# A Bernstein polynomial approach for the estimation of cumulative distribution functions in the presence of missing data

Rihab Gharbi[a], Wissem Jedidi[b], Salah Khardani[a], Frédéric Ouimet[c,*]

[a] *Département de Mathématiques, Université de Tunis El Manar, Tunisia*
[b] *Department of Statistics and Operations Research, College of Science, King Saud University, Saudi Arabia*
[c] *Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Canada*

**Abstract**

We study nonparametric estimation of univariate cumulative distribution functions (CDFs) pertaining to data missing at random. The proposed estimators smooth the inverse probability weighted (IPW) empirical CDF with the Bernstein operator, yielding monotone, $[0, 1]$-valued curves that automatically adapt to bounded supports. We analyze two versions: a pseudo estimator that uses known propensities and a feasible estimator that uses propensities estimated nonparametrically from discrete auxiliary variables, the latter scenario being much more common in practice. For both, we derive pointwise bias and variance expansions, establish the optimal polynomial degree $m$ with respect to the mean integrated squared error, and prove the asymptotic normality. A key finding is that the feasible estimator has a smaller variance than the pseudo estimator by an explicit nonnegative correction term. We also develop an efficient degree selection procedure via least-squares cross-validation. Monte Carlo experiments demonstrate that, for moderate to large sample sizes, the Bernstein-smoothed feasible estimator outperforms both its unsmoothed counterpart and an integrated version of the IPW kernel density estimator proposed by Dubnicka (2009) in the same context. A real-data application to fasting plasma glucose from the 2017–2018 NHANES survey illustrates the method in a practical setting. All code needed to reproduce our analyses is readily accessible on GitHub.

*Keywords:* Asymptotics, Bernstein estimator, cumulative distribution function estimation, inverse probability weighting, missing at random, missing data, nonparametric estimation.
*2020 MSC:* Primary: 62G05; Secondary: 62E20, 62G08, 62G20

## 1. Introduction

The cumulative distribution function (CDF) and its inverse (the quantile function) are fundamental objects in probability and statistics, providing a complete description of a random variable's distribution. They offer insights into the entire distribution, including tail behavior, that summary measures like the mean or variance cannot capture. Accurate estimation of CDFs is crucial across numerous fields, for example, in survival analysis (where the CDF relates to survival probabilities), reliability engineering (where the CDF gives failure probabilities and reliability over time), economics (where the CDF characterizes income, wealth, or return distributions), and risk management (where the loss CDF underpins tail-risk metrics such as Value-at-Risk), where one often needs to understand the full distribution rather than just a single parameter.

In practice, statistical analysis is frequently complicated by incomplete data. Missing observations are ubiquitous in large studies, clinical trials, and surveys. If data are missing in a nontrivial fashion, standard nonparametric estimators such as the empirical CDF computed from observed cases can be biased for the true distribution unless the data are missing completely at random, that is, the missingness mechanism is independent of the data values (Little & Rubin, 2002, p.12). A more plausible assumption in many contexts is missing at random (MAR), which posits that

*Corresponding author. Email address: frederic.ouimet2@uqtr.ca
*Email addresses:* rihab.gharbi@fst.utm.tn (Rihab Gharbi), wjedidi@ksu.edu.sa (Wissem Jedidi), salah.khardani@fst.utm.tn (Salah Khardani), frederic.ouimet2@uqtr.ca (Frédéric Ouimet)

the probability an observation is missing may depend on fully observed auxiliary variables but not on the value of the missing observation itself (Rubin, 1976). Under MAR, one can obtain unbiased estimates by appropriately reweighting or imputing the observed data. A prominent approach is inverse probability weighting (IPW), inspired by the Horvitz–Thompson estimator from survey sampling (Horvitz & Thompson, 1952). In IPW, each observed case is weighted by the inverse of its observation probability (called propensity score), often estimated from covariates (Rosenbaum & Rubin, 1983), yielding a weighted empirical CDF (sometimes called a pseudo-empirical CDF). When propensity scores are known or consistently estimated, this IPW empirical CDF is asymptotically unbiased for the full-population CDF.

Alternative nonparametric strategies for MAR data include imputation-based methods and hybrid approaches. For instance, Cheng & Chu (1996) proposed a kernel regression imputation to consistently estimate the CDF and quantiles, proving strong uniform consistency and asymptotic normality. Hybrid strategies, in particular augmented IPW estimators, have been developed to improve efficiency by combining weighting with outcome imputation. Wang & Qin (2010) introduced an augmented IPW empirical-likelihood approach for CDFs with missing responses that delivers asymptotic Gaussian limits for the process. These methods underscore that, under MAR assumptions, one can recover distributional information using either weighting or imputation, or both, without fully specifying outcome models. Furthermore, these estimation strategies have also been adapted to the more complex case of nonignorable missing data (NMAR). Related work by Ding & Tang (2018) considered IPW, regression imputation, and an augmented approach for distribution and quantile estimation under a semiparametric NMAR model, finding that all three can reach the same asymptotic variance under suitable conditions.

However, the IPW-based empirical CDF, like the ordinary empirical CDF, is a step function. When the underlying distribution is continuous, it is often desirable to smooth such step-function estimates to improve efficiency, reduce mean squared error (MSE), and produce a smooth, continuous CDF estimate. Smoothing can yield substantial gains; some smooth CDF estimators have been shown to asymptotically outperform the raw empirical CDF in mean integrated squared error (MISE). Kernel smoothing is a popular approach in the density estimation context; see, e.g., Rosenblatt (1956); Parzen (1962), or Silverman (1986); Wand & Jones (1995) for book treatments. This approach extends naturally to CDFs by integrating the kernel density estimate, resulting in the classical smooth CDF estimator introduced by Nadaraya (1964), which replaces the indicator steps of the empirical CDF with integrated kernels. An alternative strategy involves direct regression smoothing of the empirical CDF. For example, Cheng & Peng (2002) proposed a local linear estimator that can achieve a smaller asymptotic MISE than the integrated kernel estimator, although it does not guarantee monotonicity. These kernel methods have been adapted to handle missing-data scenarios; for example, Dubnicka (2009) developed an IPW kernel density estimator (KDE) under MAR. For CDFs under MAR, a natural idea is to smooth the IPW variant using similar techniques. For instance, one of the competitors considered in our Monte Carlo study (Section 4) is an integrated version of Dubnicka's estimator.

Smoothing the CDF can improve pointwise variance and yield well-behaved estimates, but one must address the well-known boundary bias that traditional kernel estimators exhibit on bounded supports; see, e.g., Zhang *et al.* (2020). If the support of the distribution is [0,1], integrated-KDE CDF estimators tend to incur bias near 0 and 1 due to a spill-over effect created by the fixed kernel. To mitigate this, various boundary-correction techniques have been proposed, including reflecting the data at boundaries or using boundary kernels tailored for the edges (Tenreiro, 2013). A user-friendly and highly effective strategy employs asymmetric kernels, which inherently respect support limits and ensure monotonicity by construction; see, e.g., Lafaye de Micheaux & Ouimet (2021); Mansouri *et al.* (2023). Alternatively, monotonicity constraints can be imposed on smoothing techniques that do not naturally guarantee it. For instance, Xue & Wang (2010) proposed a constrained polynomial spline approach that smooths the empirical CDF while enforcing monotonicity, thereby guaranteeing a valid CDF estimate.

In this paper, the focus is on CDFs supported on the unit interval $[0, 1]$ (more general bounded

supports can often be transformed to $[0,1]$). For such cases, Bernstein polynomials offer an elegant and effective smoothing technique. Bernstein polynomials (Bernstein, 1912) were originally introduced as a constructive proof of the Weierstrass approximation theorem (Weierstraß, 1885), which asserts the existence of uniform polynomial approximations for continuous functions over a closed interval. When used for smoothing an empirical CDF, Bernstein operators (see, e.g., Bustamante, 2017) produce estimates that automatically respect the boundaries, mapping $[0,1]$ into $[0,1]$, and are genuine CDFs, since they are monotone non-decreasing and bounded between 0 and 1 by construction. They also enjoy shape-preserving properties, meaning the smoothed curve inherits the qualitative shape constraints of a CDF. The idea of leveraging Bernstein polynomials for nonparametric estimation dates back to Vitale (1975), who studied a Bernstein estimator for density functions on a bounded interval. Babu *et al.* (2002) and Babu & Chaubey (2006) later demonstrated the utility of Bernstein polynomials for CDF estimation, applying them to smooth the empirical CDF and the associated normalized histogram under strongly mixing data. Subsequent work has analyzed the statistical properties of Bernstein estimators in detail. Leblanc (2012a,b) derived higher-order bias and variance expansions for the Bernstein CDF estimator and showed that it has asymptotically negligible boundary bias and can be more efficient than the (unsmoothed) empirical CDF. There have also been several developments and extensions of Bernstein-type smoothers. For example, Jmaei *et al.* (2017); Slaoui (2022) develop and analyze a recursive approach to Bernstein CDF estimation, including a Robbins–Monro style estimator and moderate deviation theory. Erdoğan *et al.* (2019) introduced an alternative CDF estimator using rational Bernstein polynomials, which generalizes the classical Bernstein approach to improve flexibility. For distributions supported on $[0,\infty)$, Hanebeck & Klar (2021) proposed using Szasz–Mirakyan positive linear operators, an analog of Bernstein operators for $[0,\infty)$, to construct smooth CDF estimators for lifetime or survival data, and they showed that this method outperforms the empirical CDF in MSE and MISE as well. Moreover, Bernstein polynomials have been adapted to other complex settings. For example, Belalia *et al.* (2017) estimated conditional CDFs by smoothing regression estimators, and Khardani (2024) extended Bernstein-polynomial CDF and quantile estimation to right-censored data, highlighting their broad applicability. All of these approaches share a common goal of producing a smoother and more efficient estimate of the CDF while enforcing the shape constraints that are intrinsic to CDFs.

Despite the rich literature on handling missing data and on smooth CDF estimation separately, relatively little attention has been given to smoothing CDF estimators in the presence of missing data. This work aims to fill that gap by developing and analyzing a nonparametric CDF estimator for MAR data that marries the IPW principle with Bernstein polynomial smoothing. In other words, the IPW-adjusted empirical CDF, which corrects bias due to MAR missingness, is smoothed using the Bernstein operator (see (2.1) below) to obtain a smooth and shape-constrained estimate. This approach inherits the double benefit of unbiasedness under MAR, from the weighting, and reduced variance plus automatic boundary adaptation, from the Bernstein smoothing.

The remainder of the paper is organized as follows. Section 2 introduces the statistical framework, defines the Bernstein operator, formulates the MAR setting, describes propensity-score estimation from discrete covariates, and constructs the Bernstein-smoothed IPW estimators $\widetilde{F}_{n,m}$ (when propensities are known) and $\widehat{F}_{n,m}$ (when propensities are estimated). Section 3 presents the results: pointwise bias and variance, optimal choices of $m$ based on MSE and MISE, and asymptotic normality. Results are given first for the pseudo estimator with known propensities (Section 3.1) and then for the feasible estimator with estimated propensities (Section 3.2), where we quantify the variance reduction from estimating the propensity. Section 4 reports a Monte Carlo study comparing the smoothed and unsmoothed estimators across sample sizes, including the integrated-IPW KDE of Dubnicka (2009) as a benchmark. Section 5 applies the feasible Bernstein-smoothed estimator to a fasting plasma glucose dataset. Section 6 summarizes the contributions and outlines directions for future research. All proofs are collected in Section 7. For reproducibility, Appendix A links to the R code used to generate the figures, the simulation results, and the real-data application, and Appendix B provides a list of acronyms used throughout.

## 2. The setup

For any continuous function $\varphi : [0,1] \mapsto \mathbb{R}$, the Bernstein operator,

$$\mathcal{B}_m(\varphi)(y) = \sum_{k=0}^{m} \varphi(k/m) \binom{m}{k} y^k (1-y)^{m-k}, \quad y \in [0,1], \quad m \in \mathbb{N}, \tag{2.1}$$

defines a sequence of bounded polynomials, $\{\mathcal{B}_m(\varphi)\}_{m\in\mathbb{N}}$, which converges uniformly to $\varphi$. The weight function that smooths out the discrete values of $\varphi$ over the mesh $\{0, 1/m, 2/m, \ldots, 1\}$ corresponds to the probability mass function of a binomial distribution as a function of its success probability parameter $y$. This weight function is denoted, for all $y \in [0,1]$ and $k \in \{0, \ldots, m\}$, by

$$b_{m,k}(y) = \binom{m}{k} y^k (1-y)^{m-k}.$$

Consider the following setting. In the population under study, there are a continuous response variable $Y$ subject to missingness and a discrete auxiliary variable $\boldsymbol{X} = (X_1, \ldots, X_d)$, for some $d \in \mathbb{N}$, which is fully observed. This is a common scenario in practice, for example with categorical demographics in surveys. Take a random sample of $n$ independent and identically distributed (iid) units from the population. For each unit $i \in \{1, \ldots, n\}$ in the sample, denote by $Y_i$ and $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})$ the $i$th observation of the response and auxiliary variables, respectively. Let the Bernoulli random variable

$$\delta_i = \mathbb{1}_{\{Y_i \text{ is observed}\}} = \begin{cases} 1, & \text{if } Y_i \text{ is observed,} \\ 0, & \text{otherwise,} \end{cases}$$

indicate whether $Y_i$ is observed or not. We assume that the distribution of $\delta_i$ given $\boldsymbol{X}_i$ is defined by the *propensity score*:

$$\pi_i = \pi_i(\boldsymbol{X}_i) = \mathsf{P}\left(\delta_i = 1 \mid Y_i, \boldsymbol{X}_i\right) = \mathsf{P}\left(\delta_i = 1 \mid \boldsymbol{X}_i\right).$$

Note that $\delta_i$ is independent of $Y_i$ conditionally on $\boldsymbol{X}_i$.

