Investigation of a Non-parametric Quantile Transformation Technique for Density Estimation

Caimin Ayres

March 31, 2025

Abstract

This paper introduces a non-parametric density estimation method based on quantile transformation, which models the data distribution as a smooth monotonic transformation of a reference distribution. Unlike standard approaches such as kernel density estimation, the proposed method incorporates a reference distribution, directly into the estimation process. This reflects how many statistical procedures, particularly in regression, rely on reference distributions for diagnostics and inference. By aligning empirical quantiles of the data with those of the reference distribution, a transformation function is estimated using B-splines. The estimated transformation is then used to recover the data density through a change-of-variable formula. Simulation studies on Gamma, t, and Weibull distributions evaluate the estimator's performance against classical methods in terms of integrated squared error, consistency, and relative efficiency. Results show improved accuracy, highlighting the value of incorporating reference structure into non-parametric estimation.

Acknowledgements

I would like to sincerely thank Professor Finbarr O'Sullivan for his invaluable guidance, unwavering patience, and generous dedication of time throughout the course of this project. His insight and support were instrumental at every stage of the work.

Contents

Chapter 1

Introduction

In regression and generalised linear model situations, practitioners routinely work with parametric reference distributions that provide foundational assumptions. The quality of inferences for regression coefficients $\hat{\beta}$ depends critically on these distributional assumptions. Although $\hat{\beta}$ has asymptotic behaviour regardless of the error distribution, when residuals deviate from the assumed distribution, inferences such as confidence intervals and hypothesis tests may be unreliable. For instance, when the error distribution has heavier tails than the assumed normal distribution, t-tests produce confidence intervals that are too narrow, leading to inflated Type I error rates. While diagnostic tools like Q-Q plots leverage these reference distributions, existing density estimation methods typically approach the problem without explicitly incorporating such prior knowledge. This disconnect between how models are diagnosed and how densities are estimated suggests an opportunity: methods that systematically explore departures from an assumed parametric reference distribution could provide more nuanced insights into how observed data deviate from baseline assumptions.

Various approaches exist to assess the accuracy of the assumption of the error distribution. These include the Q-Q plot, Shapiro-Wilk test (Shapiro and Wilk, 1965), Kolmogorov-Smirnov test (Kolmogorov, 1933 and Smirnov, 1948), and even just plotting the histogram. Both Q-Q plots and histograms are not exact sciences and leave room for subjective interpretation. The aforementioned Shapiro–Wilk and Kolmogorov–Smirnov tests, while providing a more formal approach, ultimately yield a binary outcome (reject or fail to reject normality), thereby limiting the available diagnostic detail. In addition, the Kolmogorov–Smirnov test can be sensitive to deviations in certain regions of the distribution and typically assumes that the parameters of the reference distribution are known , whereas the Shapiro–Wilk test is specifically designed to assess normality and may perform poorly with small sample sizes. Consequently, these diagnostic tools have significant limitations that warrant improvement.

When confronted with non-normally distributed errors, practitioners often resort to transforming the response variable. This approach has a long history, with variance-stabilising transformations (Bartlett, 1947), power transformations such as the Box-Cox transformation (Box and Cox, 1964), and specific transformations like square root or logarithmic transformations being commonly employed. The

underlying principle is that by applying an appropriate transformation, the distribution of the transformed response becomes more amenable to standard analysis techniques.

However, transformations present their own challenges. Perhaps most significantly, they change the scale of analysis, which can complicate the interpretation of results. Bland & Altman (Bland and Altman, 1996) noted that transforming data produces results in terms of geometric rather than arithmetic means, fundamentally changing what is being measured and compared. As demonstrated by Feng et al., 2014, log-transforms can alter the mean of the data even after transforming back to the original scale. This transformation bias, discussed by Sakia (Sakia, 1992), can sometimes be accounted for with bias adjustment, though not without complications. Taylor (Taylor, 1986) demonstrated that approximating the mean after Box-Cox transformation becomes particularly inaccurate when the transformation parameter is close to zero or when error variance is large.

Box & Cox (Box and Cox, 1964) themselves acknowledged these interpretation challenges, which led to extensive debate on whether inferences should be made in the original or transformed scale (Box and Cox, 1982; Hinkley and Runger, 1984). Carroll & Ruppert (Carroll and Ruppert, 1981) quantified this as a 'cost' to estimating the transformation parameter when making inferences in the original scale, acknowledging that important information in the original scale may be obscured or lost through transformation. Despite these well-documented issues, it is common practice for the justification of using transformations to be based more in tradition rather than on robust theoretical justification.

