Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Partition-Based Probability Measures for Discontinuous and Mixed Distributions

STHITADHI DAS

sthdas999@gmail.com

Brainware University https://orcid.org/0009-0001-2484-9530

Research Article

Keywords: Partition-based probability measure, discontinuous distribution, mixed probability distribution, weak convergence, uniform CDF convergence, Wasserstein distance, statistical consistency, adaptive partitioning

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11539234/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Partition-Based Probability Measures for Discontinuous and Mixed Distributions

Sthitadhi Das
Department of Mathematics
Brainware University, Kolkata, India

Abstract

Conventional parametric probability models provide parsimonious representations of probability laws but may be inadequate when the underlying distribution exhibits abrupt changes in its density, multiple components, isolated atoms, or a combination of discrete and continuous probability. This paper develops a partition-based framework for constructing and estimating probability measures without requiring a single prescribed functional form for the entire distribution. The construction assigns probability masses to measurable cells and permits discrete, continuous, and mixed conditional distributions within the cells. We investigate the approximation of arbitrary probability measures by partition-induced discrete measures, establish weak convergence under partition refinement, and derive statistical consistency for estimated partition distributions. Particular attention is given to the distinction between weak convergence and uniform convergence of cumulative distribution functions when the target distribution has atoms. An atom-preserving approximation scheme is introduced for which uniform CDF approximation is controlled by the maximum probability of the cells used to approximate the nonatomic component. We further establish finite-partition asymptotic normality, derive a Wasserstein error bound, and formulate conditions for consistency when partition boundaries are estimated. A simulation design is proposed to investigate the finite-sample behaviour of the construction under continuous, discontinuous-density, discrete, and mixed probability laws. The results provide a unified mathematical basis for studying partition-based approximation and statistical inference for probability distributions with irregular structure.

Keywords: Partition-based probability measure; discontinuous distribution; mixed probability distribution; weak convergence; uniform CDF convergence; Wasserstein distance; statistical consistency; adaptive partitioning.

Mathematics Subject Classification (2020): 60A10; 60B10; 62G05; 62G20.

1 Introduction

Probability distributions are fundamental to statistical modelling, estimation, prediction, and inference. In many applications, the underlying distribution is represented by a parametric family whose mathematical structure is specified in advance. Normal, exponential, gamma, Weibull, and beta distributions provide useful models for a broad range of phenomena. Their parsimonious representations facilitate estimation and inference, particularly when the assumed model is

appropriate. However, a single parametric family may not adequately represent distributions with multiple modes, abrupt changes in density, structural point masses, or heterogeneous components. Such structures arise in reliability analysis, environmental measurements, insurance claims, survival data, and populations containing both zero-valued and continuously distributed observations. A density may be discontinuous even though its CDF is continuous, whereas a discrete probability component produces genuine jumps in the CDF. Mixed distributions may exhibit both features simultaneously.

Nonparametric methods provide alternatives to restrictive parametric assumptions. Empirical distribution functions, histogram estimators, kernel estimators, and adaptive partitioning methods have been extensively studied; see, for example, Devroye and Lugosi (2001), Györfi et al. (2002), and Wasserman (2006). These approaches differ in their approximation properties and in the assumptions required for their statistical analysis. In particular, density estimation with respect to Lebesgue measure does not directly describe atomic probability, and weak convergence of probability measures does not necessarily imply convergence of their CDFs at discontinuity points. These observations motivate a general construction based on measurable partitions of the sample space. Instead of prescribing a single global density, the construction allocates probability to disjoint regions and permits a different conditional probability law within each region. The resulting model can represent discrete, continuous, and mixed distributions within a common measure-theoretic framework.

The purpose of this paper is to formulate the construction and develop its approximation and statistical convergence properties. The principal questions are as follows. First, under what conditions does a partition-induced probability measure approximate an arbitrary target probability measure? Second, when cell probabilities are estimated from data, what conditions ensure statistical consistency? Third, how does the presence of atoms affect convergence of the corresponding CDFs? Finally, can an atom-preserving construction achieve uniform CDF approximation while retaining a transparent separation between approximation error and sampling error? The principal theoretical emphasis is on the last two questions. Weak convergence is suitable for comparing probability measures of different types, but it may conceal substantial discrepancies at jump points. We therefore develop an atom-preserving approximation scheme in which the atomic component is retained at its original locations and the nonatomic component is approximated through increasingly fine partitions. This leads to a direct uniform CDF error bound.

The main contributions are fourfold. A general partition-based probability construction is formulated; approximation and statistical consistency results are established; uniform CDF convergence is investigated for distributions with atoms; and finite-partition inference and convergence-rate bounds are derived. A framework for data-dependent partition selection is also discussed.

The remainder of the paper is organized as follows. Section 2 introduces the probability construction. Section 3 studies approximation by partition refinement. Section 4 develops statistical estimation and convergence rates. Section 5 investigates discontinuities and atom-preserving approximation. Section 6 presents finite-partition inference and adaptive partitioning. Section 7 describes a simulation study. Section 8 concludes the paper.

2 Construction of partition-based probability measures

2.1 General probability construction

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{X}$ be a measurable subset of $\mathbb{R}$ equipped with its Borel sigma-field. The target distribution is represented by a probability measure P on $\mathcal{X}$.

For a positive integer K, consider a finite measurable partition

$$\mathcal{P}_K = \{A_1, \ldots, A_K\},$$

where $A_j \cap A_k = \varnothing,\ j \neq k$, and $\bigcup_{j=1}^{K} A_j = \mathcal{X}$. The corresponding cell probabilities are

$$p_j = P(A_j)\ ;\ j = 1, \ldots, K.$$

They satisfy $p_j \geq 0,\ \sum_{j=1}^{K} p_j = 1$.

Let Q_j be a probability measure supported on A_j. The associated partition-based probability measure is defined by

$$P_{\mathcal{P}}(B) = \sum_{j=1}^{K} p_j Q_j(B), \qquad B \in \mathcal{B}(\mathcal{X}). \tag{2.1}$$

The construction separates the allocation of probability across cells from the conditional distribution within each cell. The probabilities p_j determine the mass assigned to the different regions, while Q_j determines the local probability structure.

Proposition 2.1 (Validity of the construction)**.** *For every finite measurable partition $\mathcal{P}_K$, every collection of nonnegative cell probabilities satisfying $\sum_{j=1}^{K} p_j = 1$, and every collection of probability measures Q_j supported on A_j, the set function $P_{\mathcal{P}}$ defined in* (2.1) *is a probability measure.*

2.2 Discrete partition approximation

A particularly simple construction replaces each cell by a representative point. Choose $a_j \in A_j$ and let δ_{a_j} denote the Dirac probability measure at a_j. Define

$$P_K = \sum_{j=1}^{K} p_j \delta_{a_j}. \tag{2.2}$$

This is a discrete probability measure, irrespective of whether the target distribution is continuous, discrete, or mixed. For a real-valued random variable X with distribution P, the corresponding quantization map is $T_K(x) = a_j \quad \text{if } x \in A_j$.

If $X \sim P$, then $T_K(X) \sim P_K$.

The construction in (2.2) is useful for studying probability-measure approximation. It should be distinguished from a piecewise-constant density, which is a different object and requires an appropriate reference measure.

2.3 Partition-based density construction

Suppose the target distribution is absolutely continuous with respect to Lebesgue measure and the cells are intervals of positive finite lengths h_j. A piecewise-constant density can be defined by

$$f_K(x) = \sum_{j=1}^{K} \frac{p_j}{h_j} \mathbf{1}_{A_j}(x). \tag{2.3}$$

Since the intervals form a partition of the support,

$$\int f_K(x)\, dx = \sum_{j=1}^{K} \frac{p_j}{h_j} h_j = 1.$$

Thus f_K is a valid probability density.

This density construction is appropriate for continuous probability components. It cannot represent an atomic component by an ordinary Lebesgue density. For mixed distributions, the measure construction in (2.1) is therefore more general.

2.4 Partition-based CDF

For the discrete approximation in (2.2), the CDF is

$$F_K(x) = P_K((-\infty, x]) = \sum_{j=1}^{K} p_j \mathbf{1}\{a_j \le x\}. \tag{2.4}$$

The target CDF is

$$F(x) = P((-\infty, x]). \tag{2.5}$$

The distinction between F_K and F is central to the convergence analysis developed below.

3 Approximation under partition refinement

Let $\{\mathcal{P}_K\}$ be a sequence of partitions of a compact interval $\mathcal{X} \subset \mathbb{R}$. Write

$$\delta_K = \max_{1 \le j \le K} \operatorname{diam}(A_j)$$

for the maximum cell diameter. The partitions are assumed to become increasingly fine, so that $\delta_K \to 0$.

Theorem 3.1 (Weak convergence of the exact partition approximation)**.** *Suppose P is supported on a compact interval $\mathcal{X}$. If $\delta_K \to 0$, then $P_K \Rightarrow P$. Moreover, for every continuous function g on $\mathcal{X}$,*

$$\left| \int g\, dP_K - \int g\, dP \right| \le \omega_g(\delta_K), \tag{3.1}$$

where $\omega_g(t) = \sup_{\substack{x,y \in \mathcal{X} \\ |x-y| \le t}} |g(x) - g(y)|$ is the modulus of continuity of g.

Theorem 3.1 does not require the target distribution to possess a density. It therefore applies to continuous, discrete, and mixed probability measures. For unbounded supports, a corresponding result requires additional tail control or tightness conditions.

3.1 CDF approximation error

The weak convergence result does not, by itself, guarantee convergence of CDFs at every point. We first give an explicit expression for the CDF approximation error.

Proposition 3.2 (CDF error under quantization)**.** *Let $T_K(x) = a_j$ for $x \in A_j$. Then*

$$F_K(x) - F(x) = \int \left[\mathbf{1}\{T_K(y) \le x\} - \mathbf{1}\{y \le x\}\right] dP(y). \tag{3.2}$$

Consequently,

$$|F_K(x) - F(x)| \le P\left(\mathbf{1}\{T_K(X) \le x\} \ne \mathbf{1}\{X \le x\}\right). \tag{3.3}$$

Corollary 3.3 (CDF convergence at continuity points)**.** *Under the assumptions of Theorem 3.1,*

$$F_K(x) \longrightarrow F(x)$$

at every continuity point x of F.

At a jump point of F, weak convergence alone does not ensure convergence of the CDF values. For example, if $P = \delta_0$ and $P_K = \delta_{1/K}$, then $P_K \Rightarrow \delta_0$, but $F_K(0) = 0$ whereas $F(0) = 1$. This motivates an atom-preserving construction in Section 5.

4 Statistical estimation and convergence rates

4.1 Empirical cell probabilities

Let $X_1, \dots, X_n$ be i.i.d. observations from P. For a partition $\mathcal{P}_K$, define

$$N_j = \sum_{i=1}^{n} \mathbf{1}\{X_i \in A_j\}, \qquad \widehat{p}_j = \frac{N_j}{n}.$$

The empirical partition-based measure is

$$\widehat{P}_{n,K} = \sum_{j=1}^{K} \widehat{p}_j \delta_{a_j}, \tag{4.1}$$

with CDF

$$\widehat{F}_{n,K}(x) = \sum_{j=1}^{K} \widehat{p}_j \mathbf{1}\{a_j \le x\}. \tag{4.2}$$

The estimation error can be decomposed into a sampling component and a partition approxi-

mation component. For example, for a bounded continuous function g,

$$\left|\int g\,d\widehat{P}_{n,K} - \int g\,dP\right| \le \left|\int g\,d\widehat{P}_{n,K} - \int g\,dP_K\right| + \left|\int g\,dP_K - \int g\,dP\right|.$$

The first term is controlled by the empirical cell probabilities, while the second is controlled by partition refinement.

Lemma 4.1 (Expected cell-probability error)**.** *For a fixed partition with K cells,*

$$\mathbb{E}\left[\sum_{j=1}^{K} |\widehat{p}_j - p_j|\right] \le \sqrt{\frac{K}{n}}. \tag{4.3}$$

Theorem 4.2 (Weak consistency of the estimated partition measure)**.** *Suppose P is supported on a compact interval and $\mathcal{P}_{K_n}$ is a deterministic sequence of partitions such that*

$$\delta_{K_n} \longrightarrow 0, \qquad \frac{K_n}{n} \longrightarrow 0.$$

Then

$$\widehat{P}_{n,K_n} \Rightarrow P \quad \textit{in probability.}$$

4.2 Wasserstein convergence

For probability measures P and Q on $\mathbb{R}$ with finite first moments, define the Wasserstein-1 distance by

$$W_1(P,Q) = \inf_{\pi\in\Pi(P,Q)} \int_{\mathbb{R}^2} |x-y|\,d\pi(x,y),$$

where $\Pi(P,Q)$ denotes the collection of couplings of P and Q. On the real line, it also admits the representation

$$W_1(P,Q) = \int_{\mathbb{R}} |F_P(x) - F_Q(x)|\,dx.$$

Theorem 4.3 (Wasserstein error bound)**.** *Suppose the support of P is contained in an interval of length L. Let δ_K be the maximum cell diameter. Then*

$$W_1(P,\widehat{P}_{n,K}) \le \delta_K + L\sum_{j=1}^{K} |\widehat{p}_j - p_j|. \tag{4.4}$$

Consequently,

$$W_1(P,\widehat{P}_{n,K}) = O_{\mathbb{P}}\left(\delta_K + \sqrt{\frac{K}{n}}\right). \tag{4.5}$$

For equal-width partitions of an interval of length L, $\delta_K = L/K$. The preceding bound becomes

$$W_1(P,\widehat{P}_{n,K}) = O_{\mathbb{P}}\left(\frac{L}{K} + \sqrt{\frac{K}{n}}\right).$$

Balancing the two terms gives the illustrative choice $K \asymp n^{1/3}$ and the bound $W_1(P, \widehat{P}_{n,K}) = O_{\mathbb{P}}(n^{-1/3})$ for fixed L. This is a rate implied by the stated bound, not a claim of minimax optimality.

5 Discontinuities and atom-preserving approximation

5.1 Atomic and nonatomic components

Every probability measure P on $\mathbb{R}$ admits a decomposition

$$P = P_{\mathrm{at}} + P_{\mathrm{na}},$$

where

$$P_{\mathrm{at}} = \sum_{r \geq 1} \alpha_r \delta_{z_r}, \qquad \alpha_r = P(\{z_r\}),$$

and P_{na} is nonatomic. The total mass of P_{at} may be less than one; the remaining mass is carried by P_{na}.

The CDF of P is

$$F(x) = F_{\mathrm{at}}(x) + F_{\mathrm{na}}(x),$$

where

$$F_{\mathrm{at}}(x) = \sum_{r \geq 1} \alpha_r \mathbf{1}\{z_r \leq x\}.$$

The jump size at any point x is

$$\Delta F(x) = F(x) - F(x-) = P(\{x\}).$$

The key idea is to preserve the atomic part at its original locations and approximate only the nonatomic component.

