \mathbb{N}_0^d \cap m\mathcal{S}_d} ((k_i \wedge \ell_i)/m - x_i) P_{\mathbf{k},m}(\mathbf{x}) P_{\ell,m}(\mathbf{x}) \\ &\quad + n^{-1} \sum_{i,j \in [d]} \mathcal{O}(m^{-1} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j)) + \mathcal{O}(n^{-1} m^{-2}). \end{aligned}$$

By combining Lemma 6 and Lemma 4, we get the asymptotics of the variance of the c.d.f. estimator near the boundary.

Theorem 4 (Variance of $F_{n,m}^*(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (13). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i > 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) &= n^{-1}m^{-1}V_{\mathcal{J}}(\mathbf{x}) + n^{-1}F(\mathbf{x})(1 - F(\mathbf{x}))\mathbb{1}_{\{\mathcal{J}=\emptyset\}} \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}}(n^{-1}m^{-2}) + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(n^{-1}m^{-1/2})\mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \end{aligned}$$

where

$$V_{\mathcal{J}}(\mathbf{x}) := \sum_{i \in [d]} \frac{\partial}{\partial x_i} F(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} \left\{ \begin{array}{l} \lambda_i (1 - e^{-2\lambda_i}(I_0(2\lambda_i) + I_1(2\lambda_i)))\mathbb{1}_{\{i \in \mathcal{J}\}} \\ -m^{1/2} \sqrt{\pi^{-1}x_i(1-x_i)}\mathbb{1}_{\{i \in [d] \setminus \mathcal{J}\}} \end{array} \right\}. \quad (15)$$

In the case where $x_i = 0$ for some $i \in \mathcal{J} \neq \emptyset$, notice that $\mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) = 0$ because $F_{n,m}^*(\mathbf{x}) = 0$ a.s.

By combining Theorem 3 and Theorem 4, we get the asymptotics of the mean squared error of the c.d.f. estimator near the boundary.

Corollary 3 (MSE of $F_{n,m}^*(\mathbf{x})$ near the boundary of \mathcal{S}_d). Assume (13). For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i > 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ is independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \text{MSE}(F_{n,m}^*(\mathbf{x})) &= n^{-1}m^{-1}V_{\mathcal{J}}(\mathbf{x}) + n^{-1}F(\mathbf{x})(1 - F(\mathbf{x}))\mathbb{1}_{\{\mathcal{J}=\emptyset\}} + m^{-2}B_{\mathcal{J}}^2(\mathbf{x}) \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}}(n^{-1}m^{-2}) + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(n^{-1}m^{-1/2})\mathbb{1}_{\{\mathcal{J} \neq [d]\}} \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-4}) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-5/2})\mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \end{aligned}$$

where $B_{\mathcal{J}}(\mathbf{x})$ is defined in (14) and $V_{\mathcal{J}}(\mathbf{x})$ is defined in (15). In the case where $x_i = 0$ for some $i \in \mathcal{J} \neq \emptyset$, notice that $\text{MSE}(F_{n,m}^*(\mathbf{x})) = 0$ because $F_{n,m}^*(\mathbf{x}) = 0$ a.s. Furthermore, as pointed out by (Leblanc, 2012b, p.2772) for $d = 1$, there is no optimal m with respect to the MSE when $\mathcal{J} \neq \emptyset$. This is also true here. The remaining case $\mathcal{J} = \emptyset$ (i.e., when \mathbf{x} is far from the boundary in every component) was already treated in Corollary 1 of Ouimet (2021a).

6 Proof of the results for the density estimator $\tilde{f}_{n,m}$

6.1 Proof of Lemma 1

Take $\delta_n \searrow 0$ slow enough as $n \rightarrow \infty$ (for example $\delta_n \geq m^{-1/4}$) that standard concentration bounds for the binomial distribution yield

$$\sum_{\substack{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d \\ \|\mathbf{k}/m - \mathbf{x}\|_1 > \delta_n}} P_{\mathbf{k},m}(\mathbf{x}) \leq \sum_{\ell \in [d]} \sum_{\substack{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d \\ |k_{\ell}/m - x_{\ell}| > \delta_n/d}} P_{\mathbf{k},m}(\mathbf{x}) \leq C_d e^{-c_d \delta_n^2 m} = o(m^{-2}), \quad (16)$$

for some appropriate constants $c_d, C_d > 0$. For any \mathbf{k} such that $\|\mathbf{k}/m - \mathbf{x}\|_1 \leq \delta_n$, we can use Taylor expansions and our assumption that f is twice continuously differentiable to obtain