The goal is to estimate the unknown CDF $F$ of the observations $Y_1, \ldots, Y_n$, which are MAR. We investigate two versions of the estimator based on the knowledge of the propensity scores.

The first version assumes propensities are known, which is the case, for example, when the missingness mechanism is designed. In this case, $F$ can be estimated by the following IPW pseudo-empirical estimator:

$$\widetilde{F}_n(y) = n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)} \mathbb{1}_{\{Y_i \le y\}}. \tag{2.2}$$

The second and more realistic version addresses the fact that, in practice, the propensity scores are usually unknown. The discreteness of $\boldsymbol{X}$ simplifies propensity score estimation for MAR adjustment, allowing us to rely on the following nonparametric estimator:

$$\hat{\pi}_i(\boldsymbol{X}_{1:n}) = \frac{\sum_{j=1}^{n} \delta_j \, \mathbb{1}_{\{\boldsymbol{X}_j = \boldsymbol{X}_i\}}}{\sum_{k=1}^{n} \mathbb{1}_{\{\boldsymbol{X}_k = \boldsymbol{X}_i\}}}. \tag{2.3}$$

For continuous auxiliary data, this could be a Nadaraya–Watson estimator; see Dubnicka (2009, Eq. (4)). The CDF $F$ can then be estimated using the feasible empirical estimator:

$$\widehat{F}_n(y) = n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_i(\boldsymbol{X}_{1:n})} \mathbb{1}_{\{Y_i \le y\}}.$$

The aim of this paper is to show that we can smooth $\widetilde{F}_n$ and $\widehat{F}_n$ using Bernstein polynomials to improve their performance. Applying the Bernstein operator yields an oracle or pseudo estimator

$\widetilde{F}_{n,m}$ when propensities are known, and a feasible estimator $\widehat{F}_{n,m}$ when propensities are estimated nonparametrically from the data. More specifically, define, for all $y \in [0,1]$ and $m \in \mathbb{N}$,

$$\begin{aligned}
\widetilde{F}_{n,m}(y) &= \mathcal{B}_m(\widetilde{F}_n)(y) = \sum_{k=0}^{m} \widetilde{F}_n(k/m) b_{m,k}(y) = n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y), \\
\widehat{F}_{n,m}(y) &= \mathcal{B}_m(\widehat{F}_n)(y) = \sum_{k=0}^{m} \widehat{F}_n(k/m) b_{m,k}(y) = n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}_i(\boldsymbol{X}_{1:n})} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y).
\end{aligned} \tag{2.4}$$

## 3. Results

In this section, we present theoretical properties of the proposed Bernstein estimators. We derive asymptotic expansions for the pointwise bias and variance, determine the optimal polynomial degree $m$ that minimizes the MSE and MISE, and establish asymptotic normality for both the pseudo and feasible estimators defined in (2.4).

*Assumptions.* Beyond the setup presented in Section 2, we assume the following:

(A1) The propensity scores are bounded from below, i.e., $\min_i \min_{\boldsymbol{x}} \pi_i(\boldsymbol{x}) \equiv \pi_{\min} > 0$.

(A2) The distribution of $\boldsymbol{X}$ is supported on finitely many values, so that $\min_{\boldsymbol{x}} p(\boldsymbol{x}) \equiv p_{\min} > 0$.

(A3) The CDF $F \equiv F_{Y_1}$ is twice continuously differentiable on $[0,1]$.

(A4) The CDF $F_{Y_1|\boldsymbol{X}_1}$ is continuously differentiable on $[0,1]$.

*Notation.* Throughout the paper, $f \equiv F'$ denotes the density of $Y$, and similarly $f_{Y_1|\boldsymbol{X}_1} = F'_{Y_1|\boldsymbol{X}_1}$. The degree parameter $m = m(n)$ is always assumed to be a function of the sample size $n$, and further $m \to \infty$ whenever $n \to \infty$. For real sequences $(U_n)$ and $(V_n)$, the notation $U_n = \mathcal{O}(V_n)$ means that $\limsup_{n\to\infty} |U_n/V_n| \le C < \infty$, where the nonnegative constant $C$ may depend on $d$, $\pi_{\min}$, $p_{\min}$, $F$, or $F_{Y_1|\boldsymbol{X}_1}$, but no other variable unless explicitly written as a subscript. Typically, the dependence is on the variable $y$, in which case one writes $U_n = \mathcal{O}_y(V_n)$. In some places, the Vinogradov notation $U_n \ll V_n$ is used to indicate $U_n = \mathcal{O}(V_n)$ more concisely. Similarly, the notation $U_n = \mathrm{o}(V_n)$ means that $\lim_{n\to\infty} |U_n/V_n| = 0$. Subscripts indicate which parameters the convergence rate can depend on. More generally, if $U_n, V_n$ are random variables, we write $U_n = \mathcal{O}_{L^2}(V_n)$ if $\limsup_{n\to\infty} \mathsf{E}[|U_n/V_n|^2] \le C < \infty$ and $U_n = \mathrm{o}_{L^2}(V_n)$ if $\lim_{n\to\infty} \mathsf{E}[|U_n/V_n|^2] = 0$. If the rate depends on a variable such as $y$, we add a subscript.

### *3.1. Asymptotic properties of $\widetilde{F}_{n,m}$*

We begin by analyzing the pseudo estimator $\widetilde{F}_{n,m}$, which utilizes known propensity scores. This serves as a benchmark for the feasible estimator analyzed later.

The following proposition establishes the asymptotics of the pointwise bias. Since the underlying pseudo-empirical estimator $\widetilde{F}_n$ is unbiased for $F$, the bias of the smoothed estimator $\widetilde{F}_{n,m}$ is entirely attributable to the Bernstein smoothing process, confirming the standard bias rate of $\mathcal{O}(m^{-1})$.

**Proposition 1** (Bias)**.** *Suppose that Assumption (A3) holds. Then, uniformly for $y \in [0,1]$, we have, as $n \to \infty$,*

$$\mathsf{Bias}(\widetilde{F}_{n,m}) = \mathsf{E}[\widetilde{F}_{n,m}(y)] - F(y) = m^{-1} B(y) + \mathrm{o}(m^{-1}),$$

*where*

$$B(y) = \frac{1}{2} y(1-y) f'(y). \tag{3.1}$$

Next, we examine the pointwise variance of the pseudo estimator. The expansion reveals two main components: the leading term $n^{-1}\sigma^2(y)$, corresponding to the variance of the unsmoothed IPW empirical CDF, and a negative correction term of order $\mathcal{O}(n^{-1}m^{-1/2})$. This correction term explicitly quantifies the variance reduction achieved by the Bernstein smoothing.

**Proposition 2** (Variance). *Suppose that Assumptions (A1), (A3), and (A4) hold. Then, for any $y \in (0,1)$, we have, as $n \to \infty$,*

$$\mathsf{Var}(\widetilde{F}_{n,m}(y)) = n^{-1}\sigma^2(y) - n^{-1}m^{-1/2}V(y) + \mathrm{o}_y(n^{-1}m^{-1/2}),$$

*where*

$$\sigma^2(y) = \mathsf{E}\left[\frac{F_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)}\right] - \{F(y)\}^2, \qquad V(y) = \sqrt{\frac{y(1-y)}{\pi}}\mathsf{E}\left[\frac{f_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)}\right]. \tag{3.2}$$

The MSE is an immediate consequence of the bias and variance expansions derived above. Furthermore, analyzing the MSE allows us to determine the optimal polynomial degree $m_{\mathrm{opt}}(y)$ that balances the trade-off between the squared bias term (of order $m^{-2}$) and the variance reduction term (of order $n^{-1}m^{-1/2}$).

**Corollary 3** (Mean squared error). *Suppose that Assumptions (A1), (A3), and (A4) hold. Then, for any $y \in (0,1)$, we have, as $n \to \infty$,*

$$\begin{aligned}\mathrm{MSE}[\widetilde{F}_{n,m}(y)] &= \mathsf{Var}(\widetilde{F}_{n,m}(y)) + \{\mathsf{Bias}(\widetilde{F}_{n,m})\}^2 \\ &= n^{-1}\sigma^2(y) - n^{-1}m^{-1/2}V(y) + m^{-2}B^2(y) + \mathrm{o}_y(n^{-1}m^{-1/2}) + \mathrm{o}(m^{-2}).\end{aligned}$$

*In particular, if $B(y)V(y) \neq 0$, the asymptotically optimal choice of $m$, with respect to the* MSE*, is*

$$m_{\mathrm{opt}}(y) = n^{2/3}\left[4B^2(y)/V(y)\right]^{2/3},$$

*in which case, as $n \to \infty$,*

$$\mathrm{MSE}[\widetilde{F}_{m_{\mathrm{opt}},n}(y)] = n^{-1}\sigma^2(y) - n^{-4/3}\left[27V^4(y)/\{256B^2(y)\}\right]^{1/3} + \mathrm{o}_y(n^{-4/3}).$$

To obtain a global measure of accuracy, we integrate the MSE over the support $[0,1]$. This yields the MISE and allows for the determination of a globally optimal degree $m_{\mathrm{opt}}$. This optimal degree achieves a convergence rate of $n^{-4/3}$ for the second-order term, demonstrating a clear improvement over the unsmoothed estimator $\widetilde{F}_n$.

**Corollary 4** (Mean integrated squared error). *Suppose that Assumptions (A1), (A3), and (A4) hold. Then, as $n \to \infty$, we have*

$$\begin{aligned}\mathrm{MISE}[\widetilde{F}_{n,m}] &= \int_0^1 \mathrm{MSE}[\widetilde{F}_{n,m}(y)]\mathrm{d}y \\ &= n^{-1}\int_0^1 \sigma^2(y)\mathrm{d}y - n^{-1}m^{-1/2}\int_0^1 V(y)\mathrm{d}y + m^{-2}\int_0^1 B^2(y)\mathrm{d}y \\ &\quad + \mathrm{o}(n^{-1}m^{-1/2}) + \mathrm{o}(m^{-2}).\end{aligned}$$

*In particular, if $\int_0^1 B^2(y)\mathrm{d}y \times \int_0^1 V(y)\mathrm{d}y \neq 0$, the asymptotically optimal choice of $m$, with respect to the* MISE*, is*

$$m_{\mathrm{opt}} = n^{2/3}\left[\frac{4\int_0^1 B^2(y)\mathrm{d}y}{\int_0^1 V(y)\mathrm{d}y}\right]^{2/3},$$

*in which case, as $n \to \infty$,*

$$\mathrm{MISE}[\widetilde{F}_{m_{\mathrm{opt}},n}] = n^{-1}\int_0^1 \sigma^2(y)\mathrm{d}y - n^{-4/3}\left[\frac{27\{\int_0^1 V(y)\mathrm{d}y\}^4}{256\int_0^1 B^2(y)\mathrm{d}y}\right]^{1/3} + \mathrm{o}(n^{-4/3}).$$

Next, we establish the asymptotic normality of the pseudo estimator, which is essential for constructing confidence intervals. The proof is a straightforward verification of the Lindeberg condition for double arrays.

**Theorem 5** (Asymptotic normality). *Suppose that Assumptions (A1), (A3), and (A4) hold, and assume that $\sigma^2(y) > 0$ in (3.2). Then, for any $y \in (0,1)$ and $\lambda \in (0,\infty)$, we have, as $n \to \infty$,*

$$n^{1/2}\{\widetilde{F}_{n,m}(y) - F(y)\} \xrightarrow{\text{law}} \begin{cases}\mathcal{N}(0,\sigma^2(y)), & \text{if } n^{1/2}m^{-1} \to 0, \\ \mathcal{N}(\lambda B(y),\sigma^2(y)), & \text{if } n^{1/2}m^{-1} \to \lambda.\end{cases}$$

*3.2. Asymptotic properties of $\widehat{F}_{n,m}$*

We now turn to the feasible estimator $\widehat{F}_{n,m}$, which uses estimated propensity scores. We analyze its properties by characterizing its relationship with the pseudo estimator $\widetilde{F}_{n,m}$. We first show that the process of estimating the propensity scores introduces only a negligible amount of additional bias of order $\mathcal{O}(n^{-1})$, which can be positive or negative.