5.2 Atom-preserving construction

Let $\mathcal{P}_K = \{A_1, \ldots, A_K\}$ be a finite partition of the nonatomic part of the support, and choose $a_j \in A_j$. Define

$$q_K = \max_{1 \leq j \leq K} P_{\mathrm{na}}(A_j).$$

The atom-preserving approximation is

$$P_K^\star = P_{\mathrm{at}} + \sum_{j=1}^{K} P_{\mathrm{na}}(A_j) \delta_{a_j}. \tag{5.1}$$

The atomic part is retained exactly, and the nonatomic component is replaced by a discrete approximation.

Theorem 5.1 (Uniform CDF approximation with atom preservation)**.** *Suppose the target*

probability measure has the decomposition

$$P = P_{\mathrm{at}} + P_{\mathrm{na}},$$

and construct $P_K^\star$ *as in* (5.1)*. Then*

$$\sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| \le q_K. \tag{5.2}$$

Consequently, if $q_K \to 0$*, then*

$$\sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| \to 0,$$

even when F *has jumps.*

Theorem 5.1 distinguishes geometric refinement from probability refinement. Small geometric diameters alone do not provide the displayed uniform bound for an arbitrary measure if atoms are moved between cells. Preserving atoms and controlling the maximum nonatomic cell mass directly address the issue.

5.3 Uniform convergence of the estimated CDF

Suppose the atomic locations are known, and the remaining support is partitioned into K interval cells. Estimate all cell probabilities, including the probabilities of the atomic cells, using empirical frequencies. Denote the resulting atom-preserving estimator by $\widehat{P}_{n,K}^\star$, with CDF $\widehat{F}_{n,K}^\star$.

Theorem 5.2 (Uniform statistical consistency)**.** *Under the preceding construction,*

$$\sup_{x\in\mathbb{R}} |\widehat{F}_{n,K}^\star(x) - F(x)| \le q_K + \sum_{j=1}^{K} |\widehat{p}_j - p_j|, \tag{5.3}$$

where the sum includes all estimated cell probabilities, including any separately represented atoms. If

$$q_{K_n} \to 0, \qquad \frac{K_n}{n} \to 0,$$

then

$$\sup_{x\in\mathbb{R}} |\widehat{F}_{n,K_n}^\star(x) - F(x)| \xrightarrow{\mathbb{P}} 0.$$

The theorem presumes that the atomic component is represented at its true locations. When atom locations are unknown, consistent identification and estimation require additional assumptions and a separate analysis. The theorem does not establish that arbitrary unknown atoms can be detected without restrictions.

5.4 Estimation of jump sizes

Lemma 5.3 (Consistency and asymptotic normality of a fixed jump)**.** *Suppose* x_0 *is a fixed point with* $p_0 = P(\{x_0\})$*. Let* $\widehat{\Delta}_n(x_0) = \widehat{F}_n(x_0) - \widehat{F}_n(x_0-)$*, where* $\widehat{F}_n$ *is the empirical CDF. Then,* $\widehat{\Delta}_n(x_0) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbf{1}\{X_i = x_0\}$*. Consequently,* $\widehat{\Delta}_n(x_0) \xrightarrow{\text{a.s.}} p_0$*.*

If $0 < p_0 < 1$, *then* $\sqrt{n}\{\widehat{\Delta}_n(x_0) - p_0\} \xrightarrow{d} N(0, p_0(1-p_0))$.

5.5 Stability with respect to probability estimation

Proposition 5.4 (CDF stability)**.** *Let* P_K *and* Q_K *be supported on the same ordered representative points* $a_1, \ldots, a_K$, *with respective probabilities* p_j *and* q_j. *Then*

$$\sup_{x \in \mathbb{R}} |F_{P_K}(x) - F_{Q_K}(x)| \leq \sum_{j=1}^{K} |p_j - q_j|.$$

6 Finite-partition inference and adaptive refinement

6.1 Asymptotic normality for a fixed partition

Theorem 6.1 (Multinomial asymptotic normality)**.** *For a fixed partition with* K *cells, let*

$$\widehat{\boldsymbol{p}} = (\widehat{p}_1, \ldots, \widehat{p}_K)^\top, \qquad \boldsymbol{p} = (p_1, \ldots, p_K)^\top.$$

Then, $\sqrt{n}(\widehat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{d} N_K(\mathbf{0}, \Sigma)$, *where* $\Sigma = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top$.

For a fixed x, let $\boldsymbol{c}(x) = (\mathbf{1}\{a_1 \leq x\}, \ldots, \mathbf{1}\{a_K \leq x\})^\top$. It follows that

$$\sqrt{n}\{\widehat{F}_{n,K}(x) - F_K(x)\} \xrightarrow{d} N\left(0, \boldsymbol{c}(x)^\top \Sigma \boldsymbol{c}(x)\right).$$

Because the representative points are ordered, the selected cells form a prefix of the partition, and the limiting variance is $F_K(x)\{1 - F_K(x)\}$. This result concerns a fixed partition. When the number of cells increases with sample size, a growing-dimensional central limit theorem or an empirical-process argument is needed.

6.2 Estimated partition boundaries

Let $b_0 < b_1 < \cdots < b_K$ denote the boundaries of a fixed partition, and let $\widehat{b}_j$ be estimators of the internal boundaries.

Proposition 6.2 (Consistency under estimated boundaries)**.** *Suppose* K *is fixed,*

$$\max_{1 \leq j < K} |\widehat{b}_j - b_j| \xrightarrow{\mathbb{P}} 0,$$

and $P(\{b_j\}) = 0, \qquad j = 1, \ldots, K-1$. *Then the empirical probabilities of the cells formed by the estimated boundaries converge in probability to the corresponding true cell probabilities.*

If a true boundary coincides with an atom, the zero-boundary- mass condition fails. In that case, the partition convention and the treatment of the atom must be specified explicitly. For a refining sequence of estimated partitions, further conditions are required to control the maximum cell diameter, the number of cells, and the cumulative boundary-estimation error.

6.3 A bias–variance principle for partition selection

The preceding results suggest choosing partition complexity by balancing approximation and estimation errors. A generic criterion is

$$\mathcal{R}_n(K) = \mathcal{A}(K) + \mathcal{E}_n(K), \tag{6.1}$$

where $\mathcal{A}(K)$ measures the approximation error and $\mathcal{E}_n(K)$ measures the estimation error.

For the Wasserstein bound in Theorem 4.3, one may use the upper-bound proxy

$$\mathcal{R}_n(K) = \delta_K + L\sqrt{\frac{K}{n}}.$$

This is a conservative selection criterion rather than a universal optimality result. An adaptive method should also consider whether the partition separates important atoms and whether the cells around suspected discontinuities are sufficiently fine.

A rigorous optimal partition theorem would require a specified class of target distributions, a loss function, and a precise approximation class. These elements determine whether a rate is minimax optimal or merely an upper bound.

7 Simulation study design

7.1 Objectives

The simulation study is intended to assess how partition refinement and atom preservation affect distributional approximation and statistical estimation. It should investigate whether the theoretical error bounds are informative in finite samples and how the proposed method compares with conventional empirical and parametric estimators.

No numerical results are asserted in this section. They should be reported only after the simulation has been implemented and the output verified.

7.2 Data-generating distributions

The following distributions provide complementary test cases.

7.2.1 Model I: Uniform distribution.

Generate observations from $U(0, 1)$. This provides a continuous benchmark with no density discontinuity.

7.2.2 Model II: Discontinuous density.

Consider

$$f_2(x) = \begin{cases} 1/3, & 0 \le x < 1, \\ 2/3, & 1 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

The density integrates to one and has a jump at $x = 1$. Its CDF is continuous.

7.2.3 Model III: Discrete distribution.

Let $P(X = 0) = 0.4,\ P(X = 1) = 0.6$. This model evaluates the preservation and estimation of atomic probabilities.

7.2.4 Model IV: Mixed distribution.

Let $P(X = 0) = 0.3$ and let the remaining probability 0.7 follow a uniform distribution on $(0, 1)$. This model has an atom at zero and a continuous component.

7.2.5 Model V: Multimodal continuous distribution.

Use a finite mixture of normal distributions with separated component means. This assesses the ability of a partition to represent multiple modes without specifying a single parametric family.

7.3 Estimators to compare

The study should compare the following procedures:

(i) the empirical CDF;

(ii) a partition-based discrete probability measure;

(iii) a piecewise-constant density estimator, where a density exists;

(iv) an atom-preserving partition estimator when the atomic locations are known;

(v) an ordinary histogram with a data-selected bin width;

(vi) a correctly specified parametric estimator for each benchmark model, where available.

The comparison between the ordinary partition estimator and the atom-preserving estimator is particularly important for the discrete and mixed models.

7.4 Performance criteria

For replication r, define the uniform CDF error by $D_r = \sup_{x \in \mathcal{G}} |\widehat{F}_r(x) - F(x)|$, where $\mathcal{G}$ is a sufficiently fine deterministic grid augmented with all known discontinuity locations. The grid approximation to the supremum should be checked for sensitivity to grid refinement.

For models with a finite first moment, compute $W_1(\widehat{P}_r, P) = \int_{\mathbb{R}} |\widehat{F}_r(x) - F(x)|\, dx$. For density estimators, use integrated squared error $\mathrm{ISE}_r = \int \{\widehat{f}_r(x) - f(x)\}^2\, dx$, where the target possesses a density. The simulation should also record the estimated jump sizes, their biases, their mean squared errors, and the computational cost of each procedure.

7.5 Sample sizes and partition complexity

Consider sample sizes $n \in \{50, 100, 200, 500, 1000\}$ and several partition sizes satisfying both moderate and fine resolutions. For each configuration, repeat the experiment independently, for example, $B = 2000$ times. The empirical convergence study should examine whether the estimation error decreases as n increases and whether overly fine partitions introduce excessive sampling variability. A separate experiment should investigate the effect of preserving known atoms when the target distribution contains point masses.

7.6 Results & Interpretation

The finite-sample performance of the proposed partition-based probability construction was investigated through 2000 Monte Carlo replications for each sample size $n \in \{50, 100, 200, 500, 1000\}$. The analysis considered five distributions with different structural characteristics: a uniform distribution, a distribution with a discontinuous density, a discrete distribution, a mixed distribution, and a multimodal continuous distribution. The principal performance criteria were the uniform error of the estimated cumulative distribution function (CDF), the first Wasserstein distance, the integrated squared error (ISE) of density estimators, and the mean squared error of estimated atomic probabilities. The numerical findings are discussed below, with particular attention to the effects of sample size, partition resolution, and explicit preservation of atoms.

7.6.1 Uniform distribution

The uniform distribution serves as a continuous benchmark for examining the approximation error induced by partitioning and the effect of sample size on distributional and density estimation. Figure 1 presents the mean CDF error for the empirical CDF, the correctly specified parametric estimator, and partition-based approximations with different numbers of cells. The corresponding numerical results, including the mean Wasserstein-1 distance and the integrated squared error (ISE) of the density estimators, are summarized in Table 1. Together, these results illustrate the effects of sample size and partition resolution on the quality of the estimated distribution.

The empirical CDF exhibited a consistent improvement as the sample size increased. Its mean CDF error decreased from 0.12003 at $n = 50$ to 0.02729 at $n = 1000$, while the mean Wasserstein-1 distance declined from 0.04503 to 0.00998, respectively (Table 1). This pattern is consistent with the convergence of the empirical distribution function to the underlying distribution as the sample size increases. The reduction in Wasserstein distance also indicates that the empirical distribution becomes closer to the target distribution in terms of the average transport distance. The downward trend in Figure 1 provides a visual confirmation of the reduction in CDF error with increasing sample size.

The partition-based estimators show that the number of cells plays an important role in the accuracy of the discrete approximation. At $n = 50$, increasing the number of cells from $K = 2$ to $K = 8$ reduced the mean CDF error from 0.30681 to 0.15700. At $n = 1000$, the corresponding error decreased from 0.11900 for $K = 5$ to 0.04833 for $K = 20$. The Wasserstein-1 distance followed the same pattern at $n = 1000$, decreasing from 0.05080 for $K = 5$ to 0.01577 for $K = 20$ (Table 1). These findings indicate that partition refinement improves the representation of the

continuous distribution by providing a finer spatial resolution. Nevertheless, the partition-based approximations remain less accurate than the empirical CDF for the reported configurations, emphasizing that finer partitioning does not automatically eliminate the approximation error associated with replacing a continuous distribution by a discrete measure.

The correctly specified parametric estimator achieved smaller errors than both the empirical CDF and the reported partition-based estimators. Its mean CDF error decreased from 0.02997 at $n = 50$ to 0.00149 at $n = 1000$, while its mean Wasserstein-1 distance decreased from 0.01332 to 0.00066 (Table 1). This result illustrates the efficiency attainable when the parametric family is correctly specified. The comparison should, however, be interpreted in light of the benchmark design: the parametric estimator benefits from knowledge of the correct distributional form, whereas partition-based methods provide a more flexible representation that does not require the target distribution to belong to a prespecified parametric family. Accordingly, the purpose of the partition-based construction is not necessarily to outperform a correctly specified parametric estimator, but to provide a systematic approximation that can also be used when the distributional form is unknown or more complicated.

The density-estimation results reveal a different relationship between partition resolution and estimation accuracy. At $n = 1000$, the mean ISE of the piecewise-constant density estimator was 0.00394, 0.00895, and 0.01901 for $K = 5$, $K = 10$, and $K = 20$, respectively (Table 1). Thus, in this experiment, the coarser partition with $K = 5$ produced the smallest ISE among these three choices. This does not contradict the improvement in CDF error and Wasserstein distance under partition refinement: the metrics assess different properties of the estimators. A finer partition can represent the distribution at greater spatial resolution, while simultaneously producing less stable estimates of the density within individual cells. With a fixed sample size, increasing the number of cells generally reduces the number of observations available to estimate each cell probability, potentially increasing the variability of the estimated density heights. The results therefore demonstrate that partition selection should account for the intended inferential target and should balance approximation resolution against sampling variability.

Overall, the uniform-distribution experiment demonstrates that increasing the sample size improves empirical distribution estimation and that partition refinement generally reduces the discrepancy between the partition-based discrete measure and the target distribution. However, the density-estimation results show that the finest partition need not minimize ISE. Figure 1 and Table 1 jointly highlight the distinction between approximating a distribution through a discrete measure and estimating its underlying density. This distinction is important when selecting the partition size for practical applications.

Table 1: Selected simulation results for the uniform distribution. The table reports mean CDF error, mean Wasserstein-1 distance, and mean integrated squared error (ISE), where available.