$$\begin{aligned} m^d \int_{\left(\frac{\mathbf{k}}{m}, \frac{\mathbf{k}+1}{m}\right]} f(\mathbf{y}) d\mathbf{y} - f(\mathbf{x}) &= f(\mathbf{k}/m) - f(\mathbf{x}) + \frac{1}{2m} \sum_{i \in [d]} \frac{\partial}{\partial x_i} f(\mathbf{k}/m) + \frac{1}{m^2} \sum_{i,j \in [d]} \left(\frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{k}/m) (1 + o(1)) \\ &= \frac{1}{m} \sum_{i \in [d]} (k_i - mx_i) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{2m^2} \sum_{i,j \in [d]} (k_i - mx_i)(k_j - mx_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) (1 + o(1)) \\ &\quad + \frac{1}{2m} \sum_{i \in [d]} \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{2m^2} \sum_{i,j \in [d]} (k_j - mx_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) (1 + o(1)) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{m^2} \sum_{i,j \in [d]} \left(\frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) (1 + o(1)) \\
& = \frac{1}{m} \sum_{i \in [d]} (k_i - (m-1)x_i) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{m} \sum_{i \in [d]} \left(\frac{1}{2} - x_i \right) \frac{\partial}{\partial x_i} f(\mathbf{x}) \\
& + \frac{1}{2m^2} \sum_{i,j \in [d]} (k_i - (m-1)x_i)(k_j - (m-1)x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& - \frac{1}{2m^2} \sum_{i,j \in [d]} [2x_i(k_j - (m-1)x_j) - x_i x_j] \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) + \frac{1}{2m^2} \sum_{i,j \in [d]} (k_j - (m-1)x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \frac{1}{m^2} \sum_{i,j \in [d]} \left(\frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} - \frac{1}{2} x_j \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \frac{1}{m^2} \sum_{i \in [d]} o(1 + |k_i - (m-1)x_i| + |k_i - (m-1)x_i|^2).
\end{aligned}$$

If we multiply the last expression by $P_{\mathbf{k}, m-1}(\mathbf{x})$ and sum over $\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d$, then the joint moments from Lemma 7 and Jensen's inequality yield

$$\begin{aligned}
\mathbb{E}[\tilde{f}_{n,m}(\mathbf{x})] - f(\mathbf{x}) & = \frac{1}{m} \sum_{i \in [d]} \left(\frac{1}{2} - x_i \right) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{m-1}{2m^2} \sum_{i,j \in [d]} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \frac{1}{m^2} \sum_{i,j \in [d]} \left\{ \frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} - \frac{1}{2} x_j + \frac{1}{2} x_i x_j \right\} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \frac{1}{m^2} \sum_{i \in [d]} o \left(1 + \sqrt{(m-1)x_i(1-x_i)} + (m-1)x_i(1-x_i) \right).
\end{aligned}$$

The conclusion follows.

6.2 Proof of Theorem 1

Take \mathbf{x} as in the statement of the theorem. Using the notation from (6), we have

$$\begin{aligned}
\Delta_1(\mathbf{x}) & = \sum_{i \in \mathcal{J}} \left(\frac{1}{2} - \frac{\lambda_i}{m} \right) \frac{\partial}{\partial x_i} f(\mathbf{x}) + \sum_{i \in \mathcal{J}} \frac{\lambda_i}{2m} \frac{\partial^2}{\partial x_i^2} f(\mathbf{x}) + \sum_{i \in [d] \setminus \mathcal{J}} \left(\frac{1}{2} - x_i \right) \frac{\partial}{\partial x_i} f(\mathbf{x}) - \sum_{\substack{i \in \mathcal{J} \\ j \in [d] \setminus \mathcal{J}}} \frac{\lambda_i x_j}{m} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \sum_{i,j \in [d] \setminus \mathcal{J}} \frac{1}{2} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) + o_{\lambda_{\mathcal{J}}}(m^{-2}),
\end{aligned} \tag{17}$$

and

$$\begin{aligned}
\Delta_2(\mathbf{x}) & = \sum_{(i,j) \in [d]^2 \setminus ([d] \setminus \mathcal{J})^2} \left(\frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} \right) \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \\
& + \sum_{i,j \in [d] \setminus \mathcal{J}} \left\{ \frac{1}{6} \mathbb{1}_{\{i=j\}} + \frac{1}{8} \mathbb{1}_{\{i \neq j\}} - \frac{1}{2} x_i \mathbb{1}_{\{i=j\}} - \frac{1}{2} x_j + x_i x_j \right\} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) + o_{\lambda_{\mathcal{J}}}(m^{-1}).
\end{aligned} \tag{18}$$

For all $i, j \in [d]$, note that

$$\begin{aligned}
\frac{\partial}{\partial x_i} f(\mathbf{x}) & = \frac{\partial}{\partial x_i} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \sum_{j \in \mathcal{J}} \frac{\lambda_j}{m} \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} (1 + o_{\lambda_{\mathcal{J}}}(1)), \\
\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) & = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \Big|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} (1 + o_{\lambda_{\mathcal{J}}}(1)),
\end{aligned} \tag{19}$$

Then, from (17), (18) (19) and Lemma 1, we can easily deduce the conclusion.