**Proposition 6** (Bias)**.** *Suppose that Assumptions (A1)–(A3) hold. Then, uniformly for $y \in [0,1]$, we have, as $n \to \infty$,*

$$\begin{aligned}\mathsf{Bias}(\widehat{F}_{n,m}(y)) = \mathsf{E}[\widehat{F}_{n,m}(y)] - F(y) &= \mathsf{Bias}(\widetilde{F}_{n,m}(y)) + \mathcal{O}(n^{-1}) \\ &= m^{-1}B(y) + \mathrm{o}(m^{-1}) + \mathcal{O}(n^{-1}),\end{aligned}$$

*where $B$ is defined in* (3.1)*.*

A key finding is that the feasible estimator exhibits a notable variance reduction compared to the pseudo estimator. Proposition 7 reveals that $\widehat{F}_{n,m}$ has a smaller asymptotic variance than $\widetilde{F}_{n,m}$ by an explicit nonnegative correction term $n^{-1}C(y)$. This highlights a beneficial interaction between smoothing and nonparametric propensity estimation in this context. Let us remark that Dubnicka (2009, Eq. (8)) obtained a similar variance reduction in the KDE setting.

**Proposition 7** (Variance)**.** *Suppose that Assumptions (A1)–(A4) hold, and define*

$$\begin{aligned}C(y) &= \mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\{F_{Y_1|\boldsymbol{X}_1}(y)\}^2\right], \\ \nu^2(y) &= \sigma^2(y) - C(y) \\ &= \mathsf{E}\left[\frac{F_{Y_1|\boldsymbol{X}_1}(y)(1-F_{Y_1|\boldsymbol{X}_1}(y))}{\pi_1(\boldsymbol{X}_1)}\right] + \mathsf{E}\left[\{F_{Y_1|\boldsymbol{X}_1}(y)\}^2\right] - \{F(y)\}^2.\end{aligned} \tag{3.3}$$

*Then, for any $y \in (0,1)$, we have, as $n \to \infty$,*

$$\begin{aligned}\mathsf{Var}(\widehat{F}_{n,m}(y)) &= \mathsf{Var}(\widetilde{F}_{n,m}(y)) - n^{-1}C(y) + \mathcal{O}(n^{-1}m^{-1}) \\ &= n^{-1}\nu^2(y) - n^{-1}m^{-1/2}V(y) + \mathrm{o}_y(n^{-1}m^{-1/2}),\end{aligned}$$

*where $V(y)$ is defined in* (3.2)*.*

Combining the bias and the reduced variance, we obtain the MSE for the feasible estimator. Since the leading bias term and the variance reduction term due to smoothing remain the same as for the pseudo estimator, the asymptotically optimal degree $m_{\mathrm{opt}}(y)$ is unchanged. However, the resulting MSE is smaller due to the lower asymptotic variance: typically $\nu^2(y) < \sigma^2(y)$.

**Corollary 8** (Mean squared error)**.** *Suppose that Assumptions (A1)–(A4) hold. Then, for any $y \in (0,1)$, we have, as $n \to \infty$,*

$$\begin{aligned}\mathrm{MSE}[\widehat{F}_{n,m}(y)] &= \mathsf{Var}(\widehat{F}_{n,m}(y)) + \{\mathsf{Bias}(\widehat{F}_{n,m}(y))\}^2 \\ &= \mathrm{MSE}[\widetilde{F}_{n,m}(y)] - n^{-1}C(y) + \mathcal{O}(n^{-1}m^{-1}) + \mathcal{O}(n^{-2}) \\ &= n^{-1}\nu^2(y) - n^{-1}m^{-1/2}V(y) + m^{-2}B^2(y) + \mathrm{o}_y(n^{-1}m^{-1/2}) + \mathrm{o}(m^{-2}) + \mathcal{O}(n^{-2}).\end{aligned}$$

*In particular, if $B(y)V(y) \neq 0$, the asymptotically optimal choice of $m$, with respect to the* MSE*, is*

$$m_{\mathrm{opt}}(y) = n^{2/3}\left[4B^2(y)/V(y)\right]^{2/3},$$

*in which case, as $n \to \infty$,*

$$\mathrm{MSE}[\widehat{F}_{m_{\mathrm{opt}},n}(y)] = n^{-1}\nu^2(y) - n^{-4/3}\left[27\,V^4(y)/\{256B^2(y)\}\right]^{1/3} + \mathrm{o}_y(n^{-4/3}).$$

The global performance, measured by MISE, also improves for the feasible estimator. The optimal global degree $m_{\text{opt}}$ remains consistent with the pseudo estimator, confirming the overall superiority of the feasible estimator in this setting.

**Corollary 9** (Mean integrated squared error)**.** *Suppose that Assumptions (A1)–(A4) hold. Then, as $n \to \infty$, we have*

$$\begin{aligned}\text{MISE}[\widehat{F}_{n,m}] &= \int_0^1 \text{MSE}[\widehat{F}_{n,m}(y)]\mathrm{d}y \\ &= \text{MISE}[\widetilde{F}_{n,m}] - n^{-1}\int_0^1 C(y)\mathrm{d}y + \mathcal{O}(n^{-1}m^{-1}) + \mathcal{O}(n^{-2}) \\ &= n^{-1}\int_0^1 \nu^2(y)\mathrm{d}y - n^{-1}m^{-1/2}\int_0^1 V(y)\mathrm{d}y + m^{-2}\int_0^1 B^2(y)\mathrm{d}y \\ &\quad + \mathrm{o}(n^{-1}m^{-1/2}) + \mathrm{o}(m^{-2}) + \mathcal{O}(n^{-2}).\end{aligned}$$

*In particular, if $\int_0^1 B^2(y)\mathrm{d}y \times \int_0^1 V(y)\mathrm{d}y \neq 0$, the asymptotically optimal choice of $m$, with respect to the* MISE*, is*

$$m_{\text{opt}} = n^{2/3}\left[\frac{4\int_0^1 B^2(y)\mathrm{d}y}{\int_0^1 V(y)\mathrm{d}y}\right]^{2/3},$$

*in which case, as $n \to \infty$,*

$$\text{MISE}[\widehat{F}_{m_{\text{opt}},n}] = n^{-1}\int_0^1 \nu^2(y)\mathrm{d}y - n^{-4/3}\left[\frac{27\{\int_0^1 V(y)\mathrm{d}y\}^4}{256\int_0^1 B^2(y)\mathrm{d}y}\right]^{1/3} + \mathrm{o}(n^{-4/3}).$$

Finally, we establish the asymptotic normality of the feasible estimator. The conditions on the polynomial degree remain the same as for the pseudo estimator, but the asymptotic variance $\nu^2(y)$ is smaller than $\sigma^2(y)$, reflecting the efficiency gain from estimating the propensity scores.

**Theorem 10** (Asymptotic normality)**.** *Suppose that Assumptions (A1)–(A4) hold, and assume that $\nu^2(y) > 0$ in* (3.3)*. Then, for any $y \in (0,1)$ and $\lambda \in (0,\infty)$, we have, as $n \to \infty$,*

$$n^{1/2}\{\widehat{F}_{n,m}(y) - F(y)\} \xrightarrow{\text{law}} \begin{cases}\mathcal{N}(0,\nu^2(y)), & \text{if } n^{1/2}m^{-1} \to 0, \\ \mathcal{N}(\lambda B(y),\nu^2(y)), & \text{if } n^{1/2}m^{-1} \to \lambda.\end{cases}$$

## 4. Monte Carlo simulations

In this section, we conduct a modest Monte Carlo study to evaluate three families of estimators for the CDF under MAR: the unsmoothed IPW empirical CDF, its Bernstein-smoothed version, and the *integrated* IPW Gaussian KDE of Dubnicka (2009) (abbreviated I-IPW KDE). Each family is examined in two regimes: a pseudo version using known propensities and a feasible version using propensities estimated nonparametrically, as described in Section 2. All simulations were implemented in R; see Appendix A for the GitHub link to the code.

### *4.1. Data generating process and setup*

We consider a setting with a univariate continuous response $Y$ and a univariate discrete auxiliary covariate $X$ ($d$=1). This entails no loss of generality for discrete covariates, given that any finite-dimensional discrete $\boldsymbol{X}$ can be recoded into a single factor by labeling the cells of the Cartesian product of its marginal categories. The data generating process used in the code is:

- $X_i \sim \text{Bernoulli}(p_X)$ with $p_X = 0.5$;
- $Y_i \sim \text{Beta}(\alpha,\beta)$ with $(\alpha,\beta) = (2,5)$;

- Missingness is MAR with propensity $\pi(X_i)$ given by

$$\pi(x) = \begin{cases} 0.6, & x = 0, \\ 0.9, & x = 1, \end{cases} \qquad \text{so that } \mathsf{E}[\pi(X_i)] = 0.75 \text{ and } \pi_{\min} = 0.6.$$

For the feasible estimators, the propensity is estimated nonparametrically using

$$\widehat{\pi}(x) = \frac{\sum_{i=1}^{n} \delta_i \mathbb{1}_{\{X_i = x\}}}{\sum_{i=1}^{n} \mathbb{1}_{\{X_i = x\}}}.$$

We compare six estimators of the CDF $F$, based on the true IPW weights $W_i = \delta_i/\pi(X_i)$ in the pseudo setting and the estimated IPW weights $\widehat{W}_i = \delta_i/\widehat{\pi}(X_i)$ in the feasible setting:

(i) the unsmoothed pseudo estimator $\widetilde{F}_n$ using $W_i$;

(ii) the Bernstein-smoothed pseudo estimator $\widetilde{F}_{n,m} = \mathcal{B}_m(\widetilde{F}_n)$ (smoothing $\widetilde{F}_n$);

(iii) the pseudo I-IPW KDE, obtained by integrating the IPW KDE of Dubnicka (2009) constructed with $W_i$;

(iv) the unsmoothed feasible estimator $\widehat{F}_n$ using $\widehat{W}_i$;

(v) the Bernstein-smoothed feasible estimator $\widehat{F}_{n,m} = \mathcal{B}_m(\widehat{F}_n)$ (smoothing $\widehat{F}_n$);

(vi) the feasible I-IPW KDE, obtained by integrating the IPW KDE built with $\widehat{W}_i$.

*Simulation parameters.* We use sample sizes $n \in \{25, 50, 100, 200, 400, 800, 1600, 3200, 6400\}$ and $N = 100$ Monte Carlo replications per $n$. For each replication, we compute the integrated squared error (ISE),

$$\mathrm{ISE}(\widehat{F}) = \int_0^1 \{\widehat{F}(y) - F(y)\}^2 \, \mathrm{d}y,$$

numerically via a uniform Riemann sum over $G = 2^{10} = 512$ equally spaced grid points on $[0, 1]$. We summarize the distribution of the ISE across replications by its mean, median, interquartile range, and variance. Replications are parallelized over the available CPU cores.

### *4.2. Bandwidth selection via optimized LSCV*

For the smoothed Bernstein estimator $\widehat{F}_{n,m}$, the polynomial degree $m$ is chosen by least-squares cross-validation (LSCV). The derivations below are completely analogous for $\widetilde{F}_{n,m}$, and thus are omitted. The LSCV criterion (up to an additive constant independent of $m$) is

$$\mathrm{LSCV}(m) = \underbrace{\int_0^1 \widehat{F}_{n,m}(y)^2 \, \mathrm{d}y}_{\text{Term 1}} - \underbrace{\frac{2}{n} \sum_{i=1}^{n} \widehat{W}_i \int_{Y_i}^1 \widehat{F}_{n,m}^{(-i)}(y) \, \mathrm{d}y}_{\text{Term 2}}, \tag{4.1}$$

where $\widehat{F}_{n,m}^{(-i)}$ denotes the leave-one-out version (i.e., $\widehat{F}_{n,m}$ with the $i$th data point left out).

*Term 1.* Using bilinearity and the Bernstein basis, we obtain

$$\int_0^1 \widehat{F}_{n,m}(y)^2 \, \mathrm{d}y = \sum_{k=0}^{m} \sum_{\ell=0}^{m} \widehat{F}_n(k/m) \widehat{F}_n(\ell/m) \binom{m}{k} \binom{m}{\ell} B(k + \ell + 1, 2m - k - \ell + 1),$$

where $B(\cdot, \cdot)$ denotes the Beta function (Olver *et al.* , 2010, Section 5.12). Term 1 is evaluated in $\mathcal{O}(m^2)$ operations.