Criterion	Estimator	n	K	Mean error
CDF error	Empirical CDF	50	—	0.12003
CDF error	Empirical CDF	1000	—	0.02729
Wasserstein-1	Empirical CDF	50	—	0.04503
Wasserstein-1	Empirical CDF	1000	—	0.00998
CDF error	Parametric	50	—	0.02997
CDF error	Parametric	1000	—	0.00149
Wasserstein-1	Parametric	50	—	0.01332
Wasserstein-1	Parametric	1000	—	0.00066
CDF error	Partition	50	2	0.30681
CDF error	Partition	50	8	0.15700
CDF error	Partition	1000	5	0.11900
CDF error	Partition	1000	20	0.04833
Wasserstein-1	Partition	1000	5	0.05080
Wasserstein-1	Partition	1000	20	0.01577
Density ISE	Piecewise const.	1000	5	0.00394
Density ISE	Piecewise const.	1000	10	0.00895
Density ISE	Piecewise const.	1000	20	0.01901

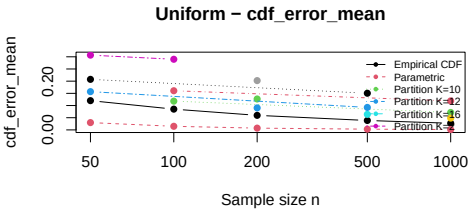


Figure 1: Mean CDF error against sample size for the uniform distribution. The empirical CDF, parametric estimator, and partition-based approximations are compared.

7.6.2 Distribution with a discontinuous density

The second simulation model has density

$$f(x) = \begin{cases} 1/3, & 0 \le x < 1, \\ 2/3, & 1 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Although the density has a jump discontinuity at $x = 1$, the corresponding cumulative distribution function (CDF) is continuous. This model therefore examines the ability of partition-based methods to approximate a continuous distribution when its density changes abruptly across adjacent regions. In particular, it allows us to investigate how sample size, partition resolution, and the placement of partition boundaries affect distributional approximation and density estimation. Figure 2 displays the mean Wasserstein-1 distance, while the numerical results for the CDF error, Wasserstein-1 distance, and density integrated squared error (ISE) are summarized in Table 2.

The empirical CDF improved as the sample size increased. Its mean CDF error decreased from 0.11955 at $n = 50$ to 0.02727 at $n = 1000$, while the mean Wasserstein-1 distance declined from 0.08664 to 0.01924, respectively (Table 2). These results indicate that the empirical distribution increasingly approximates the target distribution as more observations become available. The correctly specified parametric estimator also exhibited decreasing errors: its mean CDF error fell from 0.05364 at $n = 50$ to 0.01216 at $n = 1000$. As in the uniform benchmark, this estimator benefits from the correct specification of the underlying distributional form, whereas

the partition-based construction does not require such a parametric assumption.

The partition-based results demonstrate the importance of partition resolution. At $n = 1000$, increasing the number of cells from $K = 5$ to $K = 20$ reduced the mean CDF error from 0.12790 to 0.05465. The corresponding mean Wasserstein-1 distance decreased from 0.09875 to 0.03137 (Table 2). Figure 2 likewise illustrates the improvement in distributional approximation associated with increasing sample size and partition resolution. These findings support the use of finer partitions when a more accurate discrete representation of the target distribution is required. Nevertheless, the reported partition-based errors remain larger than those of the empirical CDF for the displayed configurations, showing that finer partitions reduce, but do not necessarily remove, the discrepancy between the discrete approximation and the continuous target distribution.

The placement of partition boundaries was also examined. At $n = 100$ with $K = 5$, the mean CDF error was 0.16553 for the partition aligned with the density discontinuity at $x = 1$, compared with 0.18347 for the unaligned partition. Their corresponding mean Wasserstein-1 distances were 0.11266 and 0.11587, respectively (Table 2). Thus, in this configuration, alignment improved both measures of distributional approximation. This is consistent with the fact that a boundary at $x = 1$ prevents a partition cell from combining observations from the two regions with different density levels. The benefit is, however, dependent on the partition geometry and is not universal. For example, the reported $K = 10$ aligned and unaligned results coincide, as do several other configurations in which the procedures produce the same effective partition. Consequently, the alignment effect should be interpreted for each actual partition rather than as a general guarantee of improved approximation.

The density-estimation results further distinguish distributional approximation from density recovery. The data-selected Freedman–Diaconis (FD) histogram had a mean ISE of 0.03192 at $n = 50$, decreasing to 0.00447 at $n = 1000$ (Table 2). For the piecewise-constant density estimator at $n = 1000$, the mean ISE was 0.00209 for $K = 5$ and 0.00947 for $K = 20$. Thus, the coarser of these two partitions yielded the smaller ISE in this experiment. This result does not contradict the improvement in CDF error and Wasserstein distance under partition refinement, because density ISE measures the accuracy of the estimated density heights over the domain, rather than the discrepancy between the estimated and target distribution functions. A larger number of cells offers finer spatial resolution, but it also leaves fewer observations per cell on average, which can increase the variability of the estimated density. Since the true density is piecewise constant, a partition whose boundary coincides with $x = 1$ can represent the two density levels without averaging across the discontinuity. The results therefore suggest that both partition resolution and boundary placement should be considered when estimating a discontinuous density.

Overall, this experiment shows that a jump discontinuity in the density does not prevent accurate estimation of the continuous distribution function. Increasing the sample size improves the empirical CDF and reduces the Wasserstein-1 distance, while partition refinement generally improves the distributional approximation. Boundary alignment can provide an additional improvement when it separates regions with different density levels. However, the density ISE results show that the partition with the largest number of cells need not provide the most

accurate density estimate. Figure 2 and Table 2 together emphasize that the choice of partition should reflect the estimation objective: approximating the distribution, recovering the density, or capturing a known discontinuity.

Table 2: Selected simulation results for the distribution with a discontinuous density. The table reports mean CDF error, mean Wasserstein-1 distance, and mean density integrated squared error (ISE), where available.

Criterion	Estimator	n	K	Mean error
CDF error	Empirical CDF	50	—	0.11955
CDF error	Empirical CDF	1000	—	0.02727
Wasserstein-1	Empirical CDF	50	—	0.08664
Wasserstein-1	Empirical CDF	1000	—	0.01924
CDF error	Parametric	50	—	0.05364
CDF error	Parametric	1000	—	0.01216
CDF error	Partition, aligned	100	5	0.16553
CDF error	Partition, unaligned	100	5	0.18347
Wasserstein-1	Partition, aligned	100	5	0.11266
Wasserstein-1	Partition, unaligned	100	5	0.11587
CDF error	Partition, aligned	1000	5	0.12790
CDF error	Partition, aligned	1000	20	0.05465
Wasserstein-1	Partition, aligned	1000	5	0.09875
Wasserstein-1	Partition, aligned	1000	20	0.03137
Density ISE	FD histogram	50	—	0.03192
Density ISE	FD histogram	1000	—	0.00447
Density ISE	Piecewise const.	1000	5	0.00209
Density ISE	Piecewise const.	1000	20	0.00947

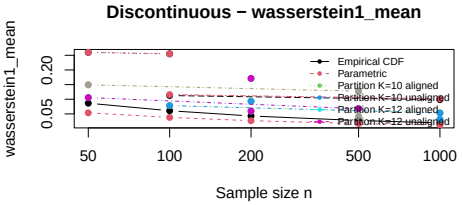


Figure 2: Mean Wasserstein-1 distance against sample size for the distribution with a discontinuous density. The plot compares the empirical CDF, parametric benchmark, and partition-based approximations.

7.6.3 Discrete distribution and preservation of atoms

The discrete model assigns probabilities 0.4 and 0.6 to the points 0 and 1, respectively. Unlike the preceding continuous models, this setting directly examines whether a partition-based construction preserves the locations of point masses. This distinction is fundamental because a discrete distribution is characterized not only by the probabilities assigned to its support points but also by the locations at which the CDF exhibits jumps. Figure 3 presents the mean uniform CDF error for the atom-preserving estimators, while Table 3 reports selected numerical results for the atom-preserving and ordinary partition-based procedures.

In the implementation considered here, the empirical CDF, the atom-preserving estimator, and the model-aware parametric estimator coincide because each uses the observed proportions at the two known support points. Their mean uniform CDF error decreased from 0.05268 at $n = 50$ to 0.01256 at $n = 1000$, while the corresponding mean Wasserstein-1 distance decreased from 0.05269 to 0.01256 (Table 3). The mean squared error (MSE) of the estimated atomic probabilities also declined, from 0.004514 to 0.0002474 over the same sample-size range. These reductions are consistent with the increasing precision of empirical proportions as estimators of the fixed atomic probabilities. The estimated atomic probabilities are unbiased under the

usual independent sampling model, and their variability decreases as the sample size increases. The overlapping curves in Figure 3 reflect the fact that these estimators use the same observed proportions and retain the same support points.

The ordinary partition-based approximation behaved differently. Its mean uniform CDF error remained at 0.6 across the reported sample sizes and partition sizes (Table 3). This persistent discrepancy arises because the ordinary construction assigns the probability associated with a cell to a representative point, which need not coincide with the location of the true atom. Consequently, the approximating CDF can place a jump at a different location from the target CDF. Even when the representative point is relatively close to the true support point, a displacement of an atom can produce a substantial uniform CDF discrepancy: the CDF difference is evaluated at every location, including points between the true and approximating jump locations. Refining the partition can reduce the displacement, but it does not guarantee exact preservation of the atomic support.

The Wasserstein-1 results demonstrate a complementary pattern. For $n = 50$, the mean distance for the ordinary partition approximation decreased from 0.27633 at $K = 2$ to 0.10878 at $K = 8$. At $n = 1000$, it decreased from 0.11014 at $K = 5$ to 0.03695 at $K = 20$ (Table 3). Thus, increasing the partition resolution improved the transport-based approximation even though the uniform CDF error remained unchanged in the reported configurations. The two criteria measure different aspects of discrepancy. The uniform CDF error is sensitive to the maximum difference between the two distribution functions, particularly around displaced jumps, whereas the Wasserstein-1 distance reflects the amount of probability mass that must be transported and the distances over which it is moved. As a result, the transport distance can decline when representative points approach the true atoms, even when the uniform CDF error remains large.

These findings highlight why preserving atoms is important when constructing approximations to discrete distributions. An estimator that retains the true support points and estimates their probabilities directly avoids the artificial displacement of CDF jumps introduced by ordinary cell-representative constructions. In the present experiment, the improvement of the atom-preserving estimators with sample size is attributable to more accurate estimation of the atomic probabilities, rather than to an increase in the number of partition cells. By contrast, refining the ordinary partition improves the Wasserstein approximation but does not resolve the persistent uniform CDF discrepancy in the configurations considered.

Overall, the discrete-distribution experiment demonstrates that distributional approximation requires attention to both probability weights and support locations. The atom-preserving estimators achieve improving CDF and Wasserstein accuracy as the sample size increases, whereas the ordinary partition-based construction exhibits a substantial uniform CDF error despite improvements in Wasserstein distance under partition refinement. Figure 3 and Table 3 therefore support the use of an explicit atom-preservation mechanism when the target distribution contains point masses.

Table 3: Selected simulation results for the discrete distribution with masses 0.4 at 0 and 0.6 at 1. The table reports mean uniform CDF error, mean Wasserstein-1 distance, and the mean squared error of the estimated atomic probabilities (MSE).

Criterion	Estimator	n	K	Mean error
Uniform CDF error	Atom-preserving	50	—	0.05268
Uniform CDF error	Atom-preserving	1000	—	0.01256
Wasserstein-1	Atom-preserving	50	—	0.05269
Wasserstein-1	Atom-preserving	1000	—	0.01256
Atomic probability MSE	Atom-preserving	50	—	0.004514
Atomic probability MSE	Atom-preserving	1000	—	0.0002474
Uniform CDF error	Ordinary partition	Reported range	Various	0.60000
Wasserstein-1	Ordinary partition	50	2	0.27633
Wasserstein-1	Ordinary partition	50	8	0.10878
Wasserstein-1	Ordinary partition	1000	5	0.11014
Wasserstein-1	Ordinary partition	1000	20	0.03695

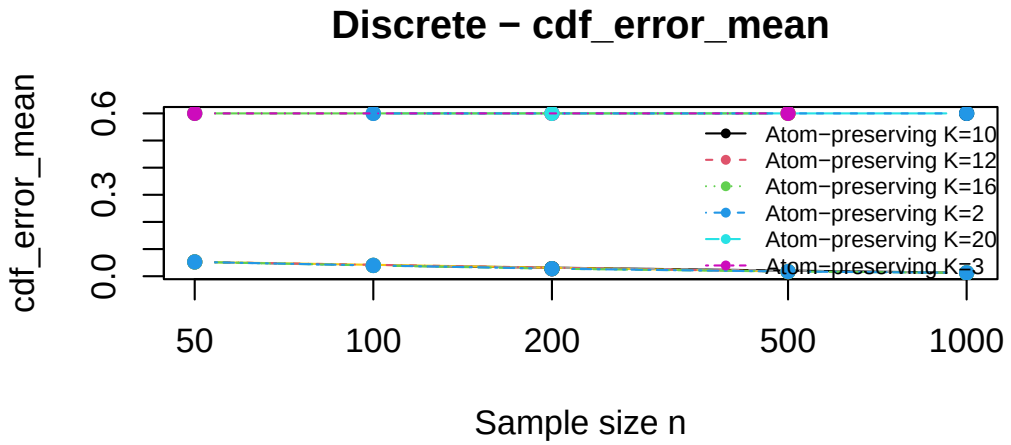


Figure 3: Mean uniform CDF error against sample size for the discrete distribution. The empirical CDF and the atom-preserving estimators retain the two support points and estimate their probabilities from the observed proportions. Their overlapping curves reflect the equivalence of these procedures in the implementation considered. Selected numerical results are reported in Table 3.

7.6.4 Mixed distribution

The mixed model consists of an atom with probability 0.3 at zero and an absolutely continuous uniform component on $(0, 1]$ with total probability 0.7. Equivalently, its distribution can be written as

$$F(x) = \begin{cases} 0, & x < 0, \\ 0.3 + 0.7x, & 0 \le x \le 1, \\ 1, & x > 1, \end{cases}$$

where the jump at zero has size 0.3, followed by a continuous increase over the interval $(0, 1]$. This model examines whether partition-based methods can accommodate a distribution containing both a singular atomic component and an absolutely continuous component. The numerical results for the empirical CDF, the model-aware parametric estimator, the ordinary partition approximation, and the estimated atomic probability are summarized in Table 4. Figure 4 displays the mean squared error (MSE) of the estimated probability mass at zero as a function of sample size.

The empirical CDF improved with increasing sample size. Its mean CDF error decreased from 0.11485 at $n = 50$ to 0.02650 at $n = 1000$, while its mean Wasserstein-1 distance decreased from 0.04678 to 0.01067, respectively (Table 4). The model-aware parametric estimator also exhibited a reduction in mean CDF error, from 0.05133 at $n = 50$ to 0.01203 at $n = 1000$. These results show the improvement in distributional estimation as the sample size grows. The smaller errors of the parametric estimator are consistent with the advantage of using the correct mixed-distribution structure. This comparison should be understood as a benchmark against a model-aware estimator rather than as evidence that a parametric approach will necessarily be preferable when the mixture structure or component distributions are unknown.