6.3 Proof of Lemma 2

By the independence of the observations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, we have

$$\mathbb{V}\text{ar}(\tilde{f}_{n,m}(\mathbf{x})) = n^{-1} \left\{ m^{2d} \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d(\frac{\mathbf{k}}{m}, \frac{\mathbf{k}+1}{m})} \int f(\mathbf{y}) d\mathbf{y} P_{\mathbf{k},m-1}^2(\mathbf{x}) - \mathbb{E}[\tilde{f}_{n,m}(\mathbf{x})]^2 \right\}. \quad (20)$$

From Lemma 1, we already know that $\mathbb{E}[\tilde{f}_{n,m}(\mathbf{x})] = f(\mathbf{x}) + \mathcal{O}(m^{-1})$, uniformly for $\mathbf{x} \in \mathcal{S}_d$. We can also expand the integral using a Taylor expansion:

$$\begin{aligned} & m^d \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d(\frac{\mathbf{k}}{m}, \frac{\mathbf{k}+1}{m})} \int f(\mathbf{y}) d\mathbf{y} P_{\mathbf{k},m-1}^2(\mathbf{x}) \\ &= f(\mathbf{x}) \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k},m-1}^2(\mathbf{x}) + \frac{1}{m} \sum_{i \in [d]} \mathcal{O} \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} |k_i - mx_i| P_{\mathbf{k},m-1}^2(\mathbf{x}) \right) + \mathcal{O}(m^{-1}). \end{aligned} \quad (21)$$

The Cauchy-Schwarz inequality yields

$$\begin{aligned} & \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} |k_i - mx_i| P_{\mathbf{k},m-1}^2(\mathbf{x}) \\ & \leq \sqrt{\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} |k_i - mx_i|^2 P_{\mathbf{k},m-1}(\mathbf{x})} \sqrt{\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k},m-1}^3(\mathbf{x})}. \end{aligned} \quad (22)$$

By putting (20), (21) and (22) together, we get the conclusion.

6.4 Proof of Theorem 2

By Lemma 2, Lemma 5, and (26) in Lemma 7, we have

$$\begin{aligned} \mathbb{V}\text{ar}(\tilde{f}_{n,m}(\mathbf{x})) &= \frac{m^{(d+|\mathcal{J}|)/2}}{n} f(\mathbf{x}) \left\{ \psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) \prod_{i \in \mathcal{J}} e^{-2\lambda_i} I_0(2\lambda_i) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + \mathcal{O}_{\mathbf{x}_{[d] \setminus \mathcal{J}}}(\mathbf{1}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}} \right\} \\ &+ \frac{m^{d-1}}{n} \sum_{i \in [d]} \left\{ \mathcal{O}_{\lambda_{\mathcal{J}}}(1 + \mathbb{1}_{\{i \in \mathcal{J}\}} m^{1/2}) \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-(d-|\mathcal{J}|)/2}) + \mathcal{O}(1) \right\}. \end{aligned}$$

Using the Taylor expansion $f(\mathbf{x}) = f(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1})$, we get

$$\begin{aligned} \mathbb{V}\text{ar}(\tilde{f}_{n,m}(\mathbf{x})) &= \frac{m^{(d+|\mathcal{J}|)/2}}{n} \left(f(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) \right) \\ &\cdot \left\{ \psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) \prod_{i \in \mathcal{J}} e^{-2\lambda_i} I_0(2\lambda_i) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + \mathcal{O}_{\mathbf{x}_{[d] \setminus \mathcal{J}}}(\mathbf{1}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}} \right\} \\ &+ \frac{m^{(d+|\mathcal{J}|)/2}}{n} \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-1} + \mathbb{1}_{\{\mathcal{J} \neq [d]\}} m^{-1/2}). \end{aligned}$$

The conclusion follows.