*Term 2.* By Fubini's theorem and the identity $F(y) = \mathsf{E}[\mathbb{1}_{\{y \geq Y\}}]$,

$$\int_0^1 \widehat{F}_{n,m}(y) F(y)\, \mathrm{d}y = \mathsf{E}\left[\int_Y^1 \widehat{F}_{n,m}(y)\, \mathrm{d}y\right],$$

which motivates the leave-one-out estimator in (4.1). Expanding the Bernstein operator and integrating the Bernstein basis functions gives

$$\int_{Y_i}^1 b_{m,k}(y)\, \mathrm{d}y = \frac{1 - \mathrm{pbeta}(Y_i; k+1, m-k+1)}{m+1},$$

where $\mathrm{pbeta}(\cdot\,; \alpha, \beta)$ denotes the CDF of the $\mathrm{Beta}(\alpha, \beta)$ distribution, so that

$$\int_{Y_i}^1 \widehat{F}_{n,m}^{(-i)}(y)\, \mathrm{d}y = \sum_{k=0}^{m} \widehat{F}_n^{(-i)}(k/m) \frac{1 - \mathrm{pbeta}(Y_i; k+1, m-k+1)}{m+1}.$$

Here, $\widehat{F}_n^{(-i)}(k/m)$ can be obtained without recomputing from scratch via

$$\widehat{F}_n^{(-i)}(k/m) = \frac{n\widehat{F}_n(k/m) - \widehat{W}_i \mathbb{1}_{\{Y_i \leq k/m\}}}{n-1}.$$

Thus Term 2 is computed in $\mathcal{O}(nm)$ operations. Overall, $\mathrm{LSCV}(m)$ is evaluated in $\mathcal{O}(m^2 + nm)$ operations for each $m$. We search $m$ over a data-dependent grid

$$m \in \{m_{\min}, \ldots, m_{\max}(n)\}, \qquad m_{\max}(n) = \min(\lfloor cn^{2/3} \rfloor,\ m_{\mathrm{cap}},\ n),$$

with $(m_{\min}, m_{\mathrm{cap}}, c) = (1, 300, 3)$ by default in our R script. The selected $m^\star$ minimizes (4.1).

For the integrated-IPW KDE competitor, the kernel bandwidth $h$ is chosen by an entirely analogous LSCV scheme; see Section 2.2 of Dubnicka (2009) for details. Obviously, in our setting, we replace the IPW KDE there by its integrated version to get the corresponding CDF estimator (the I-IPW KDE mentioned earlier). Since the optimization and computational structure mirror the Bernstein case, we omit further details.

### *4.3. Results and discussion*

The simulation results are summarized in Table 1 for the pseudo estimators and Table 2 for the feasible estimators. For each sample size $n$, we report the median and interquartile range of the ISEs (top block) and the mean and variance of the ISE (bottom block), each split into Unsmoothed (the IPW empirical CDF), Bernstein (the Bernstein-smoothed IPW empirical CDF), and I-IPW KDE columns.

*Pseudo estimators (known propensity).* Across all sample sizes, the integrated-IPW KDE of Dubnicka (2009) dominates the alternatives. It achieves the smallest mean and median ISE and, simultaneously, the smallest dispersion (interquartile range and variance of the ISE). Bernstein smoothing improves markedly over the unsmoothed IPW empirical CDF (as expected) but remains uniformly behind the I-IPW KDE in this oracle setting.

*Feasible estimators (estimated propensity).* When propensities are estimated from the discrete $X$ (the more common case in practice), the ranking depends on the sample size. For small $n$ (e.g., $n = 25, 50$), the I-IPW KDE yields the lowest mean and median ISE, with interquartile range and variance that are comparable to those of the Bernstein method. For moderate to large $n$ (from about $n \geq 200$; results are mixed for $n = 100$), the Bernstein estimator becomes preferable: it attains the smallest mean and median ISE and typically the second smallest dispersion (interquartile range and variance), just behind the unsmoothed IPW empirical CDF. This pattern aligns with our theory for the feasible case, where estimating the propensity scores reduces the leading variance and allows the boundary-adaptive Bernstein smoothing to deliver the best overall risk at larger $n$.

| $n$ | Median ISE | | | Interquartile Range ISE | | |
|---|---|---|---|---|---|---|
| | Unsmoothed | Bernstein | I-IPW KDE | Unsmoothed | Bernstein | I-IPW KDE |
| 25 | 0.0096555 | 0.0070414 | **0.0030089** | 0.0127329 | 0.0112873 | **0.0085635** |
| 50 | 0.0047962 | 0.0043267 | **0.0011669** | 0.0075497 | 0.0084605 | **0.0037742** |
| 100 | 0.0022661 | 0.0022514 | **0.0007979** | 0.0031720 | 0.0028093 | **0.0015168** |
| 200 | 0.0012260 | 0.0010812 | **0.0005476** | 0.0020964 | 0.0019265 | **0.0013304** |
| 400 | 0.0006381 | 0.0005614 | **0.0002186** | 0.0009655 | 0.0008969 | **0.0005294** |
| 800 | 0.0002860 | 0.0002664 | **0.0001286** | 0.0004353 | 0.0003667 | **0.0002541** |
| 1600 | 0.0001638 | 0.0001569 | **0.0000879** | 0.0002642 | 0.0002915 | **0.0001743** |
| 3200 | 0.0000647 | 0.0000625 | **0.0000458** | 0.0001016 | 0.0001147 | **0.0000600** |
| 6400 | 0.0000350 | 0.0000326 | **0.0000207** | 0.0000523 | 0.0000504 | **0.0000335** |
| $n$ | Mean ISE | | | Variance ISE ($\times 10^3$) | | |
| | Unsmoothed | Bernstein | I-IPW KDE | Unsmoothed | Bernstein | I-IPW KDE |
| 25 | 0.0143201 | 0.0133532 | **0.0071434** | 0.2847269 | 0.3125405 | **0.0978971** |
| 50 | 0.0086517 | 0.0077342 | **0.0034877** | 0.0903042 | 0.0816562 | **0.0290106** |
| 100 | 0.0031670 | 0.0029573 | **0.0015708** | 0.0101113 | 0.0098666 | **0.0043262** |
| 200 | 0.0019346 | 0.0018317 | **0.0009948** | 0.0025334 | 0.0025783 | **0.0011975** |
| 400 | 0.0009751 | 0.0009097 | **0.0004582** | 0.0012254 | 0.0012337 | **0.0005215** |
| 800 | 0.0004473 | 0.0004261 | **0.0002247** | 0.0002172 | 0.0002237 | **0.0000579** |
| 1600 | 0.0002607 | 0.0002544 | **0.0001523** | 0.0000703 | 0.0000735 | **0.0000252** |
| 3200 | 0.0001115 | 0.0001097 | **0.0000585** | 0.0000113 | 0.0000113 | **0.0000028** |
| 6400 | 0.0000635 | 0.0000622 | **0.0000319** | 0.0000051 | 0.0000049 | **0.0000009** |

Table 1: Pseudo-estimators (known propensity). Top: Median ISE and Interquartile Range ISE. Bottom: Mean ISE and Variance ISE. Columns subdivided into Unsmoothed, Bernstein-smoothed, and Integrated-IPW KDE. Bold indicates the minimum for each $n$ within each group.

| $n$ | Median ISE | | | Interquartile Range ISE | | |
|---|---|---|---|---|---|---|
| | Unsmoothed | Bernstein | I-IPW KDE | Unsmoothed | Bernstein | I-IPW KDE |
| 25 | 0.0036403 | 0.0032637 | **0.0019051** | **0.0043712** | 0.0069910 | 0.0057725 |
| 50 | 0.0017426 | 0.0009735 | **0.0008677** | 0.0022932 | 0.0026344 | **0.0020441** |
| 100 | 0.0008157 | 0.0006579 | **0.0005637** | **0.0010321** | 0.0010690 | 0.0010337 |
| 200 | 0.0004531 | **0.0003054** | 0.0004395 | 0.0005402 | **0.0005118** | 0.0007073 |
| 400 | 0.0002851 | 0.0002310 | **0.0001647** | **0.0002725** | 0.0002801 | 0.0003617 |
| 800 | 0.0001152 | **0.0000942** | 0.0001209 | **0.0001116** | 0.0001456 | 0.0002016 |
| 1600 | 0.0000641 | **0.0000571** | 0.0000678 | **0.0000769** | 0.0000874 | 0.0001231 |
| 3200 | 0.0000269 | **0.0000240** | 0.0000384 | **0.0000205** | 0.0000241 | 0.0000581 |
| 6400 | 0.0000145 | **0.0000134** | 0.0000188 | **0.0000137** | 0.0000143 | 0.0000258 |
| $n$ | Mean ISE | | | Variance ISE ($\times 10^3$) | | |
| | Unsmoothed | Bernstein | I-IPW KDE | Unsmoothed | Bernstein | I-IPW KDE |
| 25 | 0.0050741 | 0.0049555 | **0.0043381** | **0.0216479** | 0.0248313 | 0.0283421 |
| 50 | 0.0025418 | 0.0019601 | **0.0016662** | 0.0046908 | 0.0042419 | **0.0038480** |
| 100 | 0.0011946 | **0.0010309** | 0.0010398 | **0.0008924** | 0.0010181 | 0.0013722 |
| 200 | 0.0006369 | **0.0005393** | 0.0005946 | **0.0003167** | 0.0003213 | 0.0003481 |
| 400 | 0.0003485 | **0.0002904** | 0.0003163 | 0.0000875 | **0.0000872** | 0.0001319 |
| 800 | 0.0001436 | **0.0001235** | 0.0001630 | 0.0000121 | **0.0000104** | 0.0000191 |
| 1600 | 0.0000848 | **0.0000793** | 0.0001094 | **0.0000038** | 0.0000040 | 0.0000099 |
| 3200 | 0.0000348 | **0.0000316** | 0.0000493 | **0.0000007** | 0.0000007 | 0.0000019 |
| 6400 | 0.0000173 | **0.0000165** | 0.0000268 | **0.0000001** | 0.0000001 | 0.0000005 |

Table 2: Feasible estimators (estimated propensity). Top: Median ISE and Interquartile Range ISE. Bottom: Mean ISE and Variance ISE. Columns subdivided into Unsmoothed, Bernstein-smoothed, and Integrated-IPW KDE. Bold indicates the minimum for each $n$ within each group.

Overall, the experiments indicate the following practical guidance. If propensities are known, the I-IPW KDE is the strongest performer in our design. In the practically relevant feasible scenario, the I-IPW KDE is attractive at small $n$, while for moderate and large $n$ the Bernstein-smoothed estimator is the recommended choice: it offers the best central accuracy (mean/median ISE) and near-best dispersion.

## 5. Real-data application

We highlight the method's utility in the US National Health and Nutrition Examination Survey (NHANES) 2017–2018 cycle (National Center for Health Statistics, 2020); see Endres *et al.* (2025) to retrieve the dataset. The continuous outcome is fasting plasma glucose (`LBXGLU`, mg/dL). Glucose is observed only for participants in the fasting laboratory subsample; for others it is missing. Let $\delta_i = \mathbb{1}_{\{\text{glucose observed for subject } i\}}$. Conditioning on fully observed design covariates, missingness is plausibly MAR. The propensities are unknown, so we use the feasible estimators $\widehat{F}_n$ and $\widehat{F}_{n,m}$ defined in Section 2.

The auxiliary variable $X$ is a discrete 4-level factor formed as the cross of the NHANES exam period indicator `RIDEXMON` and sex `RIAGENDR`. In NHANES coding, `RIDEXMON`$\in \{1, 2\}$ denotes examination period (1: November–April, 2: May–October) and `RIAGENDR`$\in \{1, 2\}$ denotes sex (1: Male, 2: Female). We map the four cells to

$$X = 1 : \text{November–April, Male}, \qquad X = 2 : \text{November–April, Female},$$
$$X = 3 : \text{May–October, Male}, \qquad X = 4 : \text{May–October, Female}.$$

To apply Bernstein smoothing on the unit interval while preserving order, we rescale glucose to $[0, 1]$ using the transformation:

$$Y_i = \min\left\{ \max\left\{ \frac{\texttt{LBXGLU}_i - a}{b - a}, 0 \right\}, 1 \right\}, \qquad (a, b) = (40, 460) \text{ mg/dL}.$$

As in Section 4, the Bernstein degree $m$ is selected by leave-one-out LSCV.

Table 3 reports the effective sample size after requiring $X$ to be observed, the overall observed fraction $n^{-1}\sum_{i=1}^{n} \delta_i$, and the selected degree $m^*$.

| $n$ | Observed rate (%) | $m^*$ |
|---|---|---|
| 3036 | 95.2 | 516 |

Table 3: NHANES 2017–2018 real-data application (feasible estimator with estimated propensity) on the fasting-lab subsample. The table reports the subsample size used (with $X$ observed), overall observed rate, and the LSCV-chosen Bernstein degree $m^*$.

Figure 1 overlays the feasible IPW empirical CDF and its Bernstein-smoothed counterpart. Smoothing yields a monotone, boundary-adaptive CDF with visibly reduced roughness, consistent with the efficiency gains predicted by our theory; the selected $m^*$ balances bias and variance effectively. The workflow could be extended by refining $X$ (for example, adding age groups) or by reporting quantiles back on the original mg/dL scale via the inverse of the monotone rescaling.