The ordinary partition-based approximation exhibited a substantial uniform CDF discrepancy. Its mean uniform CDF error was 0.47483 for $K = 2$ at $n = 50$, compared with 0.31736 for $K = 20$ at $n = 1000$. The corresponding mean Wasserstein-1 distances were 0.16882 and 0.02100, respectively (Table 4). These results illustrate that a partition-based approximation can have a large uniform CDF error even when its Wasserstein distance is relatively small. In a mixed distribution, the placement of the atomic mass is particularly important because relocating the mass at zero can shift the location of the CDF jump. Refining the continuous partition can improve the spatial approximation, but it does not by itself guarantee exact preservation of the atom. The numerical comparison across these configurations should nevertheless be interpreted with care, since both sample size and partition resolution differ between the two reported cases.

The estimation of the atomic probability provides a separate and more direct assessment of the mixed-distribution construction. The mean squared error of the estimated jump probability at zero decreased from 0.004195 at $n = 50$ to 0.0002257 at $n = 1000$ (Table 4). Figure 4 shows this reduction in estimation error as the sample size increases. When the atom is located at a known point, its probability can be estimated directly by the observed proportion of observations at that point. Under independent sampling, this empirical proportion is an unbiased estimator of the true atomic probability, and its variance decreases with sample size. The results therefore support direct estimation of the atomic mass rather than treating it as part of the continuous component.

The atom-preserving CDF implementation for this mixed model, however, requires correction before its distributional errors can be interpreted. In particular, the probabilities assigned to cells within the continuous component must be calculated using the observations belonging to that component and normalized by the number of such observations. These conditional cell probabilities can then be multiplied by the estimated total probability of the continuous component. The atomic probability must be included separately, with the resulting atomic and continuous contributions summing to one. Similarly, density estimation must distinguish

the point mass at zero from the absolutely continuous component: the density represents only the latter and must not treat the atom as an ordinary continuous observation. Until these corrections are implemented and the simulations are recomputed, the reported atom-preserving CDF errors and density ISE values are not used to draw conclusions about the distributional accuracy of the atom-preserving procedure.

Overall, the mixed-distribution experiment demonstrates the importance of treating atomic and continuous components separately. The empirical CDF and the model-aware parametric estimator improve with sample size, while the ordinary partition approximation illustrates the potential discrepancy caused by representing atomic mass at cell representatives. The declining MSE of the estimated jump probability further confirms that the atomic mass can be estimated increasingly accurately from observed frequencies. Figure 4 and Table 4 summarize these findings. A complete assessment of the atom-preserving CDF and density estimators, however, must await correction of the mixed-component normalization and subsequent recalculation of the corresponding simulation results.

Table 4: Selected simulation results for the mixed distribution with an atom of probability 0.3 at zero and a continuous uniform component of total probability 0.7. Atom-preserving CDF errors and density ISE are omitted because the current implementation requires correction.

Criterion	Estimator	n	K	Mean error
CDF error	Empirical CDF	50	—	0.11485
CDF error	Empirical CDF	1000	—	0.02650
Wasserstein-1	Empirical CDF	50	—	0.04678
Wasserstein-1	Empirical CDF	1000	—	0.01067
CDF error	Parametric	50	—	0.05133
CDF error	Parametric	1000	—	0.01203
CDF error	Ordinary partition	50	2	0.47483
CDF error	Ordinary partition	1000	20	0.31736
Wasserstein-1	Ordinary partition	50	2	0.16882
Wasserstein-1	Ordinary partition	1000	20	0.02100
Atomic probability MSE	Empirical atom frequency	50	—	0.004195
Atomic probability MSE	Empirical atom frequency	1000	—	0.0002257

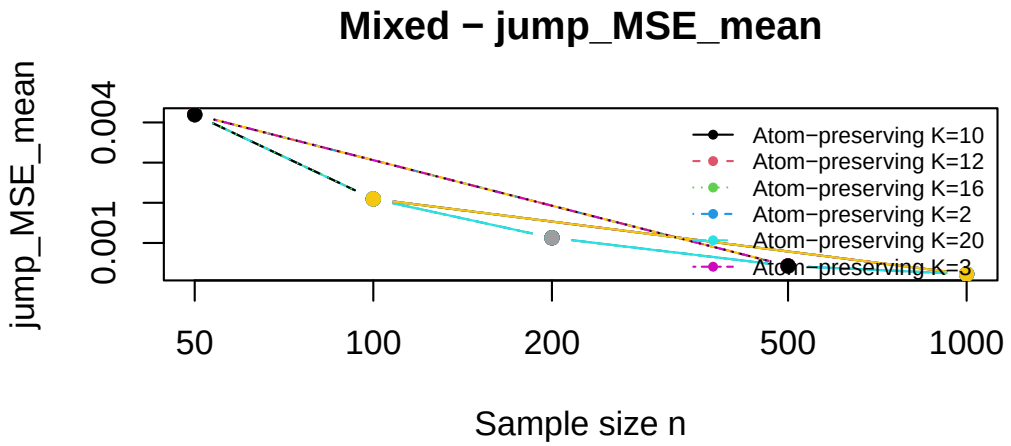


Figure 4: Mean squared error of the estimated atomic probability at zero against sample size for the mixed distribution. The decreasing MSE reflects the increasing precision of the empirical estimate of the point mass. The corresponding numerical results are reported in Table 4.

7.6.5 Multimodal continuous distribution

The multimodal model evaluates the ability of partition-based methods to approximate a continuous distribution characterized by separated regions of high probability density. The empirical CDF error decreased from 0.11887 at $n = 50$ to 0.02695 at $n = 1000$, while the corresponding Wasserstein-1 distance decreased from 0.28650 to 0.06532. The correctly specified mixture benchmark yielded mean uniform CDF errors of 0.07296 and 0.01651, respectively, at these sample sizes. Thus, the mixture benchmark outperformed the empirical CDF in terms of the reported CDF-error criterion. The numerical results for distribution-function approximation and density estimation are summarized in Table 5.

Table 5: Simulation results for the multimodal continuous distribution.

Method	Configuration	n	CDF error / ISE	W_1 distance
Empirical CDF	–	50	0.11887	0.28650
Empirical CDF	–	1000	0.02695	0.06532
Mixture benchmark	Correctly specified	50	0.07296	0.25224
Mixture benchmark	Correctly specified	1000	0.01651	0.05740
Partition-based CDF	$K = 2$	50	0.53800	0.73086
Partition-based CDF	$K = 8$	50	0.24912	–
Partition-based CDF	$K = 5$	1000	0.31170	0.45242
Partition-based CDF	$K = 20$	1000	0.10382	0.14679
Data-selected histogram	–	50	0.12526	–
Data-selected histogram	–	1000	0.03732	–
Piecewise-constant density	$K = 5$	1000	0.04469	–
Piecewise-constant density	$K = 20$	1000	0.00919	–

Notes. For distribution-function estimators, the fourth column reports the mean uniform CDF error. For density estimators, it reports the mean integrated squared error (ISE). The Wasserstein-1 distance is reported for the distribution-function estimators. A dash indicates that the corresponding result was not available in the simulation summary.

The distribution-function results reveal a clear effect of partition refinement. At $n = 50$, the mean uniform CDF error decreased from 0.53800 for $K = 2$ to 0.24912 for $K = 8$. At $n = 1000$, it decreased from 0.31170 for $K = 5$ to 0.10382 for $K = 20$. The Wasserstein-1 distance exhibited a similar pattern, decreasing from 0.45242 to 0.14679 as the partition size increased from $K = 5$ to $K = 20$ at $n = 1000$. These results suggest that finer partitions preserve more information about the locations and relative magnitudes of separated modes. By contrast, a coarse partition may aggregate observations from distinct high-density regions into the same cell, thereby losing important features of the underlying distribution. Despite the improvement under refinement, the partition-based estimator did not attain the accuracy of the correctly specified mixture benchmark under the reported CDF-error criterion.

The density-estimation results exhibit a complementary pattern. The data-selected histogram had mean ISE 0.12526 at $n = 50$, decreasing to 0.03732 at $n = 1000$. For the piecewise-constant density estimator at $n = 1000$, the mean ISE decreased from 0.04469 for $K = 5$ to 0.00919 for $K = 20$. Thus, for the reported configurations, a finer partition provided a more accurate density approximation. The decline in histogram ISE with sample size also indicates improved estimation as more observations became available. Figure 5 displays the mean ISE across sample sizes for the histogram and piecewise-constant density estimators.

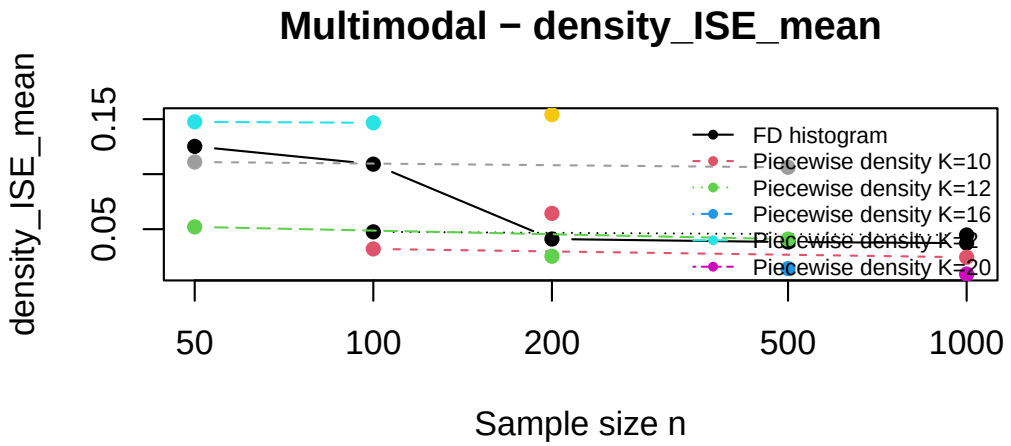


Figure 5: Mean integrated squared error against sample size for density estimators under the multimodal continuous model. The figure compares the data-selected histogram with piecewise-constant density estimators using different partition sizes.

Overall speaking, the multimodal experiment illustrates both the flexibility and the limitations of partition-based estimation. A sufficiently fine partition can represent separated regions of high density without imposing a single-component parametric form. However, increasing the number of cells does not guarantee uniformly better performance across all estimation criteria. In the reported results, the correctly specified mixture benchmark performed better for distribution-function estimation, whereas finer partitions improved the piecewise-constant density estimator. These findings emphasize the importance of distinguishing distribution-function approximation from density estimation when assessing partition-based methods.

7.6.6 Overall interpretation

The numerical experiments reveal several important features of partition-based probability approximation across uniform, discontinuous, discrete, mixed, and multimodal distributions. First, increasing the sample size consistently improves the empirical distribution-function estimator, as reflected in the decreasing uniform CDF error and Wasserstein-1 distance across the models considered. Second, refining the partition generally improves the distributional approximation by reducing the spatial error induced by representing observations within each cell by a finite collection of representative points. However, the magnitude of this improvement depends on the underlying distribution, the partition geometry, and the error criterion employed. In density estimation, finer partitions provide greater flexibility to capture local features, including discontinuities and separated modes, but may also increase sampling variability; consequently, the optimal partition size need not be the largest one considered. Third, the experiments demonstrate the importance of explicitly preserving atomic locations when the target distribution contains point masses. In the discrete model, the ordinary partition-based approximation exhibited a substantial uniform CDF error because the relocation of atomic probability mass altered the locations of the distribution-function jumps, whereas the atom-preserving estimator coincided with the empirical CDF when the support was known and estimated the atomic probabilities

consistently. A similar limitation was observed for the ordinary partition approximation in the mixed model, where continuous and atomic components must be treated separately; however, the atom-preserving CDF results for this model require recomputation following correction of the implementation. The distinction between uniform CDF error and Wasserstein-1 distance is also important for interpreting the numerical findings. Uniform CDF error measures the largest absolute discrepancy between two distribution functions and can therefore be particularly sensitive to errors near jump discontinuities, whereas the Wasserstein-1 distance reflects the cumulative spatial cost of transporting probability mass and is consequently sensitive to the distances over which that mass is displaced. Improvements under one criterion therefore need not correspond to proportional improvements under the other. Taken together, the experiments support partition refinement as a means of increasing distributional resolution and demonstrate that preserving exact atomic locations is essential when the target contains point masses whose jump locations must be retained. They also indicate that the performance of the proposed construction depends jointly on sample size, partition resolution, partition alignment, and the nature of the approximation target. Accordingly, conclusions concerning the mixed-model atom-preserving CDF estimator and its density-estimation results should remain provisional until the relevant implementation errors have been corrected and the simulations recomputed.

8 Conclusion

This study develops and investigates a partition-based framework for approximating probability distributions with different structural characteristics, including continuous distributions, discontinuous densities, discrete distributions with point masses, mixed distributions, and multimodal continuous distributions. The central idea is to represent a probability law through a finite collection of partition cells and associated probability masses, thereby providing a flexible construction for approximating distribution functions and, under suitable conditions, probability densities. The numerical experiments demonstrate that the effectiveness of this construction depends on the sample size, the number and geometry of the partition cells, the nature of the underlying distribution, and the approximation criterion.

The simulation results indicate that increasing the sample size improves distribution-function estimation, as reflected in the decreasing uniform CDF error and Wasserstein-1 distance for the empirical estimator across the models considered. Partition refinement also generally improves distributional approximation by reducing the spatial discrepancy arising from the representation of probability mass at selected locations. This effect is particularly evident in the multimodal experiment, where finer partitions provide a more detailed representation of separated regions of high probability density. For density estimation, the results further show that increasing the partition resolution can substantially reduce integrated squared error in the configurations examined. Nevertheless, the relationship between partition size and estimation accuracy is not universally monotone: finer partitions may increase sampling variability, and the balance between approximation error and estimation error must be considered when selecting the partition resolution.

An important finding concerns distributions containing point masses. The discrete-distribution

experiment illustrates that an ordinary partition approximation may fail to preserve the locations of distribution-function jumps when atomic probability is redistributed among representative points. By contrast, when the support is known, an atom-preserving construction can retain the specified atomic locations and estimate their probabilities using empirical proportions. This distinction motivates treating atomic and absolutely continuous components separately when approximating mixed distributions. However, the atom-preserving CDF results and the density-estimation results for the mixed model require recomputation after correction of the implementation. Consequently, no definitive finite-sample performance claim is made for those components until the corrected simulations have been completed.

The experiments also emphasize that distributional approximation should be evaluated using complementary criteria. The uniform CDF error measures the largest discrepancy between distribution functions and is particularly informative about errors in jump heights and locations, whereas the Wasserstein-1 distance quantifies distributional discrepancy in terms of the spatial displacement of probability mass. Integrated squared error, in turn, assesses the accuracy of density estimation and is applicable only when the target and estimated densities are defined with respect to a common dominating measure. These criteria capture different aspects of approximation quality and should not be interpreted interchangeably.