7 Proof of the results for the c.d.f. estimator $F_{n,m}^*$

7.1 Proof of Lemma 3

Take $\delta_n \searrow 0$ slow enough as $n \rightarrow \infty$ that the contribution coming from points \mathbf{k}/m outside the bulk of the multinomial distribution (i.e., $\|\mathbf{k}/m - \mathbf{x}\|_1 > \delta_n$) is negligible, exactly as we did in (16). Then, for any \mathbf{k} such

that $\|\mathbf{k}/m - \mathbf{x}\|_1 \leq \delta_n$, we can use a Taylor expansion and our assumption that F is three-times continuously differentiable to obtain

$$\begin{aligned} F(\mathbf{k}/m) &= F(\mathbf{x}) + \sum_{i \in [d]} (k_i/m - x_i) \frac{\partial}{\partial x_i} F(\mathbf{x}) + \frac{1}{2} \sum_{i,j \in [d]} (k_i/m - x_i)(k_j/m - x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \\ &\quad + \frac{1}{6} \sum_{i,j,\ell \in [d]} (k_i/m - x_i)(k_j/m - x_j)(k_\ell/m - x_\ell) \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) + o(\|\mathbf{k}/m - \mathbf{x}\|_1^3). \end{aligned} \quad (23)$$

If we multiply the last expression by $P_{\mathbf{k},m}(\mathbf{x})$, sum over $\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d$, and then take the expectation on both sides, we get

$$\begin{aligned} \mathbb{E}[F_{n,m}^*(\mathbf{x})] &= F(\mathbf{x}) + \sum_{i \in [d]} \mathbb{E}[\xi_i/m - x_i] \frac{\partial}{\partial x_i} F(\mathbf{x}) + \frac{1}{2} \sum_{i,j \in [d]} \mathbb{E}[(\xi_i/m - x_i)(\xi_j/m - x_j)] \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \\ &\quad + \frac{1}{6} \sum_{i,j,\ell \in [d]} \mathbb{E}[(\xi_i/m - x_i)(\xi_j/m - x_j)(\xi_\ell/m - x_\ell)] \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \\ &\quad + \sum_{i,j,\ell \in [d]} o(\mathbb{E}[|\xi_i/m - x_i| |\xi_j/m - x_j| |\xi_\ell/m - x_\ell|]), \end{aligned}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_d) \sim \text{Multinomial}(m, \mathbf{x})$. From the multinomial joint central moments in Lemma 7, we get

$$\begin{aligned} \mathbb{E}[F_{n,m}^*(\mathbf{x})] &= F(\mathbf{x}) + \frac{1}{2m} \sum_{i,j \in [d]} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \\ &\quad + \frac{1}{6m^2} \sum_{i,j,\ell \in [d]} \left(2x_i x_j x_\ell - \mathbb{1}_{\{i=j\}} x_i x_\ell - \mathbb{1}_{\{j=\ell\}} x_i x_j \right. \\ &\quad \left. - \mathbb{1}_{\{i=\ell\}} x_j x_\ell + \mathbb{1}_{\{i=j=\ell\}} x_i \right) \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \\ &\quad + \frac{1}{m^3} \sum_{i,j,\ell \in [d]} o(\mathbb{E}[|\xi_i - m x_i| |\xi_j - m x_j| |\xi_\ell - m x_\ell|]). \end{aligned}$$

We apply Holder's inequality on the error term to get the conclusion.

7.2 Proof of Theorem 3

Take \mathbf{x} as in the statement of the theorem. For all $i, j, \ell \in [d]$, note that

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) &= \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} + \sum_{\ell \in \mathcal{J}} \frac{\lambda_\ell}{m} \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} (1 + o_{\lambda_g}(1)), \\ \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) &= \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} + o_{\lambda_g}(1). \end{aligned}$$

By Lemma 3 and the second and fourth moments in Lemma 7, we deduce

$$\begin{aligned} \text{Bias}(F_{n,m}^*(\mathbf{x})) &= \frac{1}{m} \sum_{i,j \in [d] \setminus \mathcal{J}} \frac{1}{2} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} \\ &\quad + \frac{1}{m^2} \left\{ \frac{1}{6} \sum_{i,j,\ell \in [d] \setminus \mathcal{J}} \left(2x_i x_j x_\ell - \mathbb{1}_{\{i=j\}} x_i x_\ell - \mathbb{1}_{\{j=\ell\}} x_i x_j \right. \right. \\ &\quad \left. \left. - \mathbb{1}_{\{i=\ell\}} x_j x_\ell + \mathbb{1}_{\{i=j=\ell\}} x_i \right) \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} \right. \\ &\quad \left. + \sum_{i,j \in [d] \setminus \mathcal{J}} \sum_{\ell \in \mathcal{J}} \frac{\lambda_\ell}{2} (x_i \mathbb{1}_{\{i=j\}} - x_i x_j) \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} \right. \\ &\quad \left. + \frac{1}{2} \sum_{i \in \mathcal{J}} \sum_{j \in [d] \setminus \mathcal{J}} (\lambda_i \mathbb{1}_{\{i=j\}} - 2\lambda_i x_j) \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x}) \Big|_{\mathbf{x}_g = \mathbf{0}} \right\} \\ &\quad + o_{\lambda_g}(m^{-3}) + o_{\lambda_g}(m^{-3/2}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \end{aligned}$$

This ends the proof.