## 6. Summary and outlook

This paper developed a smooth, shape-preserving approach to CDF estimation under MAR by applying the Bernstein operator $\mathcal{B}_m$ to IPW empirical CDFs. The estimator is monotone and $[0, 1]$-valued by construction and exploits the bounded support to mitigate boundary bias. We studied two variants, a pseudo estimator with known propensities and a feasible estimator with propensities estimated nonparametrically from discrete $\boldsymbol{X}$, and we derived pointwise and integrated risk expansions, optimal degrees $m$, and asymptotic normality. A central finding was a strict variance improvement for the feasible estimator, where the oracle variance $\sigma^2(y)$ is replaced

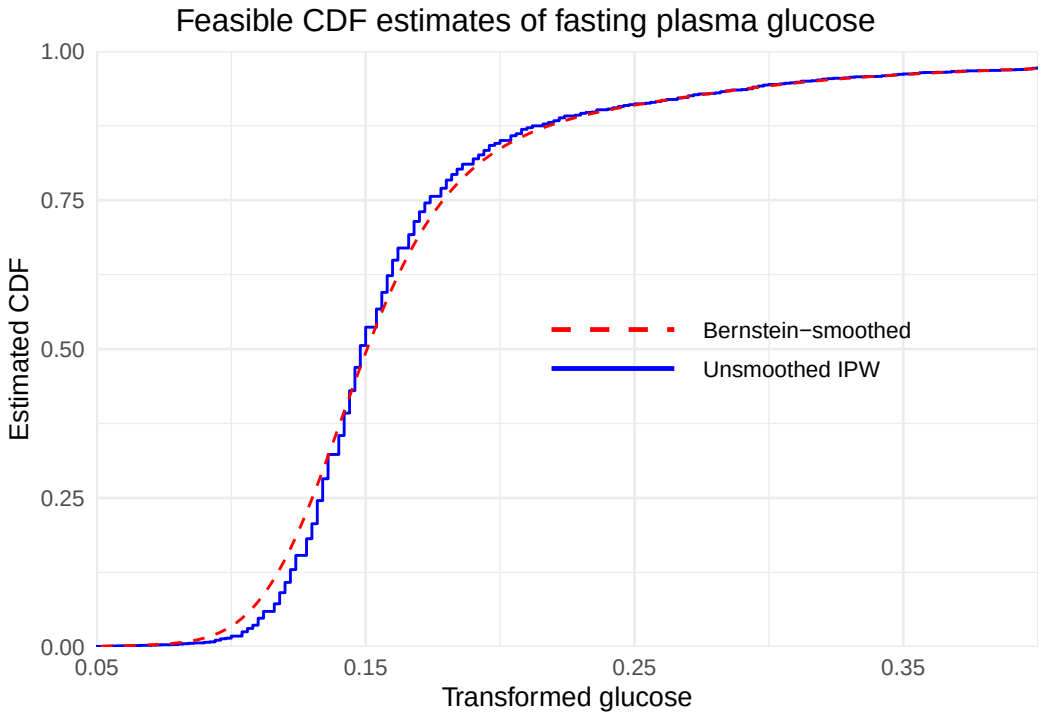


Figure 1: Feasible CDF of fasting plasma glucose: unsmoothed IPW versus Bernstein-smoothed (LSCV-chosen degree $m^*$). The horizontal axis only shows $[0.05, 0.40]$ for a clearer view.

by $\nu^2(y) = \sigma^2(y) - C(y)$ with $C(y) \geq 0$ (Propositions 2 and 7). The MSE expansions show the usual $n^{-1}$ variance and $m^{-2}$ bias together with a variance reduction term $-n^{-1}m^{-1/2}$ that rewards moderate smoothing. Our Monte Carlo results and the NHANES application indicate that smoothing improves finite-sample performance.

In practice, the workflow we advocate is straightforward. Map the response variable $Y$ to $[0, 1]$ by a monotone transformation when the support is known or can be sensibly capped; estimate $\widehat{\pi}(\boldsymbol{X})$ nonparametrically when $\boldsymbol{X}$ is discrete, or by kernel methods when $\boldsymbol{X}$ is continuous; enforce a small floor $\pi_{\min}$, or trim or stabilize the weights, to control extreme IPW weights from very small propensities; select $m$ by LSCV. One LSCV evaluation has a cost of $\mathcal{O}(m^2 + nm)$ operations, and parallelizing across candidate $m$ and over Monte Carlo replications keeps wall-clock time modest. In our real-data analysis, smoothing added little overhead relative to the unsmoothed IPW curve and produced a visibly less ragged CDF estimate.

Our analysis has limitations that suggest future research directions. The variance comparison in Proposition 7 was derived for discrete $\boldsymbol{X}$. Extending it to continuous or high-dimensional covariates will need new expansions for flexible $\widehat{\pi}$ and stronger assumptions that keep propensities away from 0. Since IPW can be unstable when some propensities are very small, simple fixes such as trimming extreme weights, using stabilized weights, or overlap weighting can be applied before smoothing, with only minor changes to the proofs. We studied a univariate response; carrying the same shape-preserving ideas to several dimensions, for example via rearrangements or copulas, is promising but technically harder. For inference, we proved pointwise limits. Uniform confidence bands for $F$ and for the induced quantiles will likely require multiplier or bootstrap methods tailored to the smoothed IPW empirical CDF process. It is also natural to pair Bernstein smoothing with CDF estimators that combine outcome modeling and weighting for missingness, and to accommodate survey features such as design weights, calibration, and clustering.

Overall, coupling IPW with Bernstein smoothing yields estimators that are principled, fast, and easy to implement. They respect the parameter geometry, adapt to boundaries automatically, and dominate unsmoothed IPW empirical CDFs for finite $n$.

## 7. Proofs

*7.1. Proofs of the asymptotic properties of $\widetilde{F}_{n,m}$*

*Proof of Proposition 1.* Let $y \in [0,1]$ be given. By conditioning on $\boldsymbol{X}_1$, then using the fact that $\delta_1$ is independent of $Y_1$ conditionally on $\boldsymbol{X}_1$, and $\mathsf{E}[\delta_1 \mid \boldsymbol{X}_1] = \pi_1(\boldsymbol{X}_1)$, we have

$$\begin{aligned}
\mathsf{E}[\widetilde{F}_{n,m}(y)] &= \sum_{k=0}^{m} \mathsf{E}\left[\frac{\delta_1}{\pi_1(\boldsymbol{X}_1)} \mathbb{1}_{\{Y_1 \le k/m\}}\right] b_{m,k}(y) \\
&= \sum_{k=0}^{m} \mathsf{E}\left[\mathsf{E}\left[\frac{\delta_1}{\pi_1(\boldsymbol{X}_1)} \mathbb{1}_{\{Y_1 \le k/m\}} \mid \boldsymbol{X}_1\right]\right] b_{m,k}(y) \\
&= \sum_{k=0}^{m} \mathsf{E}\left[\frac{\mathsf{E}(\delta_1 \mid \boldsymbol{X}_1)}{\pi_1(\boldsymbol{X}_1)} \mathsf{E}\left[\mathbb{1}_{\{Y_1 \le k/m\}} \mid \boldsymbol{X}_1\right]\right] b_{m,k}(y) \qquad (7.1) \\
&= \sum_{k=0}^{m} \mathsf{E}\left[\mathbb{1}_{\{Y_1 \le k/m\}}\right] b_{m,k}(y) \\
&= \sum_{k=0}^{m} F(k/m) b_{m,k}(y).
\end{aligned}$$

Next, under Assumption (A3), a second-order Taylor expansion of $F$ around $y$ yields

$$F(k/m) = F(y) + f(y)(k/m - y) + \frac{1}{2} f'(y)(k/m - y)^2 + \mathcal{O}(|k/m - y|^3).$$

Summing over the binomial weights, $b_{m,k}(y)$, and applying the well-known moment formulas,

$$\begin{aligned}
\sum_{k=0}^{m} (k/m - y) b_{m,k}(y) &= \frac{1}{m} \sum_{k=0}^{m} (k - my) b_{m,k}(y) = 0, \\
\sum_{k=0}^{m} (k/m - y)^2 b_{m,k}(y) &= \frac{1}{m^2} \sum_{k=0}^{m} (k - my)^2 b_{m,k}(y) = \frac{y(1-y)}{m},
\end{aligned}$$

one deduces

$$\mathcal{B}_m(F)(y) = \sum_{k=0}^{m} F(k/m) b_{m,k}(y) = F(y) + \frac{y(1-y)}{2m} f'(y) + \mathrm{o}(m^{-1}). \qquad (7.2)$$

The error term is a consequence of the Cauchy-Schwarz inequality:

$$\begin{aligned}
\sum_{k=0}^{m} |k/m - y|^3 b_{m,k}(y) &\le \frac{1}{m^3} \left(\sum_{k=0}^{m} |k - my|^4 b_{m,k}(y)\right)^{1/2} \left(\sum_{k=0}^{m} |k - my|^2 b_{m,k}(y)\right)^{1/2} \\
&= \frac{1}{m^3} (3m^2 y^2 (y-1)^2 + my(1-y)(6y^2 - 6y + 1))^{1/2} (my(1-y))^{1/2} \\
&= \mathcal{O}(m^{-3/2}).
\end{aligned}$$

This concludes the proof. □

*Proof of Proposition 2.* Let $y \in (0,1)$. Because of (7.1), we can write

$$\widetilde{F}_{n,m}(y) - \mathsf{E}[\widetilde{F}_{n,m}(y)] = \widetilde{F}_{n,m}(y) - \mathcal{B}_m(F)(y) = n^{-1} \sum_{i=1}^{n} Z_{i,m}, \qquad (7.3)$$

where

$$Z_{i,m} = \sum_{k=0}^{m} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) - \mathcal{B}_m(F)(y).$$

Given that $Z_{1,m}, \ldots, Z_{n,m}$ are iid and centered, and $\delta_1^2 = \delta_1$, we deduce that

$$\begin{aligned}\mathsf{Var}(\widetilde{F}_{n,m}(y)) &= n^{-1}\mathsf{Var}(Z_{1,m}) \\ &= n^{-1}\left\{\sum_{k,\ell=0}^{m} \mathsf{E}\left[\frac{\delta_1}{\pi_1(\boldsymbol{X}_1)^2}\mathbb{1}_{\{Y_1\le(k\wedge\ell)/m\}}\right] b_{m,k}(y)b_{m,\ell}(y)\right\} - n^{-1}\{\mathcal{B}_m(F)(y)\}^2,\end{aligned}$$

where $a \wedge b \equiv \min(a,b)$. Moreover, $\delta_1$ is independent of $Y_1$ conditionally on $\boldsymbol{X}_1$, so we have

$$\begin{aligned}\mathsf{E}\left[\frac{\delta_1}{\pi_1(\boldsymbol{X}_1)^2}\mathbb{1}_{\{Y_1\le(k\wedge\ell)/m\}}\right] &= \mathsf{E}\left[\frac{\mathsf{E}[\delta_1 \mid \boldsymbol{X}_1]}{\pi_1(\boldsymbol{X}_1)^2}\mathsf{E}\left[\mathbb{1}_{\{Y_1\le(k\wedge\ell)/m\}} \mid \boldsymbol{X}_1\right]\right] \\ &= \mathsf{E}\left[\frac{1}{\pi_1(\boldsymbol{X}_1)}\mathsf{E}\left[\mathbb{1}_{\{Y_1\le(k\wedge\ell)/m\}} \mid \boldsymbol{X}_1\right]\right] \\ &= \mathsf{E}\left[\frac{1}{\pi_1(\boldsymbol{X}_1)}F_{Y_1|\boldsymbol{X}_1}((k\wedge\ell)/m)\right].\end{aligned}$$

Next, under Assumption (A4), a first-order Taylor expansion of $F_{Y_1|\boldsymbol{X}_1}$ around $y$ yields

$$F_{Y_1|\boldsymbol{X}_1}((k\wedge\ell)/m) = F_{Y_1|\boldsymbol{X}_1}(y) + f_{Y_1|\boldsymbol{X}_1}(y)((k\wedge\ell)/m - y) + \mathcal{O}(|(k\wedge\ell)/m - y|^2).$$

Putting the last three equations together with (7.2) yields

$$\begin{aligned}\mathsf{Var}(\widetilde{F}_{n,m}(y)) = n^{-1}\,\mathsf{E}\left[\frac{F_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)} - \{F(y)\}^2 + \frac{f_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)}\sum_{k,\ell=0}^{m}((k\wedge\ell)/m - y)b_{m,k}(y)b_{m,\ell}(y)\right] \\ + \mathcal{O}(n^{-1}m^{-1}),\end{aligned}$$

where the error term also implicitly uses Assumption (A1) to absorb $1/\pi_1(\boldsymbol{X}_1)$ times the Taylor series remainder in the expectation. By Lemma 4 of Ouimet (2021), it is known that

$$\sum_{k,\ell=0}^{m}((k\wedge\ell)/m - y)b_{m,k}(y)b_{m,\ell}(y) = m^{-1/2}\left\{-\sqrt{\frac{y(1-y)}{\pi}} + \mathrm{o}_y(1)\right\}.$$