Several directions remain for further investigation. The theoretical analysis can be extended to establish explicit convergence rates under different partition-refinement schemes and to characterize the trade-off between partition resolution and sampling variability. Data-adaptive partition selection, higher-dimensional constructions, and principled procedures for choosing representative points also warrant study. For mixed distributions, a theoretically justified decomposition into atomic and continuous components would provide a systematic basis for preserving jump locations while estimating the continuous component. Finally, the numerical assessment should be extended using corrected mixed-model simulations, additional distributional settings, and uncertainty summaries across repeated Monte Carlo samples.

In conclusion, partition-based probability approximation offers a flexible approach to representing distributions with diverse structural features without requiring every target law to belong to a prespecified parametric family. Its practical reliability, however, depends on an appropriate choice of partition, a clear distinction between atomic and continuous components, and evaluation using criteria suited to the target of interest. The findings support further development of partition-based methods as a general framework for distributional approximation, while identifying implementation validation, theoretical convergence analysis, and adaptive partition selection as essential steps toward a more comprehensive methodology.

Data availability

No external dataset is required for the theoretical results. The simulation data can be generated from the specified probability models. Simulation code and generated results are available in .

Acknowledgements

The author is grateful to his colleagues for providing such novel thoughts.

Conflict of interest

No conflict of interest.

Funding

No external funding to conduct this research.

References

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York.

Dudley, R. M. (2002). *Real Analysis and Probability*, 2nd ed. Cambridge University Press, Cambridge.

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Villani, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer, New York.

A Appendix: Proofs

A.1 Proof of Proposition 2.1

Let $\mathcal{P}_K = \{A_1, \ldots, A_K\}$ be a finite measurable partition of the sample space $(\mathcal{X}, \mathcal{A})$. Thus, each $A_j \in \mathcal{A}$, the sets $A_1, \ldots, A_K$ are pairwise disjoint, and $\bigcup_{j=1}^{K} A_j = \mathcal{X}$. Suppose that $p_j \geq 0$, $j = 1, \ldots, K$, with $\sum_{j=1}^{K} p_j = 1$, and let Q_j be a probability measure on $(\mathcal{X}, \mathcal{A})$ satisfying $Q_j(A_j) = 1$. In other words, Q_j assigns probability one to the j-th partition cell. For every $B \in \mathcal{A}$, define

$$P_{\mathcal{P}}(B) = \sum_{j=1}^{K} p_j Q_j(B). \tag{A.1}$$

We prove that $P_{\mathcal{P}}$ satisfies the defining properties of a probability measure.

Since each Q_j is a probability measure, we have $Q_j(B) \geq 0$ for every $B \in \mathcal{A}$. Together with $p_j \geq 0$, this gives

$$P_{\mathcal{P}}(B) = \sum_{j=1}^{K} p_j Q_j(B) \geq 0.$$

Moreover, because $Q_j(\mathcal{X}) = 1$ for every $j = 1, \ldots, K$, $P_{\mathcal{P}}(\mathcal{X}) = \sum_{j=1}^{K} p_j Q_j(\mathcal{X}) = \sum_{j=1}^{K} p_j = 1$.

In particular, $P_{\mathcal{P}}(\varnothing) = 0$, since $Q_j(\varnothing) = 0$ for every j. Thus, $P_{\mathcal{P}}$ is nonnegative and normalized.

Let $B_1, B_2, \ldots \in \mathcal{A}$ be any sequence of pairwise disjoint measurable sets. Their union $B = \bigcup_{m=1}^{\infty} B_m$ belongs to $\mathcal{A}$, because $\mathcal{A}$ is a sigma-algebra. Since each Q_j is a probability measure, its countable additivity implies

$$Q_j\left(\bigcup_{m=1}^{\infty} B_m\right) = \sum_{m=1}^{\infty} Q_j(B_m), \qquad j = 1, \ldots, K.$$

Consequently, using the definition in (2.1), we obtain

$$P_{\mathcal{P}}\left(\bigcup_{m=1}^{\infty} B_m\right) = \sum_{j=1}^{K} p_j Q_j\left(\bigcup_{m=1}^{\infty} B_m\right) = \sum_{j=1}^{K} p_j \sum_{m=1}^{\infty} Q_j(B_m) = \sum_{m=1}^{\infty} \sum_{j=1}^{K} p_j Q_j(B_m) = \sum_{m=1}^{\infty} P_{\mathcal{P}}(B_m).$$

The interchange of the finite sum over j and the infinite sum over m is valid because all terms $p_j Q_j(B_m)$ are nonnegative. Hence, $P_{\mathcal{P}}$ is countably additive on $\mathcal{A}$.

The set function $P_{\mathcal{P}}$ is nonnegative, satisfies $P_{\mathcal{P}}(\mathcal{X}) = 1$, and is countably additive on the measurable space $(\mathcal{X}, \mathcal{A})$. It therefore satisfies all the axioms of a probability measure. This establishes the validity of the construction.

A.2 Proof of Theorem 3.1

Let $\mathcal{X} = [a, b]$ be a compact interval, and let P be a probability measure supported on $\mathcal{X}$. For each K, consider a finite measurable partition

$$\mathcal{P}_K = \{A_{K1}, A_{K2}, \ldots, A_{KK}\}$$

of $\mathcal{X}$, and let $p_{Kj} = P(A_{Kj})$; $j = 1, \ldots, K$. Choose a representative point $a_{Kj} \in A_{Kj}$ for each cell and define the corresponding discrete probability measure by $P_K = \sum_{j=1}^{K} p_{Kj}\delta_{a_{Kj}}$, where $\delta_{a_{Kj}}$ denotes the Dirac probability measure concentrated at a_{Kj}. The mesh of the partition is

$$\delta_K = \max_{1 \le j \le K} \operatorname{diam}(A_{Kj}),$$

and, by assumption, $\delta_K \to 0$ as $K \to \infty$.

We first establish the stated bound for an arbitrary continuous function $g : \mathcal{X} \to \mathbb{R}$. Since $\mathcal{X}$ is compact and g is continuous, the Heine–Cantor theorem guarantees that g is uniformly continuous on $\mathcal{X}$. Consequently, its modulus of continuity $\omega_g(t) = \sup_{\substack{x,y \in \mathcal{X} \\ |x-y| \le t}} |g(x) - g(y)|$ satisfies $\lim_{t \downarrow 0} \omega_g(t) = 0$. By the definition of P_K, $\displaystyle\int_{\mathcal{X}} g(x)\, P_K(dx) = \sum_{j=1}^{K} p_{Kj} g(a_{Kj})$.

On the other hand, since the sets $A_{K1}, \ldots, A_{KK}$ form a measurable partition of $\mathcal{X}$, finite additivity of the integral gives $\displaystyle\int_{\mathcal{X}} g(x)\, P(dx) = \sum_{j=1}^{K} \int_{A_{Kj}} g(x)\, P(dx)$.

Subtracting these expressions yields

$$\begin{aligned}
&\int_{\mathcal{X}} g\, dP_K - \int_{\mathcal{X}} g\, dP \\
&\quad = \sum_{j=1}^{K} \left\{ p_{Kj} g(a_{Kj}) - \int_{A_{Kj}} g(x)\, P(dx) \right\} \\
&\quad = \sum_{j=1}^{K} \int_{A_{Kj}} \{ g(a_{Kj}) - g(x) \}\, P(dx),
\end{aligned}$$

where the last equality follows from $p_{Kj} = P(A_{Kj})$.

Applying the triangle inequality, we obtain

$$\left| \int_{\mathcal{X}} g\, dP_K - \int_{\mathcal{X}} g\, dP \right| \le \sum_{j=1}^{K} \int_{A_{Kj}} |g(a_{Kj}) - g(x)|\, P(dx).$$

For every $x \in A_{Kj}$, the representative a_{Kj} also belongs to A_{Kj}. Therefore,

$$|x - a_{Kj}| \le \operatorname{diam}(A_{Kj}) \le \delta_K.$$

By the definition of the modulus of continuity, $|g(a_{Kj}) - g(x)| \le \omega_g(\delta_K)$, $x \in A_{Kj}$. It follows

that

$$\left|\int_{\mathcal{X}} g\,dP_K - \int_{\mathcal{X}} g\,dP\right| \leq \omega_g(\delta_K) \sum_{j=1}^{K} P(A_{Kj}) \;= \omega_g(\delta_K),$$

because the partition covers $\mathcal{X}$ and $P(\mathcal{X}) = 1$. This proves $\left|\int g\,dP_K - \int g\,dP\right| \leq \omega_g(\delta_K)$, as asserted.

It remains to establish weak convergence. By definition, P_K converges weakly to P, denoted by $P_K \Rightarrow P$, if $\lim_{K\to\infty} \int_{\mathcal{X}} g\,dP_K = \int_{\mathcal{X}} g\,dP$ for every bounded continuous function g on $\mathcal{X}$. Since $\mathcal{X}$ is compact, every continuous real-valued function on $\mathcal{X}$ is bounded. The inequality already established, together with $\delta_K \to 0$, gives

$$\left|\int_{\mathcal{X}} g\,dP_K - \int_{\mathcal{X}} g\,dP\right| \leq \omega_g(\delta_K) \;\longrightarrow 0\,;\; K \to \infty,$$

by uniform continuity of g. Hence the defining criterion for weak convergence is satisfied, and

$$P_K \Rightarrow P.$$

A.3 Proof of Proposition 3.2

Proof. Let X be a random variable with distribution P and cumulative distribution function

$$F(x) = P(X \leq x), \qquad x \in \mathbb{R}.$$

Consider a measurable partition $\{A_1, \ldots, A_K\}$ of the sample space of X, and let T_K be the associated quantization map defined by

$$T_K(y) = a_j \qquad \text{if } y \in A_j,$$

where $a_j \in \mathbb{R}$ is the representative point associated with the cell A_j. The distribution function of the quantized random variable $T_K(X)$ is

$$F_K(x) = P(T_K(X) \leq x), \qquad x \in \mathbb{R}.$$

By expressing both distribution functions as expectations of indicator functions, we obtain

$$F_K(x) = \mathbb{E}[\mathbf{1}\{T_K(X) \leq x\}]$$

and

$$F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]\,.$$

Subtracting the second identity from the first gives

$$F_K(x) - F(x) = \mathbb{E}[\mathbf{1}\{T_K(X) \leq x\} - \mathbf{1}\{X \leq x\}]\,.$$

Since X has distribution P, the expectation of any integrable measurable function $g(X)$ can be

written as

$$\mathbb{E}[g(X)] = \int g(y)\, dP(y).$$

Applying this identity to the difference of indicator functions yields

$$F_K(x) - F(x) = \int \left[\mathbf{1}\{T_K(y) \le x\} - \mathbf{1}\{y \le x\}\right] dP(y),$$

which establishes (3.2).

To establish the inequality, define

$$I_x(y) = \mathbf{1}\{T_K(y) \le x\}, \qquad J_x(y) = \mathbf{1}\{y \le x\}.$$

Both $I_x(y)$ and $J_x(y)$ take values in $\{0, 1\}$. Consequently, their difference satisfies

$$|I_x(y) - J_x(y)| = \mathbf{1}\{I_x(y) \ne J_x(y)\}.$$

Indeed, the absolute difference equals zero when the two indicators agree and equals one when they disagree. Taking absolute values in (3.2) and applying the triangle inequality for integrals gives

$$\begin{aligned} |F_K(x) - F(x)| &= \left| \int \left[I_x(y) - J_x(y)\right] dP(y) \right| \\ &\le \int |I_x(y) - J_x(y)|\, dP(y) \\ &= \int \mathbf{1}\{I_x(y) \ne J_x(y)\}\, dP(y) \\ &= P\big(I_x(X) \ne J_x(X)\big). \end{aligned}$$

Substituting the definitions of I_x and J_x, we obtain

$$|F_K(x) - F(x)| \le P\left(\mathbf{1}\{T_K(X) \le x\} \ne \mathbf{1}\{X \le x\}\right),$$

which is precisely (3.3). Since x was arbitrary, both identities and the inequality hold for every $x \in \mathbb{R}$. □

A.4 Proof of Corollary 3.3

Proof. Let P_K denote the probability distribution induced by the quantization map T_K, so that

$$P_K(B) = P\big(T_K(X) \in B\big)$$

for every Borel set $B \subseteq \mathbb{R}$. Its cumulative distribution function is

$$F_K(x) = P_K((-\infty, x]).$$

By Theorem 3.1, the sequence of quantized probability measures converges weakly to P, that is,

$$P_K \Rightarrow P \qquad \text{as } K \to \infty.$$

The Portmanteau theorem implies that, for every Borel set A whose boundary has zero probability under P,

$$P_K(A) \longrightarrow P(A).$$

Fix a continuity point x of F and take $A = (-\infty, x]$. The boundary of this set is $\partial A = \{x\}$. Since F is continuous at x,

$$P(\partial A) = P(\{x\}) = F(x) - F(x-) = 0.$$

Therefore, the Portmanteau theorem gives

$$F_K(x) = P_K((-\infty, x]) \longrightarrow P((-\infty, x]) = F(x).$$

Since x was an arbitrary continuity point of F, the asserted convergence holds at every continuity point of the limiting distribution function. □

A.5 Proof of Lemma 4.1

Proof. Let $\mathcal{P}_K = \{A_1, \ldots, A_K\}$ be a fixed measurable partition, and define

$$p_j = P(A_j), \qquad \widehat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \in A_j\}, \qquad j = 1, \ldots, K,$$

where $X_1, \ldots, X_n$ are independent and identically distributed observations from P. For each j, $\widehat{p}_j$ is the empirical proportion of observations falling in A_j. Hence,

$$\mathbb{E}[\widehat{p}_j] = p_j, \qquad \mathrm{Var}(\widehat{p}_j) = \frac{p_j(1-p_j)}{n}.$$

By the Cauchy–Schwarz inequality, $\mathbb{E}|\widehat{p}_j - p_j| \le \sqrt{\mathbb{E}[(\widehat{p}_j - p_j)^2]} = \sqrt{\dfrac{p_j(1-p_j)}{n}}$.

Summing over the K cells and using linearity of expectation, we obtain

$$\mathbb{E}\left[\sum_{j=1}^{K} |\widehat{p}_j - p_j|\right] = \sum_{j=1}^{K} \mathbb{E}|\widehat{p}_j - p_j| \le \frac{1}{\sqrt{n}} \sum_{j=1}^{K} \sqrt{p_j(1-p_j)} \le \frac{1}{\sqrt{n}} \sum_{j=1}^{K} \sqrt{p_j}.$$

Applying the Cauchy–Schwarz inequality to the final sum gives $\displaystyle\sum_{j=1}^{K} \sqrt{p_j} \le \sqrt{K \sum_{j=1}^{K} p_j}$.