7.3 Proof of Lemma 4

To estimate the variance of $F_{n,m}^*(\mathbf{x})$, note that

$$F_{n,m}^*(\mathbf{x}) - \mathbb{E}[F_{n,m}^*(\mathbf{x})] = \frac{1}{n} \sum_{i=1}^n \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d} (\mathbb{1}_{(-\infty, \frac{\mathbf{k}}{m}]}(\mathbf{X}_i) - F(\mathbf{k}/m)) P_{\mathbf{k},m}(\mathbf{x}).$$

By the independence of the observations $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, we get

$$\mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) = n^{-1} \left\{ \sum_{\mathbf{k}, \ell \in \mathbb{N}_0^d \cap m\mathcal{S}_d} F((\mathbf{k} \wedge \ell)/m) P_{\mathbf{k},m}(\mathbf{x}) P_{\ell,m}(\mathbf{x}) - \mathbb{E}[F_{n,m}^*(\mathbf{x})]^2 \right\}.$$

Using the expansion in (23) and Lemma 3, the above is

$$= n^{-1} \left\{ \begin{aligned} & F(\mathbf{x}) - F^2(\mathbf{x}) + \sum_{i,j \in [d]} \mathcal{O}(m^{-1}(x_i \mathbb{1}_{\{i=j\}} - x_i x_j)) + \mathcal{O}(m^{-2}) \\ & + \sum_{i \in [d]} \frac{\partial}{\partial x_i} F(\mathbf{x}) \sum_{\mathbf{k}, \ell \in \mathbb{N}_0^d \cap m\mathcal{S}_d} ((k_i \wedge \ell_i)/m - x_i) P_{\mathbf{k},m}(\mathbf{x}) P_{\ell,m}(\mathbf{x}) \\ & + \sum_{i \in [d]} \mathcal{O}(m^{-1} \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap m\mathcal{S}_d} |k_i - m x_i|^2 P_{\mathbf{k},m}(\mathbf{x})) \end{aligned} \right\}.$$

By the second moment expression in Lemma 7, we get the conclusion.

7.4 Proof of Theorem 4

By Lemma 4, Lemma 6 and (26) in Lemma 7, we have

$$\begin{aligned} \mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) &= n^{-1} m^{-1} \sum_{i \in [d]} \frac{\partial}{\partial x_i} F(\mathbf{x}) \left\{ \begin{aligned} & -\lambda_i e^{-2\lambda_i} (I_0(2\lambda_i) + I_1(2\lambda_i)) \mathbb{1}_{\{i \in \mathcal{J}\}} \\ & -m^{1/2} \sqrt{\pi^{-1} x_i (1-x_i)} \mathbb{1}_{\{i \in [d] \setminus \mathcal{J}\}} \end{aligned} \right\} \\ &+ n^{-1} F(\mathbf{x})(1-F(\mathbf{x})) + \mathcal{O}_{\lambda_{\mathcal{J}}}(n^{-1} m^{-2}) \\ &+ \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(n^{-1} m^{-1/2}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}. \end{aligned}$$

Now, using the fact that

$$\begin{aligned} F(\mathbf{x}) &= \underbrace{F(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}}}_{=0 \text{ when } \mathcal{J} \neq \emptyset} + \sum_{i \in \mathcal{J}} \frac{\lambda_i}{m} \frac{\partial}{\partial x_i} F(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-2}), \\ \frac{\partial}{\partial x_i} F(\mathbf{x}) &= \frac{\partial}{\partial x_i} F(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \sum_{j \in \mathcal{J}} \frac{\lambda_j}{m} \frac{\partial^2}{\partial x_i \partial x_j} F(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-2}), \end{aligned}$$

we get

$$\begin{aligned} \mathbb{V}\text{ar}(F_{n,m}^*(\mathbf{x})) &= n^{-1} m^{-1} \sum_{i \in [d]} \frac{\partial}{\partial x_i} F(\mathbf{x})|_{\mathbf{x}_{\mathcal{J}}=\mathbf{0}} \left\{ \begin{aligned} & \lambda_i (1 - e^{-2\lambda_i} (I_0(2\lambda_i) + I_1(2\lambda_i))) \mathbb{1}_{\{i \in \mathcal{J}\}} \\ & -m^{1/2} \sqrt{\pi^{-1} x_i (1-x_i)} \mathbb{1}_{\{i \in [d] \setminus \mathcal{J}\}} \end{aligned} \right\} \\ &+ n^{-1} F(\mathbf{x})(1-F(\mathbf{x})) \mathbb{1}_{\{\mathcal{J}=\emptyset\}} + \mathcal{O}_{\lambda_{\mathcal{J}}}(n^{-1} m^{-2}) + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(n^{-1} m^{-1/2}) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}. \end{aligned}$$

This ends the proof.