Therefore,

$$\begin{aligned}\mathsf{Var}(\widetilde{F}_{n,m}(y)) = n^{-1}\left(\mathsf{E}\left[\frac{F_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)}\right] - \{F(y)\}^2\right) - n^{-1}m^{-1/2}\sqrt{\frac{y(1-y)}{\pi}}\mathsf{E}\left[\frac{f_{Y_1|\boldsymbol{X}_1}(y)}{\pi_1(\boldsymbol{X}_1)}\right] \\ + \mathrm{o}_y(n^{-1}m^{-1/2}).\end{aligned}$$

This concludes the proof. □

*Proof of Theorem 5.* Let $y \in (0,1)$. We decompose the standardized pseudo estimator as follows:

$$n^{1/2}\{\widetilde{F}_{n,m}(y) - F(y)\} = n^{1/2}\{\widetilde{F}_{n,m}(y) - \mathsf{E}[\widetilde{F}_{n,m}(y)]\} + n^{1/2}\{\mathsf{E}[\widetilde{F}_{n,m}(y)] - F(y)\}. \tag{7.4}$$

The second term is the scaled bias. By Proposition 1 (under Assumption (A4)),

$$n^{1/2}\{\mathsf{E}[\widetilde{F}_{n,m}(y)] - F(y)\} = n^{1/2}\{m^{-1}B(y) + \mathrm{o}(m^{-1})\} = (n^{1/2}m^{-1})B(y) + \mathrm{o}(n^{1/2}m^{-1}). \tag{7.5}$$

This term converges to 0 if $n^{1/2}m^{-1} \to 0$, and to $\lambda B(y)$ if $n^{1/2}m^{-1} \to \lambda$.

The first term on the right-hand side of (7.4) is the stochastic component. As shown in the proof of Proposition 2 (Eq. (7.3)),

$$n^{1/2}\{\widetilde{F}_{n,m}(y) - \mathsf{E}[\widetilde{F}_{n,m}(y)]\} = n^{-1/2}\sum_{i=1}^{n} Z_{i,m},$$

where

$$Z_{i,m} = \sum_{k=0}^{m} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) - \mathcal{B}_m(F)(y), \quad i \in \{1, \dots, n\}, \tag{7.6}$$

are iid centered random variables. We apply the Lindeberg–Feller central limit theorem (CLT) for double arrays; see, e.g., Serfling (1980, Section 1.9.3). By Proposition 2 (under Assumptions (A1), (A3), and (A4)),

$$\mathsf{Var}(Z_{1,m}) = n\,\mathsf{Var}(\widetilde{F}_{n,m}(y)) \to \sigma^2(y) > 0, \quad \text{as } n \to \infty \text{ (and } m \to \infty). \tag{7.7}$$

We verify the Lindeberg condition for $Z_{1,m}$: for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{\mathsf{Var}(Z_{1,m})} \mathsf{E}\left[ Z_{1,m}^2 \mathbb{1}_{\{|Z_{1,m}|^2 > \varepsilon n\, \mathsf{Var}(Z_{1,m})\}} \right] = 0. \tag{7.8}$$

We check if $Z_{1,m}$ is bounded. Since $\pi_i(\boldsymbol{X}_i) \ge \pi_{\min} > 0$ for all $i \in \{1, \dots, n\}$ by Assumption (A1), and since $\sum_{k=0}^{m} b_{m,k}(y) = 1$ and $0 \le \mathcal{B}_m(F)(y) \le 1$, we have

$$|Z_{1,m}| \le \frac{1}{\pi_{\min}} \sum_{k=0}^{m} b_{m,k}(y) + |\mathcal{B}_m(F)(y)| \le \frac{1}{\pi_{\min}} + 1 < \infty.$$

Given (7.7), the indicator $\mathbb{1}_{\{|Z_{1,m}|^2 > \varepsilon n\, \mathsf{Var}(Z_{1,m})\}}$ is equal to zero almost surely for $n$ sufficiently large, and the Lindeberg condition (7.8) is satisfied. The conclusion follows. □

*7.2. Proofs of the asymptotic properties of $\widehat{F}_{n,m}$*

*Proof of Proposition 6.* Let $y \in [0,1]$ be given. First, note that a Taylor expansion of $1/\hat{\pi}_i(\boldsymbol{X}_{1:n})$ around $1/\pi(\boldsymbol{X}_i)$ yields

$$\frac{1}{\hat{\pi}_i(\boldsymbol{X}_{1:n})} = \frac{1}{\pi_i(\boldsymbol{X}_i)} - \frac{1}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) + \sum_{j=2}^{\infty} \frac{(-1)^j}{\pi_i(\boldsymbol{X}_i)^{j+1}} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))^j.$$

From (2.4), it follows that

$$\begin{aligned} \widehat{F}_{n,m}(y) = \widetilde{F}_{n,m}(y) &- n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)^2} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \\ &+ n^{-1} \sum_{j=2}^{\infty} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{(-1)^j \delta_i}{\pi_i(\boldsymbol{X}_i)^{j+1}} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))^j \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y). \end{aligned} \tag{7.9}$$

If $p$ denotes the marginal probability mass function of each $\boldsymbol{X}_i$, and

$$\hat{p}(\boldsymbol{X}_i) = n^{-1} \sum_{k=1}^{n} \mathbb{1}_{\{\boldsymbol{X}_k = \boldsymbol{X}_i\}}$$

denotes the corresponding empirical estimator, then (2.3) implies

$$\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i) = \frac{n^{-1} \sum_{j=1}^{n} (\delta_j - \pi_i(\boldsymbol{X}_i))\, \mathbb{1}_{\{\boldsymbol{X}_j = \boldsymbol{X}_i\}}}{\hat{p}(\boldsymbol{X}_i)}.$$

A Taylor expansion of $1/\hat{p}(\boldsymbol{X}_i)$ around $1/p(\boldsymbol{X}_i)$ yields

$$\frac{1}{\hat{p}(\boldsymbol{X}_i)} = \frac{1}{p(\boldsymbol{X}_i)} - \frac{1}{p(\boldsymbol{X}_i)^2}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i)) + \sum_{j=2}^{\infty} \frac{(-1)^j}{p(\boldsymbol{X}_i)^{j+1}} (\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))^j.$$

We put this expansion in the previous equation and evaluate term by term:

$$\begin{aligned}\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i) = {} & \frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)} \\ & - \frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)^2}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i)) \\ & + n^{-1}\sum_{\ell=2}^{\infty}\sum_{j=1}^{n} \frac{(-1)^\ell}{p(\boldsymbol{X}_i)^{\ell+1}}(\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))^\ell.\end{aligned} \tag{7.10}$$

The first term on the right-hand side of (7.10) is $\mathcal{O}_{L^2}(n^{-1/2})$. Indeed, we have

$$\begin{aligned}&\mathsf{E}\left[\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)} \mid \boldsymbol{X}_{1:n}\right] \\ &\quad = \mathsf{E}\left[\frac{n^{-1}\sum_{j=1}^n \mathsf{E}[(\delta_j - \pi_i(\boldsymbol{X}_i)) \mid \boldsymbol{X}_{1:n}]\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)}\right] = 0,\end{aligned}$$

and

$$\mathsf{Var}\left(\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)} \mid \boldsymbol{X}_{1:n}\right) = \frac{n^{-2}\sum_{j=1}^n \mathsf{Var}(\delta_j \mid \boldsymbol{X}_{1:n})\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)^2} \ll \frac{n^{-1}}{p_{\min}^2},$$

where $p_{\min} \equiv \min_{\boldsymbol{x}} p(\boldsymbol{x}) > 0$ by Assumption (A2). Hence, using the conditional variance formula,

$$\begin{aligned}&\mathsf{Var}\left(\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)}\right) \\ &\quad = \mathsf{E}\left[\mathsf{Var}\left(\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)} \mid \boldsymbol{X}_{1:n}\right)\right] + 0 \ll \frac{n^{-1}}{p_{\min}^2},\end{aligned}$$

and thus

$$\mathsf{E}\left[\left(\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)}\right)^2\right] \ll \frac{n^{-1}}{p_{\min}^2}.$$

By Cauchy–Schwarz, the second moment of the second term on the right-hand side of (7.10) satisfies

$$\begin{aligned}&\mathsf{E}\left[\left(\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)^2}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))\right)^2\right] \\ &\ll \frac{1}{p_{\min}^4}\sqrt{\mathsf{E}\left[\left(n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}\right)^4\right]}\sqrt{\mathsf{E}\left[(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))^4\right]} \\ &\ll \frac{n^{-2}}{p_{\min}^4},\end{aligned}$$

(each square root being $\ll n^{-1}$) so that

$$\frac{n^{-1}\sum_{j=1}^n (\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)^2}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i)) = \mathcal{O}_{L^2}(n^{-1}), \tag{7.11}$$

with the rate being dependent on $p_{\min}$. Similarly, the residual tail sum satisfies

$$n^{-1}\sum_{\ell=2}^{\infty}\sum_{j=1}^{n} \frac{(-1)^\ell}{p(\boldsymbol{X}_i)^{\ell+1}}(\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}(\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))^\ell = \mathcal{O}_{L^2}(n^{-1}), \tag{7.12}$$

so that, from (7.10), we have

$$\mathsf{E}\left[(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))^2\right] = \mathcal{O}(n^{-1}). \tag{7.13}$$

By applying Hölder's inequality to each $j$-summand in (7.9), one deduces that the expectation of the last term in (7.9) satisfies

$$\mathsf{E}\left[n^{-1}\sum_{j=2}^{\infty}\sum_{k=0}^{m}\sum_{i=1}^{n}\frac{(-1)^j\delta_i}{\pi_i(\boldsymbol{X}_i)^{j+1}}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))^j \mathbb{1}_{\{Y_i\le k/m\}} b_{m,k}(y)\right] = \mathcal{O}(n^{-1}), \tag{7.14}$$

with the rate being dependent on $p_{\min}$ and $\pi_{\min}$.

It remains to evaluate the expectation of the first term on the right-hand side of (7.9). To do so, consider the decomposition:

$$\begin{aligned}
&n^{-1}\sum_{k=0}^{m}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))\mathbb{1}_{\{Y_i\le k/m\}} b_{m,k}(y)\\
&\quad= n^{-1}\sum_{k=0}^{m}\sum_{i=1}^{n}\frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))\mathbb{1}_{\{Y_i\le k/m\}} b_{m,k}(y)\\
&\qquad+ n^{-1}\sum_{k=0}^{m}\sum_{i=1}^{n}\frac{1}{\pi_i(\boldsymbol{X}_i)}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))\mathbb{1}_{\{Y_i\le k/m\}} b_{m,k}(y).
\end{aligned} \tag{7.15}$$

For the first term on the right-hand side of (7.15), note that, for all $i \in \{1,\dots,n\}$, we have

$$\mathsf{E}\left[(\delta_i - \pi(\boldsymbol{X}_i))^2 \mid \boldsymbol{X}_{1:n}\right] = \mathsf{Var}(\delta_i \mid \boldsymbol{X}_{1:n}) = \pi_i(\boldsymbol{X}_i)(1-\pi_i(\boldsymbol{X}_i)),$$

and, for all $j \in \{1,\dots,n\}\backslash\{i\}$,

$$\begin{aligned}
&\mathsf{E}\big[(\delta_i - \pi(\boldsymbol{X}_i))(\delta_j - \pi(\boldsymbol{X}_i))\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}} \mid \boldsymbol{X}_{1:n}\big]\\
&\quad= \mathsf{E}\left[\delta_i - \pi(\boldsymbol{X}_i) \mid \boldsymbol{X}_{1:n}\right]\mathsf{E}\left[(\delta_j - \pi(\boldsymbol{X}_i)) \mid \boldsymbol{X}_{1:n}\right]\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}} = 0.
\end{aligned}$$