Because the cells form a partition, $\displaystyle\sum_{j=1}^{K} p_j = 1$. Consequently, $\displaystyle\mathbb{E}\left[\sum_{j=1}^{K} |\widehat{p}_j - p_j|\right] \le \sqrt{\frac{K}{n}}$, which establishes (4.3). □

A.6 Proof of Theorem 4.2

Proof. Let $\mathcal{P}_{K_n} = \{A_{1,n}, \ldots, A_{K_n,n}\}$ be the deterministic partition at sample size n, with representative points $a_{1,n}, \ldots, a_{K_n,n}$. Write

$$p_{j,n} = P(A_{j,n}), \qquad \widehat{p}_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \in A_{j,n}\}.$$

The population partition measure and its empirical counterpart are

$$P_{K_n} = \sum_{j=1}^{K_n} p_{j,n}\delta_{a_{j,n}}, \qquad \widehat{P}_{n,K_n} = \sum_{j=1}^{K_n} \widehat{p}_{j,n}\delta_{a_{j,n}},$$

where δ_a denotes the unit point mass at a. We prove that $\widehat{P}_{n,K_n} \Rightarrow P$ in probability by separating the approximation error due to quantization from the sampling error due to estimation of the cell probabilities.

Let S be a compact interval supporting P, and let $f : S \to \mathbb{R}$ be any continuous function. Since S is compact, f is bounded and uniformly continuous. Define its modulus of continuity by

$$\omega_f(t) = \sup\{|f(u) - f(v)| : u, v \in S,\ |u - v| \le t\}.$$

Uniform continuity implies that $\omega_f(t) \to 0$ as $t \downarrow 0$.

For each cell $A_{j,n}$, let $a_{j,n}$ be its representative point, and suppose the partition mesh δ_{K_n} bounds the distance from every point in a cell to its representative:

$$\sup_{y \in A_{j,n}} |y - a_{j,n}| \le \delta_{K_n}, \qquad j = 1, \ldots, K_n,$$

up to irrelevant choices on cell boundaries. The difference between the integral under P_{K_n} and that under P satisfies

$$\begin{aligned}
\left| \int f\, dP_{K_n} - \int f\, dP \right| &= \left| \sum_{j=1}^{K_n} \int_{A_{j,n}} \big[f(a_{j,n}) - f(y)\big]\, dP(y) \right| \\
&\le \sum_{j=1}^{K_n} \int_{A_{j,n}} |f(a_{j,n}) - f(y)|\, dP(y) \\
&\le \omega_f(\delta_{K_n}) \sum_{j=1}^{K_n} P(A_{j,n}) \\
&= \omega_f(\delta_{K_n}).
\end{aligned}$$

Therefore, $\left| \int f\, dP_{K_n} - \int f\, dP \right| \longrightarrow 0$, because $\delta_{K_n} \to 0$.

Next, consider the sampling error:

$$\left| \int f\, d\widehat{P}_{n,K_n} - \int f\, dP_{K_n} \right| = \left| \sum_{j=1}^{K_n} f(a_{j,n})(\widehat{p}_{j,n} - p_{j,n}) \right| \le \|f\|_\infty \sum_{j=1}^{K_n} |\widehat{p}_{j,n} - p_{j,n}|,$$

where $\|f\|_\infty = \sup_{y\in S} |f(y)|$. Taking expectations and applying Lemma 4.1 to the partition $\mathcal{P}_{K_n}$ yields

$$\mathbb{E}\left[\left|\int f\, d\widehat{P}_{n,K_n} - \int f\, dP_{K_n}\right|\right] \le \|f\|_\infty \sqrt{\frac{K_n}{n}}.$$

By Markov's inequality, for every $\varepsilon > 0$,

$$P\left(\left|\int f\, d\widehat{P}_{n,K_n} - \int f\, dP_{K_n}\right| > \varepsilon\right) \le \frac{\|f\|_\infty}{\varepsilon}\sqrt{\frac{K_n}{n}} \longrightarrow 0,$$

since $K_n/n \to 0$. Thus, $\displaystyle\int f\, d\widehat{P}_{n,K_n} - \int f\, dP_{K_n} \xrightarrow{P} 0$.

Combining the approximation and sampling errors by the triangle inequality gives

$$\left|\int f\, d\widehat{P}_{n,K_n} - \int f\, dP\right| \le \left|\int f\, d\widehat{P}_{n,K_n} - \int f\, dP_{K_n}\right| + \left|\int f\, dP_{K_n} - \int f\, dP\right|.$$

The first term converges to zero in probability, and the second converges deterministically to zero. Hence, for every bounded continuous f on S, $\displaystyle\int f\, d\widehat{P}_{n,K_n} \xrightarrow{P} \int f\, dP$.

This is the bounded-continuous-function characterization of weak convergence in probability on the compact interval S. Therefore, $\widehat{P}_{n,K_n} \Rightarrow P$ in probability. □

A.7 Proof of Theorem 4.3

Proof. Let $\mathcal{P}_K = \{A_1, \ldots, A_K\}$ be the partition, with representative points $a_1, \ldots, a_K$, and write

$$p_j = P(A_j), \qquad \widehat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \in A_j\}.$$

The population and empirical partition measures are

$$P_K = \sum_{j=1}^{K} p_j \delta_{a_j}, \qquad \widehat{P}_{n,K} = \sum_{j=1}^{K} \widehat{p}_j \delta_{a_j}.$$

We first bound the Wasserstein distance between P and P_K, and then control the distance between P_K and $\widehat{P}_{n,K}$.

By the definition of the 1-Wasserstein distance,

$$W_1(P, P_K) = \inf_{\pi\in\Pi(P,P_K)} \int |x - y|\, d\pi(x, y),$$

where $\Pi(P, P_K)$ denotes the collection of all couplings of P and P_K. Consider the coupling induced by the map T_K, which sends each $x \in A_j$ to its representative a_j. Under this coupling, the transport cost is bounded by the maximum cell diameter:

$$W_1(P, P_K) \le \int |x - T_K(x)|\, dP(x) = \sum_{j=1}^{K} \int_{A_j} |x - a_j|\, dP(x) \le \delta_K \sum_{j=1}^{K} P(A_j) = \delta_K.$$

Here, the inequality uses the assumption that each representative a_j belongs to its corresponding

cell A_j, so that $|x - a_j| \le \delta_K$ for $x \in A_j$.

We next bound $W_1(P_K, \widehat{P}_{n,K})$. Since both measures are supported on the representative points $a_1, \dots, a_K$, their discrepancy arises solely from the differences between the corresponding cell probabilities. Let

$$d_j = \widehat{p}_j - p_j, \qquad D = \sum_{j=1}^{K} |d_j|.$$

Because both $(p_1, \dots, p_K)$ and $(\widehat{p}_1, \dots, \widehat{p}_K)$ are probability vectors, $\sum_{j=1}^{K} d_j = 0$. The total excess mass that must be transported from cells with surplus probability to cells with deficient probability is therefore

$$m = \sum_{j:d_j>0} d_j = \sum_{j:d_j<0} (-d_j) = \frac{1}{2} \sum_{j=1}^{K} |d_j|.$$

The support of P lies in an interval of length L; hence, the distance between any two representative points is at most L. The unmatched mass can consequently be redistributed at a cost of at most L per unit of mass. This gives

$$W_1(P_K, \widehat{P}_{n,K}) \le Lm = \frac{L}{2} \sum_{j=1}^{K} |\widehat{p}_j - p_j| \le L \sum_{j=1}^{K} |\widehat{p}_j - p_j|.$$

Applying the triangle inequality for the Wasserstein distance, we obtain

$$W_1(P, \widehat{P}_{n,K}) \le W_1(P, P_K) + W_1(P_K, \widehat{P}_{n,K}) \le \delta_K + L \sum_{j=1}^{K} |\widehat{p}_j - p_j|.$$

This establishes (4.4). To derive the stochastic rate, take expectations in the established bound. By Lemma 4.1,

$$\mathbb{E}\big[W_1(P, \widehat{P}_{n,K})\big] \le \delta_K + L\mathbb{E}\left[\sum_{j=1}^{K} |\widehat{p}_j - p_j|\right] \le \delta_K + L\sqrt{\frac{K}{n}}.$$

For any $\varepsilon > 0$, Markov's inequality therefore implies

$$P\left\{W_1(P, \widehat{P}_{n,K}) > \frac{\delta_K + L\sqrt{K/n}}{\varepsilon}\right\} \le \varepsilon.$$

Equivalently, the Wasserstein error is bounded in probability at the scale $\delta_K + L\sqrt{K/n}$. Since L is a fixed constant, this yields $W_1(P, \widehat{P}_{n,K}) = O_{\mathbb{P}}\left(\delta_K + \sqrt{\frac{K}{n}}\right)$, which proves (4.5). □

A.8 Proof of Theorem 5.1

Proof. By assumption, the target probability measure admits the decomposition

$$P = P_{\mathrm{at}} + P_{\mathrm{na}},$$

where P_{at} denotes the atomic component and P_{na} denotes the non-atomic component. Let F_{at} and F_{na} be their corresponding cumulative distribution functions, defined by

$$F_{\mathrm{at}}(x) = P_{\mathrm{at}}((-\infty, x]), \qquad F_{\mathrm{na}}(x) = P_{\mathrm{na}}((-\infty, x]).$$

The decomposition implies

$$F(x) = F_{\mathrm{at}}(x) + F_{\mathrm{na}}(x), \qquad x \in \mathbb{R}.$$

By construction, the approximation $P_K^\star$ preserves the atomic component exactly and approximates only the non-atomic component. Consequently, its distribution function can be written as

$$F_K^\star(x) = F_{\mathrm{at}}(x) + F_{\mathrm{na},K}(x),$$

where $F_{\mathrm{na},K}$ denotes the distribution-function contribution of the approximated non-atomic component. Subtracting $F(x)$ from $F_K^\star(x)$, we obtain

$$F_K^\star(x) - F(x) = F_{\mathrm{na},K}(x) - F_{\mathrm{na}}(x).$$

Thus, the atomic contribution cancels identically, and the entire CDF approximation error is determined by the approximation of the non-atomic component.

By the definition of q_K in the atom-preserving construction (5.1), the non-atomic approximation satisfies

$$\sup_{x\in\mathbb{R}} |F_{\mathrm{na},K}(x) - F_{\mathrm{na}}(x)| \le q_K.$$

Therefore,

$$\sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| = \sup_{x\in\mathbb{R}} |F_{\mathrm{na},K}(x) - F_{\mathrm{na}}(x)| \le q_K,$$

which establishes (5.2).

Finally, if $q_K \to 0$ as $K \to \infty$, then $0 \le \sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| \le q_K \longrightarrow 0$.

The squeeze theorem consequently yields $\sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| \longrightarrow 0$. The argument does not require continuity of F. In particular, the atomic component is preserved exactly, so its jumps introduce no additional approximation error. This completes the proof. □

A.9 Proof of Theorem 5.2

Proof. Let P denote the true distribution and let $\widehat{P}_{n,K}^\star$ denote the atom-preserving partition-based estimator. Write

$$F(x) = P((-\infty, x]), \qquad \widehat{F}_{n,K}^\star(x) = \widehat{P}_{n,K}^\star((-\infty, x]).$$

Let $P_K^\star$ be the corresponding population-level partition approximation, with CDF $F_K^\star$. By the preceding construction, the deterministic approximation satisfies

$$\sup_{x\in\mathbb{R}} |F_K^\star(x) - F(x)| \le q_K.$$

Moreover, the empirical estimator replaces the population cell probabilities p_j by their empirical counterparts $\widehat{p}_j$, while retaining the prescribed atomic component. For any $x \in \mathbb{R}$, the contribution of each estimated cell probability to the CDF is either zero, the full cell probability, or a fraction of that probability, depending on the precise within-cell representation. In particular, the absolute difference between the empirical and population partition CDFs is bounded by the sum of the absolute cell-probability errors. Consequently,

$$\sup_{x\in\mathbb{R}} \left| \widehat{F}^{\star}_{n,K}(x) - F^{\star}_{K}(x) \right| \leq \sum_{j=1}^{K} |\widehat{p}_j - p_j|.$$

Applying the triangle inequality gives

$$\sup_{x\in\mathbb{R}} |\widehat{F}^{\star}_{n,K}(x) - F(x)| \leq \sup_{x\in\mathbb{R}} |\widehat{F}^{\star}_{n,K}(x) - F^{\star}_{K}(x)| + \sup_{x\in\mathbb{R}} |F^{\star}_{K}(x) - F(x)| \leq q_K + \sum_{j=1}^{K} |\widehat{p}_j - p_j|,$$

which establishes (5.3).

We next establish the asserted convergence in probability. For each cell A_j, let

$$p_j = P(A_j), \qquad \widehat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \in A_j\}.$$

Since $\widehat{p}_j$ is an empirical cell proportion,

$$\mathbb{E}(\widehat{p}_j) = p_j, \qquad \mathrm{Var}(\widehat{p}_j) = \frac{p_j(1-p_j)}{n}.$$

By the Cauchy–Schwarz inequality,

$$\mathbb{E}|\widehat{p}_j - p_j| \leq \sqrt{\frac{p_j(1-p_j)}{n}} \leq \sqrt{\frac{p_j}{n}}.$$

Summing over all $K = K_n$ cells, including cells assigned to separately represented atoms, yields

$$\mathbb{E}\left[\sum_{j=1}^{K_n} |\widehat{p}_j - p_j|\right] \leq \frac{1}{\sqrt{n}} \sum_{j=1}^{K_n} \sqrt{p_j} \leq \sqrt{\frac{K_n}{n}},$$

where the last inequality follows from $\sum_{j=1}^{K_n} p_j = 1$ and the Cauchy–Schwarz inequality. Therefore, by Markov's inequality, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sum_{j=1}^{K_n} |\widehat{p}_j - p_j| > \varepsilon\right) \leq \frac{1}{\varepsilon}\sqrt{\frac{K_n}{n}} \longrightarrow 0,$$

provided $K_n/n \to 0$. Thus, $\displaystyle\sum_{j=1}^{K_n} |\widehat{p}_j - p_j| \xrightarrow{\mathbb{P}} 0$.

Combining this result with the uniform bound (5.3), we obtain

$$\sup_{x\in\mathbb{R}} |\widehat{F}^{\star}_{n,K_n}(x) - F(x)| \le q_{K_n} + \sum_{j=1}^{K_n} |\widehat{p}_j - p_j| \xrightarrow{\mathbb{P}} 0,$$

because $q_{K_n} \to 0$ and $K_n/n \to 0$. This proves the theorem. □

A.10 Proof of Lemma 5.3

Proof. Let $X_1, \dots, X_n$ be an independent and identically distributed sample from the distribution P, and let $\widehat{F}_n(x) = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}$ denote the empirical distribution function. By definition,

$$\widehat{F}_n(x_0-) = \lim_{x\uparrow x_0} \widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i < x_0\}.$$

Consequently, the empirical jump at x_0 satisfies

$$\begin{aligned}
\widehat{\Delta}_n(x_0) &= \widehat{F}_n(x_0) - \widehat{F}_n(x_0-) \\
&= \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}\{X_i \le x_0\} - \mathbf{1}\{X_i < x_0\}] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = x_0\}.
\end{aligned}$$

Thus, the empirical jump is precisely the sample proportion of observations equal to x_0.