A Technical lemmas

The first lemma is a generalization of Lemma 3 in Ouimet (2021a). It is used in the proof of Theorem 2.

Lemma 5. For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i \geq 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $m \rightarrow \infty$,

$$m^{(d-|\mathcal{J}|)/2} \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k}, m-1}^2(\mathbf{x}) = \psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) \prod_{i \in \mathcal{J}} e^{-2\lambda_i} I_0(2\lambda_i) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(1) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \quad (24)$$

$$\sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k}, m-1}^3(\mathbf{x}) = \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-(d-|\mathcal{J}|)}), \quad (25)$$

where I_0 is defined in (3) and $\psi_{[d] \setminus \mathcal{J}}$ is defined in (4).

The crucial tool for the proof is a local limit theorem for the multinomial distribution in Arenbaev (1976) that combines the multivariate Poisson approximation and the multivariate normal approximation, depending on which components of $\mathbf{x} \in \mathcal{S}_d$ are close to the boundary.

Proof of Lemma 5. Denote $\mathbf{x}_A := (x_i)_{i \in A}$ for any subset of indices $A \subseteq [d]$, and let

$$\phi_{\emptyset}(\mathbf{x}) := 1, \quad \phi_A(\mathbf{x}) := \frac{\exp\left(-\frac{1}{2} \boldsymbol{\delta}_{\mathbf{x}_A}^{\top} (\text{diag}(\mathbf{x}_A) - \mathbf{x}_A \mathbf{x}_A^{\top})^{-1} \boldsymbol{\delta}_{\mathbf{x}_A}\right)}{\sqrt{(2\pi)^{|A|}} (1 - \sum_{i \in A} x_i) \prod_{i \in A} x_i},$$

with

$$\boldsymbol{\delta}_{\mathbf{x}_A} := \left(\frac{k_i - (m-1)x_i}{\sqrt{m-1}} \right)_{i \in A}.$$

By successively applying the triangle inequality, the identity $a^2 - b^2 = (a+b)(a-b)$ and the total variation bound from Theorem 3 in Arenbaev (1976), we have, as $m \rightarrow \infty$,

$$\begin{aligned} & \left| \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k}, m-1}^2(\mathbf{x}) - \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} \left(\prod_{i \in \mathcal{J}} \frac{((m-1)x_i)^{k_i}}{k_i!} e^{-(m-1)x_i} \right)^2 \left(\frac{\phi_{[d] \setminus \mathcal{J}}(\mathbf{x})}{(m-1)^{(d-|\mathcal{J}|)/2}} \right)^2 \right| \\ & \leq 2.1 \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} \left| P_{\mathbf{k}, m-1}(\mathbf{x}) - \left(\prod_{i \in \mathcal{J}} \frac{((m-1)x_i)^{k_i}}{k_i!} e^{-(m-1)x_i} \right) \frac{\phi_{[d] \setminus \mathcal{J}}(\mathbf{x})}{(m-1)^{(d-|\mathcal{J}|)/2}} \right| \\ & = 2.1 \|\mathbf{x}_{\mathcal{J}}\|_1 \left\{ \sqrt{\frac{2}{\pi e}} + \mathcal{O}(1 \wedge (m\|\mathbf{x}_{\mathcal{J}}\|_1)^{-1/2}) \right\} + \mathcal{O} \left(\frac{\|\mathbf{x}_{[d] \setminus \mathcal{J}}^{-1/2}\|_1}{\sqrt{m}} + \frac{\|\mathbf{x}_{[d] \setminus \mathcal{J}}^{-1}\|_{\infty}}{m} \right), \end{aligned}$$

with the conventions $\mathbf{x}_A^{-1/2} := (x_i^{-1/2})_{i \in A}$, $\mathbf{x}_A^{-1} := (x_i^{-1})_{i \in A}$ and $\|\mathbf{x}_{\emptyset}^{-1/2}\|_1 = \|\mathbf{x}_{\emptyset}^{-1}\|_{\infty} = 0$. By the assumption on \mathbf{x} in the statement of the lemma, we get

$$\begin{aligned} m^{(d-|\mathcal{J}|)/2} \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} P_{\mathbf{k}, m-1}^2(\mathbf{x}) &= \sum_{\mathbf{k} \in \mathbb{N}_0^d \cap (m-1)\mathcal{S}_d} \left(\prod_{i \in \mathcal{J}} \frac{((m-1)x_i)^{k_i}}{k_i!} e^{-(m-1)x_i} \right)^2 \frac{\phi_{[d] \setminus \mathcal{J}}^2(\mathbf{x})}{(m-1)^{(d-|\mathcal{J}|)/2}} \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-1} + m^{-1/2} \mathbb{1}_{\{\mathcal{J} \neq [d]\}}). \end{aligned}$$