Hence, recalling that $\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i) = n^{-1}\sum_{j=1}^{n}(\delta_j - \pi_i(\boldsymbol{X}_i))\,\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}}/\hat{p}(\boldsymbol{X}_i)$, we deduce

$$\begin{aligned}
&\mathsf{E}\left[\frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mid \boldsymbol{X}_{1:n}\right]\\
&\quad= \mathsf{E}\left[\frac{\mathsf{E}\left[(\delta_i - \pi(\boldsymbol{X}_i))^2 \mid \boldsymbol{X}_{1:n}\right]}{n\,\pi_i(\boldsymbol{X}_i)^2\hat{p}(\boldsymbol{X}_i)}\right] + \sum_{\substack{j=1\\ j\neq i}}^{n}\mathsf{E}\left[\frac{\mathsf{E}\big[(\delta_i - \pi(\boldsymbol{X}_i))(\delta_j - \pi(\boldsymbol{X}_i))\mathbb{1}_{\{\boldsymbol{X}_j=\boldsymbol{X}_i\}} \mid \boldsymbol{X}_{1:n}\big]}{n\,\pi_i(\boldsymbol{X}_i)^2\hat{p}(\boldsymbol{X}_i)}\right]\\
&\quad= \mathsf{E}\left[\frac{\pi_i(\boldsymbol{X}_i)(1-\pi_i(\boldsymbol{X}_i))}{n\,\pi_i(\boldsymbol{X}_i)^2\hat{p}(\boldsymbol{X}_i)}\right] + 0\\
&\quad= \mathcal{O}(n^{-1}),
\end{aligned}$$

where the constant implicit in the error term depends on $p_{\min}$ and $\pi_{\min}$. Given that $\delta_i$ and $Y_i$ are independent conditionally on $\boldsymbol{X}_i$, it follows that

$$\begin{aligned}
&\mathsf{E}\left[n^{-1}\sum_{k=0}^{m}\sum_{i=1}^{n}\frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i))\mathbb{1}_{\{Y_i\le k/m\}} b_{m,k}(y)\right]\\
&\quad= \mathsf{E}\left[n^{-1}\sum_{k=0}^{m}\sum_{i=1}^{n}\mathsf{E}\left[\frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}(\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mid \boldsymbol{X}_{1:n}\right]\mathsf{E}[\mathbb{1}_{\{Y_i\le k/m\}} \mid \boldsymbol{X}_{1:n}]\, b_{m,k}(y)\right]\\
&\quad= \mathcal{O}(n^{-1}).
\end{aligned}$$

For the second term on the right-hand side of (7.15), note that $\mathsf{E}[(\delta_j - \pi_i(\boldsymbol{X}_i)) \mid \boldsymbol{X}_{1:n}]\, \mathbb{1}_{\{\boldsymbol{X}_j = \boldsymbol{X}_i\}} = 0$ for all $i, j \in \{1, \dots, n\}$, so that (7.10) implies $\mathsf{E}[\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i) \mid \boldsymbol{X}_{1:n}] = 0$, and thus

$$\begin{aligned}
&\mathsf{E}\left[ n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{1}{\pi_i(\boldsymbol{X}_i)} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \right] \\
&\quad = \mathsf{E}\left[ n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{1}{\pi_i(\boldsymbol{X}_i)} \mathsf{E}\left[\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i) \mid \boldsymbol{X}_{1:n}\right] \mathsf{E}[\mathbb{1}_{\{Y_i \le k/m\}} \mid \boldsymbol{X}_{1:n}]\, b_{m,k}(y) \right] \\
&\quad = 0.
\end{aligned}$$

Putting the last two equations together in (7.15) yields

$$\mathsf{E}\left[ n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)^2} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \right] = \mathcal{O}(n^{-1}).$$

Combined with (7.14) and the decomposition (7.9), the conclusion follows. □

*Proof of Proposition 7.* Let $y \in (0,1)$. We start by looking at the second term in the decomposition (7.9) and we decompose it as in (7.15). The first term on the right-hand side of (7.15) is $\mathcal{O}_{L^2}(n^{-1/2})$:

$$\begin{aligned}
&\mathsf{E}\left[ \left( n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \right)^2 \right] \\
&\quad \le n^{-2} \sum_{k,\ell=0}^{m} \sum_{i,j=1}^{n} \mathsf{E}\left[ \left| \frac{(\delta_i - \pi(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} \right| \right. \\
&\qquad\qquad \left. \left| \frac{(\delta_j - \pi(\boldsymbol{X}_j))}{\pi_j(\boldsymbol{X}_j)^2} (\hat{\pi}_j(\boldsymbol{X}_{1:n}) - \pi_j(\boldsymbol{X}_j)) \mathbb{1}_{\{Y_j \le \ell/m\}} \right| \right] b_{m,k}(y) b_{m,\ell}(y) \\
&\quad \ll \frac{n^{-2}}{\pi_{\min}^4} \sum_{k,\ell=0}^{m} \sum_{i,j=1}^{n} \mathsf{E}\left[ |\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)|\, |\hat{\pi}_j(\boldsymbol{X}_{1:n}) - \pi_j(\boldsymbol{X}_j)| \right] b_{m,k}(y) b_{m,\ell}(y) \\
&\quad \ll \frac{n^{-2}}{\pi_{\min}^4} \sum_{k,\ell=0}^{m} \sum_{i,j=1}^{n} \sqrt{\mathsf{E}\left[ |\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)|^2 \right]} \sqrt{\mathsf{E}\left[ |\hat{\pi}_j(\boldsymbol{X}_{1:n}) - \pi_j(\boldsymbol{X}_j)|^2 \right]} b_{m,k}(y) b_{m,\ell}(y) \\
&\quad \ll \frac{n^{-1}}{\pi_{\min}^4},
\end{aligned}$$

where the last bound follows from (7.13). For the second term on the right-hand side of (7.15), we have, using (7.10),

$$\begin{aligned}
&n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{1}{\pi_i(\boldsymbol{X}_i)} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \\
&\quad = n^{-2} \sum_{k=0}^{m} \sum_{i,j=1}^{n} \frac{(\delta_j - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \frac{\mathbb{1}_{\{\boldsymbol{X}_j = \boldsymbol{X}_i\}}}{p(\boldsymbol{X}_i)} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \\
&\qquad + n^{-2} \sum_{\ell=1}^{\infty} \sum_{k=0}^{m} \sum_{i,j=1}^{n} \frac{(-1)^\ell}{p(\boldsymbol{X}_i)^{\ell+1}} \frac{(\delta_j - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \mathbb{1}_{\{\boldsymbol{X}_j = \boldsymbol{X}_i\}} (\hat{p}(\boldsymbol{X}_i) - p(\boldsymbol{X}_i))^\ell \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y).
\end{aligned}$$

Combining (7.11) and (7.12) in the proof of Proposition 6, the second term on the right-hand side is $\mathcal{O}_{L^2}(n^{-1})$. The first term on the right-hand side is equal to

$$n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) + \mathcal{O}_{L^2}(n^{-1}),$$

by the law of large numbers in $L^2$. Putting all the above together shows that

$$\begin{aligned} & n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)^2} (\hat{\pi}_i(\boldsymbol{X}_{1:n}) - \pi_i(\boldsymbol{X}_i)) \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) \\ & \quad = n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) + \mathcal{O}_{L^2}(n^{-1}). \end{aligned} \tag{7.16}$$

Using the decomposition (7.9) and the definition of the pseudo estimator $\widetilde{F}_{n,m}$ in (2.4), we find

$$\begin{aligned} \widehat{F}_{n,m}(y) &= \widetilde{F}_{n,m}(y) - n^{-1} \sum_{k=0}^{m} \sum_{i=1}^{n} \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) + \mathcal{O}_{L^2}(n^{-1}) \\ &= n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} \mathbb{1}_{\{Y_i \le k/m\}} b_{m,k}(y) - n^{-1} \sum_{i=1}^{n} \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \\ &\quad + \mathcal{O}_{L^2}(n^{-1}) \\ &\equiv A_{n,m} - B_{n,m} + \mathcal{O}_{L^2}(n^{-1}). \end{aligned} \tag{7.17}$$

To obtain the variance of $\widehat{F}_{n,m}(y)$, it remains to compute $\mathsf{Var}(B_{n,m})$ and $\mathsf{Cov}(A_{n,m}, B_{n,m})$, given that we already have the asymptotics of $\mathsf{Var}(A_{n,m})$ from Proposition 2. Since $\delta_i$ and $Y_i$ are independent conditionally on $\boldsymbol{X}_i$, we have

$$\begin{aligned} \mathsf{Var}(B_{n,m}) &= n^{-2} \sum_{i=1}^{n} \mathsf{Var}\left( \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \right) \\ &= n^{-2} \sum_{i=1}^{n} \mathsf{E}\left[ \mathsf{Var}\left( \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \mid \boldsymbol{X}_i \right) \right] \\ &\quad + n^{-2} \sum_{i=1}^{n} \mathsf{Var}\left( \mathsf{E}\left[ \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \mid \boldsymbol{X}_i \right] \right) \\ &= n^{-2} \sum_{i=1}^{n} \mathsf{E}\left[ \frac{\mathsf{Var}(\delta_i \mid \boldsymbol{X}_i)}{\pi_i(\boldsymbol{X}_i)^2} \left( \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \right)^2 \right] \\ &\quad + n^{-2} \sum_{i=1}^{n} \mathsf{Var}\left( \frac{(\mathsf{E}[\delta_i \mid \boldsymbol{X}_i] - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) b_{m,k}(y) \right). \end{aligned}$$

Given that $\mathsf{Var}(\delta_i \mid \boldsymbol{X}_i) = \pi_i(\boldsymbol{X}_i)(1 - \pi_i(\boldsymbol{X}_i))$ and $\mathsf{E}[\delta_i \mid \boldsymbol{X}_i] - \pi_i(\boldsymbol{X}_i) = 0$, it follows that

$$\mathsf{Var}(B_{n,m}) = n^{-1} \mathsf{E}\left[ \frac{(1 - \pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)} \left( \sum_{k=0}^{m} F_{Y_1|\boldsymbol{X}_1}(k/m) b_{m,k}(y) \right)^2 \right].$$

For the covariance, note that

$$\begin{aligned} \mathsf{E}[B_{n,m}] &= n^{-1} \sum_{i=1}^{n} \mathsf{E}\left[ \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) \right] b_{m,k}(y) \\ &= n^{-1} \sum_{i=1}^{n} \mathsf{E}\left[ \mathsf{E}\left[ \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) \mid \boldsymbol{X}_i \right] \right] b_{m,k}(y) \\ &= n^{-1} \sum_{i=1}^{n} \mathsf{E}\left[ \frac{(\mathsf{E}[\delta_i \mid \boldsymbol{X}_i] - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)} \sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m) \right] b_{m,k}(y) \\ &= 0, \end{aligned} \tag{7.18}$$

and thus $\mathsf{Cov}(A_{n,m}, B_{n,m}) = \mathsf{E}[A_{n,m}B_{n,m}]$. Also, the summands of $A_{n,m}$ and $B_{n,m}$ are independent for different indices $i$, so the cross terms are zero:

$$\mathsf{E}[A_{n,m}B_{n,m}] = n^{-2}\sum_{i=1}^{n}\mathsf{E}\left[\frac{\delta_i(\delta_i-\pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}\left(\sum_{k=0}^{m}\mathbb{1}_{\{Y_i\le k/m\}}b_{m,k}(y)\right)\left(\sum_{k=0}^{m}F_{Y_i|\boldsymbol{X}_i}(k/m)b_{m,k}(y)\right)\right].$$

Again, the independence between $\delta_i$ and $Y_i$, conditionally on $\boldsymbol{X}_i$, yields

$$\begin{aligned}
&\mathsf{E}[A_{n,m}B_{n,m}]\\
&= n^{-2}\sum_{i=1}^{n}\mathsf{E}\left[\frac{\mathsf{E}[\delta_i(\delta_i-\pi_i(\boldsymbol{X}_i))\mid\boldsymbol{X}_i]}{\pi_i(\boldsymbol{X}_i)^2}\mathsf{E}\left[\sum_{k=0}^{m}\mathbb{1}_{\{Y_i\le k/m\}}b_{m,k}(y)\mid\boldsymbol{X}_i\right]\left(\sum_{k=0}^{m}F_{Y_i|\boldsymbol{X}_i}(k/m)b_{m,k}(y)\right)\right]\\
&= n^{-2}\sum_{i=1}^{n}\mathsf{E}\left[\frac{\mathsf{E}[\delta_i(\delta_i-\pi_i(\boldsymbol{X}_i))\mid\boldsymbol{X}_i]}{\pi_i(\boldsymbol{X}_i)^2}\left(\sum_{k=0}^{m}F_{Y_i|\boldsymbol{X}_i}(k/m)b_{m,k}(y)\right)^2\right].
\end{aligned}$$

Since $\delta_i^2=\delta_i$ and $\mathsf{E}[\delta_i\mid\boldsymbol{X}_i]=\pi_i(\boldsymbol{X}_i)$, we have

$$\begin{aligned}
\mathsf{E}[A_{n,m}B_{n,m}] &= n^{-2}\sum_{i=1}^{n}\mathsf{E}\left[\frac{\pi_i(\boldsymbol{X}_i)(1-\pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)^2}\left(\sum_{k=0}^{m}F_{Y_i|\boldsymbol{X}_i}(k/m)b_{m,k}(y)\right)^2\right]\\
&= n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\left(\sum_{k=0}^{m}F_{Y_1|\boldsymbol{X}_1}(k/m)b_{m,k}(y)\right)^2\right]\\
&= \mathsf{Var}(B_{n,m}).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathsf{Var}(\widehat{F}_{n,m}(y)) &= \mathsf{Var}(A_{n,m}) - 2\,\mathsf{E}[A_{n,m}B_{n,m}] + \mathsf{Var}(B_{n,m})\\
&= \mathsf{Var}(\widetilde{F}_{n,m}(y)) - \mathsf{Var}(B_{n,m})\\
&= \mathsf{Var}(\widetilde{F}_{n,m}(y)) - n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\left(\sum_{k=0}^{m}F_{Y_1|\boldsymbol{X}_1}(k/m)b_{m,k}(y)\right)^2\right].
\end{aligned}\tag{7.19}$$