Define $I_i = \mathbf{1}\{X_i = x_0\}$. Since the observations are independent and identically distributed, the variables $I_1, \dots, I_n$ are independent and identically distributed Bernoulli random variables with

$$\mathbb{P}(I_i = 1) = P(X_i = x_0) = p_0.$$

It follows that

$$\widehat{\Delta}_n(x_0) = \frac{1}{n}\sum_{i=1}^{n} I_i, \qquad \mathbb{E}[I_i] = p_0, \qquad \mathrm{Var}(I_i) = p_0(1-p_0).$$

By the strong law of large numbers, $\widehat{\Delta}_n(x_0) = \dfrac{1}{n}\sum_{i=1}^{n} I_i \xrightarrow{\text{a.s.}} \mathbb{E}[I_1] = p_0$. This establishes the strong consistency of the empirical jump estimator.

Finally, suppose $0 < p_0 < 1$. The classical central limit theorem applied to the Bernoulli variables gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_i - p_0) \xrightarrow{d} N(0, p_0(1-p_0)).$$

Since $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I_i - p_0) = \sqrt{n}\{\widehat{\Delta}_n(x_0) - p_0\}$, we obtain

$$\sqrt{n}\{\widehat{\Delta}_n(x_0) - p_0\} \xrightarrow{d} N(0, p_0(1-p_0))\,.$$

Hence, both the strong consistency and the asserted asymptotic normality follow. □

A.11 Proof of Proposition 5.4

Proof. Since P_K and Q_K are supported on the same ordered representative points $a_1 < \cdots < a_K$, their distribution functions can be expressed as

$$F_{P_K}(x) = \sum_{j=1}^{K} p_j \mathbf{1}\{a_j \le x\}, \qquad F_{Q_K}(x) = \sum_{j=1}^{K} q_j \mathbf{1}\{a_j \le x\}.$$

Consequently, for every $x \in \mathbb{R}$,

$$|F_{P_K}(x) - F_{Q_K}(x)| = \left|\sum_{j=1}^{K}(p_j - q_j)\mathbf{1}\{a_j \le x\}\right| \le \sum_{j=1}^{K}|p_j - q_j|\mathbf{1}\{a_j \le x\} \le \sum_{j=1}^{K}|p_j - q_j|,$$

where the first inequality follows from the triangle inequality and the fact that each indicator takes values in $\{0,1\}$. Taking the supremum over $x \in \mathbb{R}$ on the left-hand side yields

$$\sup_{x\in\mathbb{R}}|F_{P_K}(x) - F_{Q_K}(x)| \le \sum_{j=1}^{K}|p_j - q_j|.$$

This completes the proof. □

A.12 Proof of Theorem 6.1

Proof. Let $A_1, \ldots, A_K$ be the cells of the fixed partition, where K is fixed, and suppose that $X_1, \ldots, X_n$ are independent and identically distributed observations from the distribution P. For each observation X_i, define the K-dimensional indicator vector

$$\boldsymbol{Z}_i = \big(\mathbf{1}\{X_i \in A_1\}, \ldots, \mathbf{1}\{X_i \in A_K\}\big)^\top.$$

Because the cells form a partition, exactly one component of $\boldsymbol{Z}_i$ equals one, while all the remaining components equal zero. Moreover,

$$\widehat{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i, \qquad \mathbb{E}[\boldsymbol{Z}_i] = \boldsymbol{p},$$

where $p_j = P(X_i \in A_j)$ and $\sum_{j=1}^{K} p_j = 1$.

We first calculate the covariance matrix of $\boldsymbol{Z}_i$. Since the components are indicator variables,

$$\mathbb{E}[\boldsymbol{Z}_i\boldsymbol{Z}_i^\top] = \mathrm{diag}(\boldsymbol{p}).$$

Indeed, for $j \neq \ell$, the disjointness of the partition cells implies

$$\mathbb{E}[\mathbf{1}\{X_i \in A_j\}\mathbf{1}\{X_i \in A_\ell\}] = 0,$$

whereas the diagonal entries are equal to p_j. Therefore,

$$\mathrm{Cov}(\boldsymbol{Z}_i) = \mathbb{E}[\boldsymbol{Z}_i \boldsymbol{Z}_i^\top] - \mathbb{E}[\boldsymbol{Z}_i]\mathbb{E}[\boldsymbol{Z}_i]^\top = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top = \Sigma.$$

The vectors $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent and identically distributed. Furthermore, they are bounded and hence possess finite second moments. The multivariate central limit theorem consequently gives

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{Z}_i - \boldsymbol{p}) \xrightarrow{d} N_K(\mathbf{0}, \Sigma).$$

Using the identity $\dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} (\boldsymbol{Z}_i - \boldsymbol{p}) = \sqrt{n}\,(\widehat{\boldsymbol{p}} - \boldsymbol{p})$, we conclude that

$$\sqrt{n}\,(\widehat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{d} N_K(\mathbf{0}, \Sigma).$$

□

A.13 Proof of Proposition 6.2

Proof. Let $b_0 = -\infty < b_1 < \cdots < b_{K-1} < b_K = \infty$ denote the fixed partition boundaries, and let $\widehat{b}_0 = -\infty$, $\widehat{b}_K = \infty$ with $\widehat{b}_1, \ldots, \widehat{b}_{K-1}$ denoting the estimated interior boundaries. Define the true and estimated cells, respectively, by

$$A_j = (b_{j-1}, b_j], \qquad \widehat{A}_j = (\widehat{b}_{j-1}, \widehat{b}_j], \quad j = 1, \ldots, K.$$

The corresponding true cell probabilities are $p_j = P(A_j)$, whereas the empirical probabilities based on the estimated cells are $\widehat{p}_j = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \mathbf{1}\{X_i \in \widehat{A}_j\}$.

We show that $\widehat{p}_j \xrightarrow{\mathbb{P}} p_j$ for every $j = 1, \ldots, K$.

For each cell, introduce the empirical probability based on the fixed boundaries,

$$\widetilde{p}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \in A_j\}.$$

By the weak law of large numbers,

$$\widetilde{p}_j \xrightarrow{\mathbb{P}} p_j, \qquad j = 1, \ldots, K,$$

because K is fixed and the observations are independent and identically distributed.

It remains to establish that estimating the boundaries does not asymptotically change the

cell probabilities. For any two measurable sets A and B,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \in A\} - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \in B\}\right| \leq \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \in A \triangle B\},$$

where $A \triangle B$ denotes the symmetric difference of the two sets. Hence,

$$|\widehat{p}_j - \widetilde{p}_j| \leq \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{X_i \in \widehat{A}_j \triangle A_j\}.$$

Let $d_n = \max_{1\leq \ell < K} |\widehat{b}_\ell - b_\ell|$. By assumption, $d_n \xrightarrow{\mathbb{P}} 0$. For any fixed $\eta > 0$, on the event $\{d_n \leq \eta\}$, membership in $\widehat{A}_j$ can differ from membership in A_j only if the observation lies within distance η of one of the finite cell boundaries. Consequently,

$$\{\widehat{A}_j \triangle A_j\} \subseteq \bigcup_{\ell \in \{j-1,j\}\cap\{1,\ldots,K-1\}} [b_\ell - \eta, b_\ell + \eta]$$

on this event. It follows that

$$|\widehat{p}_j - \widetilde{p}_j| \leq \sum_{\ell \in \{j-1,j\}\cap\{1,\ldots,K-1\}} \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{|X_i - b_\ell| \leq \eta\}$$

whenever $d_n \leq \eta$.

Fix $\varepsilon > 0$. Since $P(\{b_\ell\}) = 0$ for every interior boundary and K is fixed, continuity of probability from above implies

$$P(|X - b_\ell| \leq \eta) \longrightarrow 0 \qquad \text{as } \eta \downarrow 0$$

for each $\ell = 1, \ldots, K-1$. Choose $\eta > 0$ sufficiently small that the sum of these boundary-neighborhood probabilities is less than $\varepsilon/2$. For this fixed η, the weak law of large numbers ensures that the corresponding empirical proportions converge in probability to their probabilities. Also, $P(d_n > \eta) \longrightarrow 0$. Combining these facts yields

$$|\widehat{p}_j - \widetilde{p}_j| \xrightarrow{\mathbb{P}} 0.$$

Finally, by the triangle inequality, $|\widehat{p}_j - p_j| \leq |\widehat{p}_j - \widetilde{p}_j| + |\widetilde{p}_j - p_j| \xrightarrow{\mathbb{P}} 0$.

Therefore,

$$\widehat{p}_j \xrightarrow{\mathbb{P}} p_j, \qquad j = 1, \ldots, K.$$

Since K is fixed, the convergence also holds jointly: $\max_{1\leq j\leq K} |\widehat{p}_j - p_j| \xrightarrow{\mathbb{P}} 0$. This completes the proof. □