By Equation (39) in Ouimet (2021a), the above is

$$\begin{aligned} &= \prod_{i \in \mathcal{J}} e^{-2\lambda_i} (1 + \mathcal{O}(m^{-1})) \left(\sum_{k=0}^{\infty} \frac{\lambda_i^{2k}}{(k!)^2} + \mathcal{O}_{\lambda_{\mathcal{J}}}(e^{-\frac{m}{100d^2}}) \right) (\psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) + \mathcal{O}_{\lambda_{[d] \setminus \mathcal{J}}}(1)) \\ &\quad + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(m^{-1} + m^{-1/2} \mathbb{1}_{\{\mathcal{J} \neq [d]\}}) \\ &= \psi_{[d] \setminus \mathcal{J}}(\mathbf{x}) \prod_{i \in \mathcal{J}} e^{-2\lambda_i} I_0(2\lambda_i) + \mathcal{O}_{\lambda_{\mathcal{J}}}(m^{-1}) + \mathcal{O}_{\lambda_{\mathcal{J}}, \mathbf{x}_{[d] \setminus \mathcal{J}}}(1) \mathbb{1}_{\{\mathcal{J} \neq [d]\}}, \end{aligned}$$

which proves (24). Equation (25) is direct corollary to Equation (40) in Ouimet (2021a). This ends the proof. \square

The third lemma generalizes Lemma 4 in Ouimet (2021a). It is used in the proof of Theorem 4.

Lemma 6. For any $\mathbf{x} \in \mathcal{S}_d$ such that $x_i = \lambda_i/m$ for all $i \in \mathcal{J} \subseteq [d]$ ($\lambda_i > 0$ is fixed $\forall i \in \mathcal{J}$) and $x_i \in (0, 1)$ independent of m for all $i \in [d] \setminus \mathcal{J}$, we have, as $m \rightarrow \infty$,

$$\sum_{\mathbf{k}, \ell \in \mathbb{N}_0^d \cap m\mathcal{S}_d} ((k_p \wedge \ell_p)/m - x_p) P_{\mathbf{k}, m}(\mathbf{x}) P_{\ell, m}(\mathbf{x}) = \begin{cases} -m^{-1} \lambda_p e^{-2\lambda_p} (I_0(2\lambda_p) + I_1(2\lambda_p)) + \mathcal{O}_{\lambda_p}(m^{-2}), & \text{if } p \in \mathcal{J}, \\ -m^{-1/2} \sqrt{\pi^{-1} x_p (1 - x_p)} + \mathcal{O}_{x_p}(m^{-1/2}), & \text{if } p \in [d] \setminus \mathcal{J}, \end{cases}$$

where I_j , $j \in \{0, 1\}$, is defined in (3).

Proof of Lemma 6. We know that the marginal distributions of the multinomial are binomial. Therefore, when $p \in \mathcal{J}$, the result follows from Lemma 3 (B) and Lemma 7 (B) in Leblanc (2012b). When $p \in [d] \setminus \mathcal{J}$, the proof is given in Lemma 4 of Ouimet (2021a). \square

Next, we write the joint central moments (of order two, three and four) for the multinomial distribution. These moments were derived, for example, by Ouimet (2021b).

Lemma 7 (Central moments). Let $\mathbf{x} \in \mathcal{S}_d$. If $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_d) \sim \text{Multinomial}(m, \mathbf{x})$, then, for all $i, j, \ell, p \in [d]$,

$$\mathbb{E}[(\xi_i - m x_i)(\xi_j - m x_j)] = m(x_i \mathbb{1}_{\{i=j\}} - x_i x_j), \quad (26)$$

$$\mathbb{E}[(\xi_i - m x_i)(\xi_j - m x_j)(\xi_\ell - m x_\ell)] = m \begin{pmatrix} 2x_i x_j x_\ell - \mathbb{1}_{\{i=j\}} x_i x_\ell - \mathbb{1}_{\{j=\ell\}} x_i x_j \\ -\mathbb{1}_{\{i=\ell\}} x_j x_\ell + \mathbb{1}_{\{i=j=\ell\}} x_i \end{pmatrix},$$

$$\mathbb{E}[(\xi_i - m x_i)(\xi_j - m x_j)(\xi_\ell - m x_\ell)(\xi_p - m x_p)] = \mathcal{O}(m^2), \quad \text{as } m \rightarrow \infty.$$

Conflict of the interest: The author declares no conflict of interest.