Finally, by applying the Taylor expansion under Assumption (A4),

$$F_{Y_i|\boldsymbol{X}_i}(k/m) = F_{Y_i|\boldsymbol{X}_i}(y) + f_{Y_i|\boldsymbol{X}_i}(y)(k/m-y) + \mathcal{O}(|k/m-y|^2),\tag{7.20}$$

we obtain

$$\begin{aligned}
&n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\left(\sum_{k=0}^{m}F_{Y_1|\boldsymbol{X}_1}(k/m)b_{m,k}(y)\right)^2\right]\\
&= n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\{F_{Y_1|\boldsymbol{X}_1}(y)\}^2\right]\\
&\quad - 2n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}F_{Y_1|\boldsymbol{X}_1}(y)f_{Y_1|\boldsymbol{X}_1}(y)\sum_{k=0}^{m}(k/m-y)b_{m,k}(y)\right]\\
&\quad + \mathcal{O}\left(\frac{n^{-1}}{\pi_{\min}}\left(\sum_{k=0}^{m}(k/m-y)b_{m,k}(y)\right)^2\right).
\end{aligned}$$

The second term on the right-hand side is zero because $\sum_{k=0}^{m}(k-my)b_{m,k}(y)=0$. The error term is $\mathcal{O}(n^{-1}m^{-1})$ by Jensen's inequality since

$$\left(\sum_{k=0}^{m}(k/m-y)b_{m,k}(y)\right)^2 \le \sum_{k=0}^{m}(k/m-y)^2b_{m,k}(y) = m^{-2}\sum_{k=0}^{m}(k-my)^2b_{m,k}(y) = m^{-1}y(1-y).$$

Therefore,

$$n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\left(\sum_{k=0}^{m} F_{Y_1|\boldsymbol{X}_1}(k/m)b_{m,k}(y)\right)^2\right] = n^{-1}\mathsf{E}\left[\frac{(1-\pi_1(\boldsymbol{X}_1))}{\pi_1(\boldsymbol{X}_1)}\{F_{Y_1|\boldsymbol{X}_1}(y)\}^2\right] + \mathcal{O}(n^{-1}m^{-1}).$$

Putting back this expression in (7.19) yields the conclusion. □

*Proof of Theorem 10.* Let $y \in (0,1)$. We utilize the asymptotic representation (7.17) derived in the proof of Proposition 7:

$$\widehat{F}_{n,m}(y) = \widetilde{F}_{n,m}(y) - B_{n,m} + \mathcal{O}_{L^2}(n^{-1}),$$

where $B_{n,m}$ is defined as

$$B_{n,m} = n^{-1}\sum_{i=1}^{n} U_{i,m}, \quad U_{i,m} = \frac{(\delta_i - \pi_i(\boldsymbol{X}_i))}{\pi_i(\boldsymbol{X}_i)}\sum_{k=0}^{m} F_{Y_i|\boldsymbol{X}_i}(k/m)b_{m,k}(y). \tag{7.21}$$

The expression for the standardized estimator now follows:

$$n^{1/2}\{\widehat{F}_{n,m}(y) - F(y)\} = n^{1/2}\{\widetilde{F}_{n,m}(y) - B_{n,m} - F(y)\} + \mathcal{O}_{L^2}(n^{-1/2}).$$

For the main term, we know $\mathsf{E}[\widetilde{F}_{n,m}(y)] = \mathcal{B}_m(F)(y)$ by (7.1) and $\mathsf{E}[B_{n,m}] = 0$ by (7.18). We decompose it as:

$$n^{1/2}\{\widetilde{F}_{n,m}(y) - B_{n,m} - F(y)\} = n^{1/2}\{\widetilde{F}_{n,m}(y) - \mathcal{B}_m(F)(y) - B_{n,m}\} + n^{1/2}\{\mathcal{B}_m(F)(y) - F(y)\}.$$

The second term is the bias term, analyzed in the proof of Theorem 5 (Eq. (7.5)). The first term is the stochastic part. Let $W_{i,m} = Z_{i,m} - U_{i,m}$, where $Z_{i,m}$ is defined in (7.6) and $U_{i,m}$ in (7.21). Then

$$n^{1/2}\{\widetilde{F}_{n,m}(y) - \mathcal{B}_m(F)(y) - B_{n,m}\} = n^{-1/2}\sum_{i=1}^{n} W_{i,m}.$$

The $W_{i,m}$'s are iid and centered, and $\mathsf{Var}(W_{1,m}) \to \nu^2(y) > 0$ as $n \to \infty$ by Proposition 7.

We verify the Lindeberg condition for $W_{1,m}$: for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{\mathsf{Var}(W_{1,m})}\mathsf{E}\left[W_{1,m}^2 \mathbb{1}_{\{|W_{1,m}|^2 > \varepsilon n \,\mathsf{Var}(W_{1,m})\}}\right] = 0. \tag{7.22}$$

By Assumption (A1), $\pi_i(\boldsymbol{X}_i) \geq \pi_{\min} > 0$. Using $\sum_{k=0}^{m} b_{m,k}(y) = 1$ and $0 \leq \mathcal{B}_m(F)(y) \leq 1$,

$$|Z_{1,m}| \leq \pi_{\min}^{-1} + 1, \qquad |U_{1,m}| \leq \pi_{\min}^{-1},$$

so $|W_{1,m}| \leq 2\pi_{\min}^{-1} + 1 < \infty$. Since $\mathsf{Var}(W_{1,m}) \to \nu^2(y) > 0$, we have $\varepsilon n \,\mathsf{Var}(W_{1,m}) \to \infty$, hence for all sufficiently large $n$, the indicator $\mathbb{1}_{\{|W_{1,m}|^2 > \varepsilon n \,\mathsf{Var}(W_{1,m})\}}$ is equal to zero almost surely. Therefore the expectation in (7.22) is eventually 0, and the Lindeberg condition holds. □

## Appendix A. Reproducibility

The R code that generated the figures, the simulation study results, and the real-data application is available online in the GitHub repository of Gharbi *et al.* (2025).

## Appendix B. List of acronyms

| | |
|---|---|
| CDF | cumulative distribution function |
| iid | independent and identically distributed |
| IPW | inverse probability weighting/weighted |
| ISE | integrated squared error |
| LSCV | least-squares cross-validation |
| MAR | missing at random |
| MISE | mean integrated squared error |
| MSE | mean squared error |
| NHANES | National Health and Nutrition Examination Survey |
| NMAR | not missing at random (nonignorable) |

## References


Babu, G. J., & 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.

Babu, G. J., Canty, A. J., & 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.

Belalia, M., Bouezmarni, T., & Leblanc, A. 2017. Smooth conditional distribution estimators using Bernstein polynomials. *Comput. Statist. Data Anal.*, **111**, 166–182. MR3630225.

Bernstein, S. 1912. Démonstration du théorème de Weierstrass, fondée sur le calcul des probabilités. *Commun. Soc. Math. Kharkow*, **2**(13), 1–2.

Bustamante, J. 2017. *Bernstein Operators and Their Properties*. Birkhäuser/Springer, Cham. MR3585481.

Cheng, M.-Y., & Peng, L. 2002. Regression modeling for nonparametric estimation of distribution and quantile functions. *Statist. Sinica*, **12**(4), 1043–1060. MR1379049.

Cheng, P. E., & Chu, C. K. 1996. Kernel estimation of distribution functions and quantiles with missing data. *Statist. Sinica*, **6**(1), 63–78. MR1379049.

Ding, X., & Tang, N. 2018. Adjusted empirical likelihood estimation of distribution function and quantile with nonignorable missing data. *J. Syst. Sci. Complex.*, **31**(3), 820–840. MR3782668.

Dubnicka, S. R. 2009. Kernel density estimation with missing data and auxiliary variables. *Aust. N. Z. J. Stat.*, **51**(3), 247–270. MR2569798.

Endres, C., Ale, L., Gentleman, R., & Sarkar, D. 2025. *nhanesA: NHANES Data Retrieval*. R package version 1.4. doi:10.32614/CRAN.package.nhanesA.

Erdoğan, M. S., Dişibüyük, Ç., & Ege Oruç, Ö. 2019. An alternative distribution function estimation method using rational Bernstein polynomials. *J. Comput. Appl. Math.*, **353**, 232–242. MR3899096.

Gharbi, R., Jedidi, W., Khardani, S., & Ouimet, F. 2025. *BernsteinCDFMissingData*. GitHub repository available online at https://github.com/FredericOuimetMcGill.

Hanebeck, A., & Klar, B. 2021. Smooth distribution function estimation for lifetime distributions using Szasz-Mirakyan operators. *Ann. Inst. Statist. Math.*, **73**(6), 1229–1247. MR4330318.

Horvitz, D. G., & Thompson, D. J. 1952. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, **47**, 663–685. MR53460.

Jmaei, A., Slaoui, Y., & Dellagi, W. 2017. Recursive distribution estimator defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr. Stat.*, **29**(4), 792–805. MR3740720.

Khardani, S. 2024. A Bernstein polynomial approach to the estimation of a distribution function and quantiles under censorship model. *Comm. Statist. Theory Methods*, **53**(16), 5673–5686. MR4765928.

Lafaye de Micheaux, P., & Ouimet, F. 2021. A study of seven asymmetric kernels for the estimation of cumulative distribution functions. *Mathematics*, **9**(20), Paper No. 2605, 31 pp. doi:10.3390/math9202605.

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.

Little, R. J. A., & Rubin, D. B. 2002. *Statistical Analysis With Missing Data*. Second edn. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. MR1925014.

Mansouri, B., Rustin, A., & Allah Mombeni, H. 2023. Beta kernel estimator for a cumulative distribution function with bounded support. *Journal of Sciences, Islamic Republic of Iran*, **34(4)**, 349–361. http://jsciences.ut.ac.ir.

Nadaraya, E. A. 1964. Some new estimates for distribution functions. *Teor. Verojatnost. i Primenen.*, **9**, 550–554. MR166862.

National Center for Health Statistics. 2020. *2017–2018 Demographics Data – Continuous NHANES*. Centers for Disease Control and Prevention (CDC). Data file: DEMO_J; Date published: February 2020. https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Demographics&Cycle=2017-2018.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., & Clark, C. W. (eds). 2010. *NIST Handbook of Mathematical Functions*. Cambridge University Press. Available online at the NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov.

Ouimet, F. 2021. Asymptotic properties of Bernstein estimators on the simplex. *J. Multivariate Anal.*, **185**, Paper No. 104784, 20 pp. MR4287788.

Parzen, E. 1962. On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076. MR143282.

Rosenbaum, P. R., & Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), 41–55. MR742974.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–837. MR742974.

Rubin, D. B. 1976. Inference and missing data. *Biometrika*, **63**(3), 581–592. MR455196.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR0595165.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. MR848134.

Slaoui, Y. 2022. Moderate deviation principles for nonparametric recursive distribution estimators using Bernstein polynomials. *Rev. Mat. Complut.*, **35**(1), 147–158. MR4365177.

Tenreiro, C. 2013. Boundary kernels for distribution function estimation. *REVSTAT*, **11**(2), 169–190. MR3072469.

Vitale, R. A. 1975. Bernstein polynomial approach to density function estimation. *Pages 87–99 of: Statistical Inference and Related Topics*. Academic Press, New York. MR0397977.

Wand, M. P., & Jones, M. C. 1995. *Kernel Smoothing*. Monographs on Statistics and Applied Probability, vol. 60. Chapman and Hall, Ltd., London. MR1319818.

Wang, Q., & Qin, Y. 2010. Empirical likelihood confidence bands for distribution functions with missing responses. *J. Statist. Plann. Inference*, **140**(9), 2778–2789. MR2644095.

Weierstraß, K. 1885. Über die analytische Darstellung sogenanter willkürlicher Functionen einer reelen Veränderlichen. *Verl. d. Kgl. Akad. d. Wiss. Berlin*, **2**, 633–639.

Xue, L., & Wang, J. 2010. Distribution function estimation by constrained polynomial spline regression. *J. Nonparametr. Stat.*, **22**(3-4), 443–457. MR2662606.

Zhang, S., Li, Z., & Zhang, Z. 2020. Estimating a Distribution Function at the Boundary. *Austrian Journal of Statistics*, **49**(1), 1–23. doi:10.17713/ajs.v49i1.801.