B Appendix: Simulation Code

```
# ============================================================
# PARTITION-BASED PROBABILITY DISTRIBUTIONS
# Monte Carlo simulation for a theoretical statistics paper
# ============================================================

set.seed(20261027)

# ---------------------------
# 1. Simulation configuration
# ---------------------------

sample_sizes <- c(50, 100, 200, 500, 1000)
B <- 2000

# Number of cells for partition estimators.
# For computational efficiency, use K approximately n^(1/3)
# as a baseline, together with coarser and finer alternatives.
partition_factors <- c(0.5, 1, 2)

# Numerical integration / evaluation grids
grid_size <- 4001
support_eps <- 1e-8

# ------------------------------------------------------------
# 2. Data-generating distributions and true distribution functions
# ------------------------------------------------------------

r_model <- function(model, n) {
  switch(
    model,

    Uniform = runif(n, 0, 1),

    Discontinuous = {
      # Density 1/3 on [0,1), 2/3 on [1,2].
      # P(X < 1) = 1/3; P(X >= 1) = 2/3.
      ifelse(runif(n) < 1/3,
             runif(n, 0, 1),
             runif(n, 1, 2))
    },
```

```
    Discrete = sample(c(0, 1), n, replace = TRUE,
                      prob = c(0.4, 0.6)),

    Mixed = {
      is_atom <- runif(n) < 0.3
      x <- runif(n)
      x[is_atom] <- 0
      x
    },

    Multimodal = {
      # Well-separated normal mixture:
      # 0.35 N(-2, 0.45^2) + 0.65 N(2, 0.65^2)
      component <- rbinom(n, 1, 0.65)
      ifelse(component == 1,
             rnorm(n, mean = 2, sd = 0.65),
             rnorm(n, mean = -2, sd = 0.45))
    }
  )
}

true_cdf <- function(x, model) {
  switch(
    model,

    Uniform = pmin(pmax(x, 0), 1),

    Discontinuous = {
      ans <- numeric(length(x))
      ans[x < 0] <- 0
      i1 <- x >= 0 & x < 1
      i2 <- x >= 1 & x < 2
      i3 <- x >= 2
      ans[i1] <- x[i1] / 3
      ans[i2] <- 1/3 + (2/3) * (x[i2] - 1)
      ans[i3] <- 1
      ans
    },

    Discrete = 0.4 * (x >= 0) + 0.6 * (x >= 1),

    Mixed = {
```

```
      ans <- numeric(length(x))
      ans[x < 0] <- 0
      i1 <- x >= 0 & x < 1
      i2 <- x >= 1
      ans[i1] <- 0.3 + 0.7 * x[i1]
      ans[i2] <- 1
      ans
    },

    Multimodal = {
      0.35 * pnorm(x, mean = -2, sd = 0.45) +
        0.65 * pnorm(x, mean = 2, sd = 0.65)
    }
  )
}

true_density <- function(x, model) {
  switch(
    model,

    Uniform = ifelse(x >= 0 & x <= 1, 1, 0),

    Discontinuous = ifelse(
      x >= 0 & x < 1, 1/3,
      ifelse(x >= 1 & x <= 2, 2/3, 0)
    ),

    Discrete = rep(NA_real_, length(x)),

    Mixed = ifelse(x > 0 & x < 1, 0.7, 0),

    Multimodal = {
      0.35 * dnorm(x, mean = -2, sd = 0.45) +
        0.65 * dnorm(x, mean = 2, sd = 0.65)
    }
  )
}

# ------------------------------------------------------------
# 3. Common numerical evaluation grids
# ------------------------------------------------------------
```

```
model_grid <- function(model, m = grid_size) {
  switch(
    model,
    Uniform = seq(-0.05, 1.05, length.out = m),
    Discontinuous = seq(-0.05, 2.05, length.out = m),
    Discrete = seq(-0.05, 1.05, length.out = m),
    Mixed = seq(-0.05, 1.05, length.out = m),
    Multimodal = seq(-4.5, 5.0, length.out = m)
  )
}

model_density_grid <- function(model, m = grid_size) {
  switch(
    model,
    Uniform = seq(0, 1, length.out = m),
    Discontinuous = seq(0, 2, length.out = m),
    Mixed = seq(0, 1, length.out = m),
    Multimodal = seq(-4.5, 5, length.out = m),
    Discrete = numeric(0)
  )
}

# ------------------------------------------------------------
# 4. Partition construction
# ------------------------------------------------------------

make_breaks <- function(model, K, aligned = TRUE) {

  if (model == "Uniform") {
    return(seq(0, 1, length.out = K + 1))
  }

  if (model == "Discontinuous") {
    if (aligned) {
      # Ensure x = 1 is a partition boundary.
      K1 <- max(1, floor(K / 2))
      K2 <- max(1, K - K1)
      return(c(seq(0, 1, length.out = K1 + 1),
               seq(1, 2, length.out = K2 + 1)[-1]))
    } else {
      # Equal-width bins generally do not place x = 1
      # at a boundary when K is odd.
```

```
      return(seq(0, 2, length.out = K + 1))
    }
  }

  if (model == "Discrete" || model == "Mixed") {
    return(seq(0, 1, length.out = K + 1))
  }

  if (model == "Multimodal") {
    return(seq(-4.5, 5, length.out = K + 1))
  }
}

# ------------------------------------------------------------
# 5. CDF estimators
# ------------------------------------------------------------

cdf_empirical <- function(x, data) {
  ecdf(data)(x)
}

cdf_partition <- function(x, data, breaks) {
  # Assign observations to cells and move their probability
  # masses to cell midpoints.
  K <- length(breaks) - 1
  cell <- cut(data, breaks = breaks, include.lowest = TRUE,
              right = TRUE, labels = FALSE)

  # Values exactly at the leftmost endpoint are retained.
  cell[is.na(cell) & data == breaks[1]] <- 1
  cell[is.na(cell) & data == breaks[length(breaks)]] <- K

  counts <- tabulate(cell[!is.na(cell)], nbins = K)
  p_hat <- counts / length(data)
  mid <- (breaks[-1] + breaks[-length(breaks)]) / 2

  vapply(x, function(xx) sum(p_hat[mid <= xx]), numeric(1))
}

cdf_atom_preserving <- function(x, data, model, breaks) {
  # Known atoms:
  # Discrete: 0 and 1
```

```
  # Mixed: 0
  #
  # The continuous observations are partitioned separately.
  if (model == "Discrete") {
    p0 <- mean(data == 0)
    p1 <- mean(data == 1)
    return(p0 * (x >= 0) + p1 * (x >= 1))
  }

  if (model == "Mixed") {
    p0 <- mean(data == 0)
    cont <- data[data > 0]
    p_cont <- length(cont) / length(data)

    if (length(cont) == 0) {
      return(as.numeric(x >= 0))
    }

    K <- length(breaks) - 1
    cell <- cut(cont, breaks = breaks, include.lowest = TRUE,
                right = TRUE, labels = FALSE)
    cell[is.na(cell) & cont == breaks[1]] <- 1
    cell[is.na(cell) & cont == breaks[length(breaks)]] <- K

    counts <- tabulate(cell[!is.na(cell)], nbins = K)
    probs <- counts / length(data)
    mids <- (breaks[-1] + breaks[-length(breaks)]) / 2

    return(p0 * (x >= 0) +
             p_cont * vapply(
               x, function(xx) sum(probs[mids <= xx]), numeric(1)
             ))
  }

  if (model == "Discontinuous") {
    # There is a density jump but no point mass.
    # The partition is allowed to be aligned with x=1.
    return(cdf_partition(x, data, breaks))
  }

  cdf_partition(x, data, breaks)
}
```

```
cdf_parametric <- function(x, data, model) {
  # Correctly specified benchmark estimators.
  # Parameters are estimated by maximum likelihood or
  # empirical proportions, as appropriate.
  switch(
    model,

    Uniform = {
      # MLE for Uniform(a,b): min and max.
      a <- min(data)
      b <- max(data)
      if (b <= a) return(as.numeric(x >= a))
      pmin(pmax((x - a) / (b - a), 0), 1)
    },

    Discontinuous = {
      # Known support and known density structure; estimate
      # the probability in [0,1) by its sample proportion.
      p_left <- mean(data < 1)
      ans <- numeric(length(x))
      ans[x < 0] <- 0
      i1 <- x >= 0 & x < 1
      i2 <- x >= 1 & x < 2
      i3 <- x >= 2
      ans[i1] <- p_left * x[i1]
      ans[i2] <- p_left + (1 - p_left) * (x[i2] - 1)
      ans[i3] <- 1
      ans
    },

    Discrete = {
      p0 <- mean(data == 0)
      p0 * (x >= 0) + (1 - p0) * (x >= 1)
    },

    Mixed = {
      p0 <- mean(data == 0)
      cont <- data[data > 0]
      # The continuous component is Uniform(0,1).
      ans <- p0 * (x >= 0)
      ans <- ans + (1 - p0) * pmin(pmax(x, 0), 1)
```

```
      ans
    },

    Multimodal = {
      # Correct mixture family and known component standard
      # deviations; estimate mixing weight and component means.
      # To keep the benchmark stable, assign observations to
      # the nearest component centre (-2 or 2).
      z <- ifelse(abs(data + 2) < abs(data - 2), 0, 1)
      p2 <- mean(z == 1)
      mu1 <- mean(data[z == 0])
      mu2 <- mean(data[z == 1])

      p2 * pnorm(x, mean = mu2, sd = 0.65) +
        (1 - p2) * pnorm(x, mean = mu1, sd = 0.45)
    }
  )
}

# ---------------------------------------------------------------
# 6. Density estimators
# ---------------------------------------------------------------

density_partition <- function(data, breaks) {
  K <- length(breaks) - 1
  widths <- diff(breaks)

  cell <- cut(data, breaks = breaks, include.lowest = TRUE,
              right = TRUE, labels = FALSE)
  cell[is.na(cell) & data == breaks[1]] <- 1
  cell[is.na(cell) & data == breaks[length(breaks)]] <- K

  counts <- tabulate(cell[!is.na(cell)], nbins = K)
  heights <- counts / (length(data) * widths)

  list(breaks = breaks, heights = heights)
}

eval_density_partition <- function(x, estimator) {
  breaks <- estimator$breaks
  heights <- estimator$heights
  idx <- findInterval(x, breaks, rightmost.closed = TRUE)
```

```
  out <- numeric(length(x))
  ok <- idx >= 1 & idx <= length(heights)
  out[ok] <- heights[idx[ok]]
  out
}

density_histogram <- function(data, xgrid) {
  h <- hist(data, breaks = "FD", plot = FALSE, right = TRUE)
  idx <- findInterval(xgrid, h$breaks, rightmost.closed = TRUE)
  out <- numeric(length(xgrid))
  ok <- idx >= 1 & idx <= length(h$density)
  out[ok] <- h$density[idx[ok]]
  out
}

# ------------------------------------------------------------
# 7. Performance measures
# ------------------------------------------------------------

uniform_cdf_error <- function(Fhat, Ftrue) {
  max(abs(Fhat - Ftrue))
}

wasserstein_1 <- function(Fhat, Ftrue, xgrid) {
  # Numerical integral of absolute CDF difference.
  sum((abs(Fhat[-1] - Ftrue[-1]) +
         abs(Fhat[-length(Fhat)] - Ftrue[-length(Ftrue)])) / 2 *
        diff(xgrid))
}

density_ise <- function(fhat, ftrue, xgrid) {
  if (anyNA(ftrue)) return(NA_real_)
  sum(((fhat[-1] - ftrue[-1])^2 +
         (fhat[-length(fhat)] - ftrue[-length(ftrue)])^2) / 2 *
        diff(xgrid))
}

true_jump_sizes <- function(model) {
  switch(
    model,
    Discrete = c(`0` = 0.4, `1` = 0.6),
    Mixed = c(`0` = 0.3),
```

```
    c()
  )
}

estimated_jump_sizes <- function(data, model) {
  switch(
    model,
    Discrete = c(`0` = mean(data == 0),
                 `1` = mean(data == 1)),
    Mixed = c(`0` = mean(data == 0)),
    c()
  )
}

# ------------------------------------------------------------
# 8. One replication
# ------------------------------------------------------------

one_replication <- function(model, n, B_rep) {

  data <- r_model(model, n)
  xgrid <- model_grid(model)
  Ftrue <- true_cdf(xgrid, model)

  results <- list()

  # Empirical CDF
  F_emp <- cdf_empirical(xgrid, data)
  results[["Empirical CDF"]] <- list(
    F = F_emp, f = NULL
  )

  # Parametric benchmark
  F_par <- cdf_parametric(xgrid, data, model)
  results[["Parametric"]] <- list(
    F = F_par, f = NULL
  )

  # Partition estimators
  K_base <- max(2, round(n^(1/3)))

  for (factor in partition_factors) {
```

```
  K <- max(2, round(K_base * factor))

  if (model == "Discontinuous") {
    for (aligned in c(TRUE, FALSE)) {
      breaks <- make_breaks(model, K, aligned = aligned)
      F_part <- cdf_partition(xgrid, data, breaks)
      nm <- paste0("Partition K=", length(breaks) - 1,
                   if (aligned) " aligned" else " unaligned")
      results[[nm]] <- list(F = F_part, f = NULL)
    }
  } else {
    breaks <- make_breaks(model, K, aligned = TRUE)
    F_part <- cdf_partition(xgrid, data, breaks)
    nm <- paste0("Partition K=", length(breaks) - 1)
    results[[nm]] <- list(F = F_part, f = NULL)
  }

  # Atom-preserving method for discrete and mixed models.
  if (model %in% c("Discrete", "Mixed")) {
    breaks <- make_breaks(model, K, aligned = TRUE)
    F_atom <- cdf_atom_preserving(xgrid, data, model, breaks)
    nm <- paste0("Atom-preserving K=", K)
    results[[nm]] <- list(F = F_atom, f = NULL)
  }

  # Piecewise-constant density, where appropriate.
  if (model %in% c("Uniform", "Discontinuous",
                   "Mixed", "Multimodal")) {
    breaks <- make_breaks(model, K, aligned = TRUE)
    den <- density_partition(data, breaks)
    fgrid <- model_density_grid(model)
    ftrue <- true_density(fgrid, model)
    fhat <- eval_density_partition(fgrid, den)

    results[[paste0("Piecewise density K=", K)]] <-
      list(F = NULL, f = list(
        estimate = fhat,
        truth = ftrue,
        grid = fgrid
      ))
  }
}
```

```
# Ordinary data-selected histogram density
if (model %in% c("Uniform", "Discontinuous",
                 "Mixed", "Multimodal")) {
  fgrid <- model_density_grid(model)
  ftrue <- true_density(fgrid, model)
  fhist <- density_histogram(data, fgrid)
  results[["FD histogram"]] <- list(
    F = NULL,
    f = list(estimate = fhist, truth = ftrue, grid = fgrid)
  )
}

# Evaluate each estimator.
out <- list()
jumps_true <- true_jump_sizes(model)
jumps_hat <- estimated_jump_sizes(data, model)

for (nm in names(results)) {
  obj <- results[[nm]]

  cdf_err <- NA_real_
  w1 <- NA_real_
  ise <- NA_real_

  if (!is.null(obj$F)) {
    cdf_err <- uniform_cdf_error(obj$F, Ftrue)
    w1 <- wasserstein_1(obj$F, Ftrue, xgrid)
  }

  if (!is.null(obj$f)) {
    ise <- density_ise(
      obj$f$estimate,
      obj$f$truth,
      obj$f$grid
    )
  }

  jump_bias <- NA_real_
  jump_mse <- NA_real_

  if (length(jumps_true) > 0) {
```

```
      common <- intersect(names(jumps_true), names(jumps_hat))
      if (length(common) > 0) {
        errors <- jumps_hat[common] - jumps_true[common]
        jump_bias <- mean(errors)
        jump_mse <- mean(errors^2)
      }
    }

    out[[length(out) + 1]] <- data.frame(
      model = model,
      n = n,
      estimator = nm,
      cdf_error = cdf_err,
      wasserstein1 = w1,
      density_ISE = ise,
      jump_bias = jump_bias,
      jump_MSE = jump_mse,
      stringsAsFactors = FALSE
    )
  }

  do.call(rbind, out)
}

# ------------------------------------------------------------
# 9. Run the full simulation
# ------------------------------------------------------------

models <- c("Uniform", "Discontinuous", "Discrete",
            "Mixed", "Multimodal")

all_results <- vector("list", length(models) *
                        length(sample_sizes) * B)

counter <- 1L
start_time <- Sys.time()

for (model in models) {
  for (n in sample_sizes) {
    cat("Running:", model, "| n =", n, "\n")

    for (b in seq_len(B)) {
```

```
      all_results[[counter]] <-
        one_replication(model, n, b)
      counter <- counter + 1L
    }
  }
}

raw_results <- do.call(rbind, all_results)
raw_results <- raw_results[!vapply(all_results, is.null, logical(1)), ]

elapsed <- difftime(Sys.time(), start_time, units = "mins")
cat("\nElapsed time (minutes):", round(as.numeric(elapsed), 2), "\n")

# ------------------------------------------------------------
# 10. Monte Carlo summaries
# ------------------------------------------------------------

mc_summary <- function(x) {
  x <- x[is.finite(x)]
  if (length(x) == 0) {
    return(c(mean = NA_real_, MCSE = NA_real_, SD = NA_real_))
  }
  c(
    mean = mean(x),
    MCSE = sd(x) / sqrt(length(x)),
    SD = sd(x)
  )
}

metrics <- c("cdf_error", "wasserstein1", "density_ISE",
             "jump_bias", "jump_MSE")

group_key <- interaction(raw_results$model,
                         raw_results$n,
                         raw_results$estimator,
                         drop = TRUE)

groups <- split(raw_results, group_key)

summary_rows <- lapply(groups, function(d) {
  row <- data.frame(
    model = d$model[1],
```

```
    n = d$n[1],
    estimator = d$estimator[1],
    stringsAsFactors = FALSE
  )

  for (metric in metrics) {
    s <- mc_summary(d[[metric]])
    row[[paste0(metric, "_mean")]] <- s["mean"]
    row[[paste0(metric, "_MCSE")]] <- s["MCSE"]
    row[[paste0(metric, "_SD")]] <- s["SD"]
  }

  row
})

summary_results <- do.call(rbind, summary_rows)
rownames(summary_results) <- NULL

# Display compact results
print(summary_results, row.names = FALSE)

# ------------------------------------------------------------
# 11. Save numerical results for manuscript tables
# ------------------------------------------------------------

write.csv(
  raw_results,
  file = "partition_simulation_raw_results.csv",
  row.names = FALSE
)

write.csv(
  summary_results,
  file = "partition_simulation_summary.csv",
  row.names = FALSE
)

# ------------------------------------------------------------
# 12. Optional: focused comparison of atom preservation
# ------------------------------------------------------------

atom_results <- subset(
```

```
  summary_results,
  model %in% c("Discrete", "Mixed") &
    grepl("Empirical CDF|Partition|Atom-preserving|Parametric",
          estimator)
)

atom_results <- atom_results[
  order(atom_results$model, atom_results$n,
        atom_results$estimator),
]

cat("\n===== ATOM-PRESERVATION COMPARISON =====\n")
print(atom_results, row.names = FALSE)

write.csv(
  atom_results,
  file = "partition_atom_preservation_summary.csv",
  row.names = FALSE
)

# ------------------------------------------------------------
# 13. Optional: compact convergence plots
# ------------------------------------------------------------

plot_metric <- function(model_name, metric = "cdf_error_mean") {
  d <- subset(summary_results, model == model_name)
  d <- d[is.finite(d[[metric]]), ]

  if (nrow(d) == 0) {
    warning("No finite values for this model and metric.")
    return(invisible(NULL))
  }

  estimators <- unique(d$estimator)
  cols <- seq_along(estimators)

  plot(
    NA,
    xlim = range(sample_sizes),
    ylim = range(d[[metric]], finite = TRUE),
    log = "x",
    xlab = "Sample size n",
```

```
    ylab = metric,
    main = paste(model_name, "-", metric)
  )

  for (i in seq_along(estimators)) {
    di <- d[d$estimator == estimators[i], ]
    di <- di[order(di$n), ]
    lines(di$n, di[[metric]], type = "b", pch = 16,
          lty = i, col = cols[i])
  }

  legend(
    "topright",
    legend = estimators,
    col = cols,
    pch = 16,
    lty = seq_along(estimators),
    cex = 0.65,
    bty = "n"
  )
}

# Example plots:
plot_metric("Uniform", "cdf_error_mean")
plot_metric("Discontinuous", "wasserstein1_mean")
plot_metric("Discrete", "cdf_error_mean")
plot_metric("Mixed", "jump_MSE_mean")
plot_metric("Multimodal", "density_ISE_mean")
```