Funding information: The author acknowledges (past) support of a postdoctoral fellowship from the NSERC (PDF) and a postdoctoral fellowship supplement from the FRQNT (B3X). The author is currently supported by a CRM-Simons postdoctoral fellowship from the Centre de recherches mathématiques and the Simons Foundation.

Acknowledgments: We thank the anonymous referees for their comments.

References

- Arenbaev, N. K. (1976). Asymptotic behavior of the multinomial distribution. *Teor. Veroyatnost. i Primenen.*, 21(4):826–831. MR0478288.
- Babu, G. J., Canty, A. J., and Chaubey, Y. P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference*, 105(2):377–392. MR1910059.
- Babu, G. J. and Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.*, 76(9):959–969. MR2270097.
- Belalia, M. (2016). On the asymptotic properties of the Bernstein estimator of the multivariate distribution function. *Statist. Probab. Lett.*, 110:249–256. MR3474765.
- Bouezmarni, T. and Rolin, J.-M. (2007). Bernstein estimator for unbounded density function. *J. Nonparametr. Stat.*, 19(3):145–161. MR2351744.
- Bouezmarni, T. and Rombouts, J. V. K. (2010). Nonparametric density estimation for multivariate bounded data. *J. Statist. Plann. Inference*, 140(1):139–152. MR2568128.
- Bouezmarni, T. and Scaillet, O. (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econom. Theor.*, 21(2):390–412. MR2179543.

- Gawronski, W. and Stadtmüller, U. (1980). On density estimation by means of Poisson's distribution. *Scand. J. Statist.*, 7(2):90–94. MR0574548.
- Gawronski, W. and Stadtmüller, U. (1981). Smoothing histograms by means of lattice and continuous distributions. *Metrika*, 28(3):155–164. MR0638651.
- Igarashi, G. and Kakizawa, Y. (2014). On improving convergence rate of Bernstein polynomial density estimator. *J. Nonparametr. Stat.*, 26(1):61–84. MR3174309.
- Igarashi, G. and Kakizawa, Y. (2020). Multiplicative bias correction for asymmetric kernel density estimators revisited. *Comput. Statist. Data Anal.*, 141:40–61. MR3979322.
- Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.*, 16(5):709–729. MR2068610.
- Kakizawa, Y. (2022). Multivariate elliptical-based Birnbaum-Saunders kernel density estimation for nonnegative data. *J. Multivariate Anal.*, 187:Paper No. 104834, 20 pp. MR4322320.
- Leblanc, A. (2009). Chung-Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.*, 21(2):133–142. MR2488150.
- Leblanc, A. (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *J. Nonparametr. Stat.*, 22(3-4):459–475. MR2662607.
- Leblanc, A. (2012a). On estimating distribution functions using Bernstein polynomials. *Ann. Inst. Statist. Math.*, 64(5):919–943. MR2960952.
- Leblanc, A. (2012b). On the boundary properties of Bernstein polynomial estimators of density and distribution functions. *J. Statist. Plann. Inference*, 142(10):2762–2778. MR2925964.
- Liu, B. and Ghosh, S. K. (2020). On empirical estimation of mode based on weakly dependent samples. *Comput. Statist. Data Anal.*, 152:107046, 21 pp. MR4130895.
- Lu, D. and Wang, L. (2021). On the rates of asymptotic normality for Bernstein polynomial estimators in a triangular array. *Methodol. Comput. Appl. Probab.*, 23(4):1519–1536. MR4335173.
- Lu, L. (2015). On the uniform consistency of the Bernstein density estimator. *Statist. Probab. Lett.*, 107:52–61. MR3412755.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York. MR1789474.
- Ouimet, F. (2021a). Asymptotic properties of Bernstein estimators on the simplex. *J. Multivariate Anal.*, 185:104784, 20 pp. MR4287788.
- Ouimet, F. (2021b). General formulas for the central and non-central moments of the multinomial distribution. *Stats*, 4(1):18–27. doi:10.3390/stats4010002.
- Ouimet, F. and Tolosana-Delgado, R. (2022). Asymptotic properties of Dirichlet kernel density estimators. *J. Multivariate Anal.*, 187:104832, 25 pp. MR4319409.
- Scott, D. W. (2015). *Multivariate Density Estimation*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition. MR3329609.
- Tenbusch, A. (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika*, 41(3-4):233–253. MR1293514.
- Vitale, R. A. (1975). Bernstein Polynomial Approach to Density Function Estimation. In *Statistical Inference and Related Topics*, pages 87–99. Academic Press, New York. MR0397977.
- Wang, L. and Lu, D. (2022). On the rates of asymptotic normality for Bernstein density estimators in a triangular array. *J. Math. Anal. Appl.*, 511(1):126063, 19 pp. MR4374864.