This paper proposes a novel approach to addressing these distributional challenges through quantile relationships. Instead of directly transforming data or relying on traditional diagnostic tools, the method examines the relationship between distributions through their quantiles. An approach that aligns naturally with statistical practice, where quantile-quantile plots are already a standard diagnostic tool.

The method is built on a fundamental insight: there exists a one-to-one function h that maps between a reference distribution X and the data distribution Y. This invertible relationship, where $Y = h(X)$ and $X = h^{-1}(Y)$, provides a natural diagnostic tool: deviations from the reference distribution are revealed through departures of h from the identity function. When the data follows the reference distribution, h will be approximately linear; any nonlinearity in h indicates specific ways in which the data distribution differs from the reference. This same mapping also enables estimation of the underlying density of the data which will be detailed in Section 2.

Density estimation itself is far from a new problem. Histograms, kernel estimators (Rosenblatt, 1956), nearest neighbours (Cover and Hart, 1967), and spline-based methods (Mächler, 1996) are all established approaches. However, these methods are typically treated as separate from regression diagnostics. The

proposed approach unifies distribution checking and non-parametric density estimation into a single framework via quantiles which is demonstrated in Section 3. The mapping h is implemented using spline smoothing techniques (Silverman, 1986; Wahba, 1990), ensuring both flexibility and controlled smoothness. This paper details the construction and fitting of the spline-based map h, illustrates its numerical behaviour through simulations, and compares it against the classical kernel density estimator and the parametric `fitdistr` function in R. The results demonstrate that this quantile-based transformation serves both as a rigorous check on distributional assumptions in regression and as a competitive density estimator in its own right.

Chapter 2

Theory

2.1 Transformation Theory

The proposed approach to density estimation and distributional diagnostics relies on establishing a mapping between distributions. This section develops the theoretical framework for such mappings through two fundamental types of transformations: random variable transformations and their corresponding density transformations.

2.1.1 Random Variable Transformation

Let X be a continuous random variable with cumulative distribution function (CDF):

$$F_X(x) = \mathbb{P}[X \leq x]$$

Consider a transformation h that is continuous and strictly monotonic everywhere, mapping X to a new random variable $Y = h(X)$. The strict monotonicity ensures that h is invertible, a property important for establishing a one-to-one relationship between distributions. The CDF of Y can be derived as:

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[h(X) \leq y] = \mathbb{P}[X \leq h^{-1}(y)] = F_X(h^{-1}(y))$$

2.1.2 Density Transformation

The probability density function (PDF) of Y is obtained by differentiating its CDF with respect to y. Applying the chain rule:

$$f_Y(y) = F'_Y(y) = \frac{d}{dy} F_X(h^{-1}(y)) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y)$$

For a strictly monotonic transformation, the density transformation formula includes the absolute value of the derivative:

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| \tag{2.1}$$

The absolute value term ensures the density remains non-negative, regardless of whether h is monotonically increasing or decreasing. This is because probability

density functions must be non-negative by definition. When h is monotonically increasing, $\frac{d}{dy}h^{-1}(y)$ is positive, and when h is monotonically decreasing, $\frac{d}{dy}h^{-1}(y)$ is negative, but the absolute value ensures the resulting density is always positive.

2.2 Quantile-Based Mapping Between Data and a Reference Distribution

The transformation theory developed in the previous section requires a specific function h to map between distributions. A natural approach to constructing this mapping utilises the quantile function, defined as the inverse of the CDF:

$$Q_X(p) = F_X^{-1}(p), \quad p \in (0, 1)$$

where $Q_X(p)$ is the point below which a fraction p of the observations from the reference distribution X lie.

Under this framework, the mapping h establishes a relationship between the quantiles of the reference distribution X and those of the target data distribution Y, enabling a direct transformation $Y = h(X)$.

Figure 2.1 demonstrates this relationship using standardised Gamma distributions as examples each compared to a standard normal reference. The left panel presents quantile-quantile (Q-Q) plots, where deviations from the diagonal line indicate departures from normality. The right panel shows the corresponding probability density functions. These plots reveal two critical properties of the quantile relationships:

- The mappings maintain smoothness and monotonicity across different parameter values, even when the distributions differ substantially from normality
- The degree and nature of nonlinearity systematically reflect the underlying distributional differences, as evidenced by the varying curves for different Gamma parameters

These properties suggest that smooth, flexible functions could effectively approximate the transformation h. In particular, the consistent smoothness of these relationships motivates the use of B-splines, which can capture the non-linear patterns while maintaining the smoothness essential for reliable density estimation.

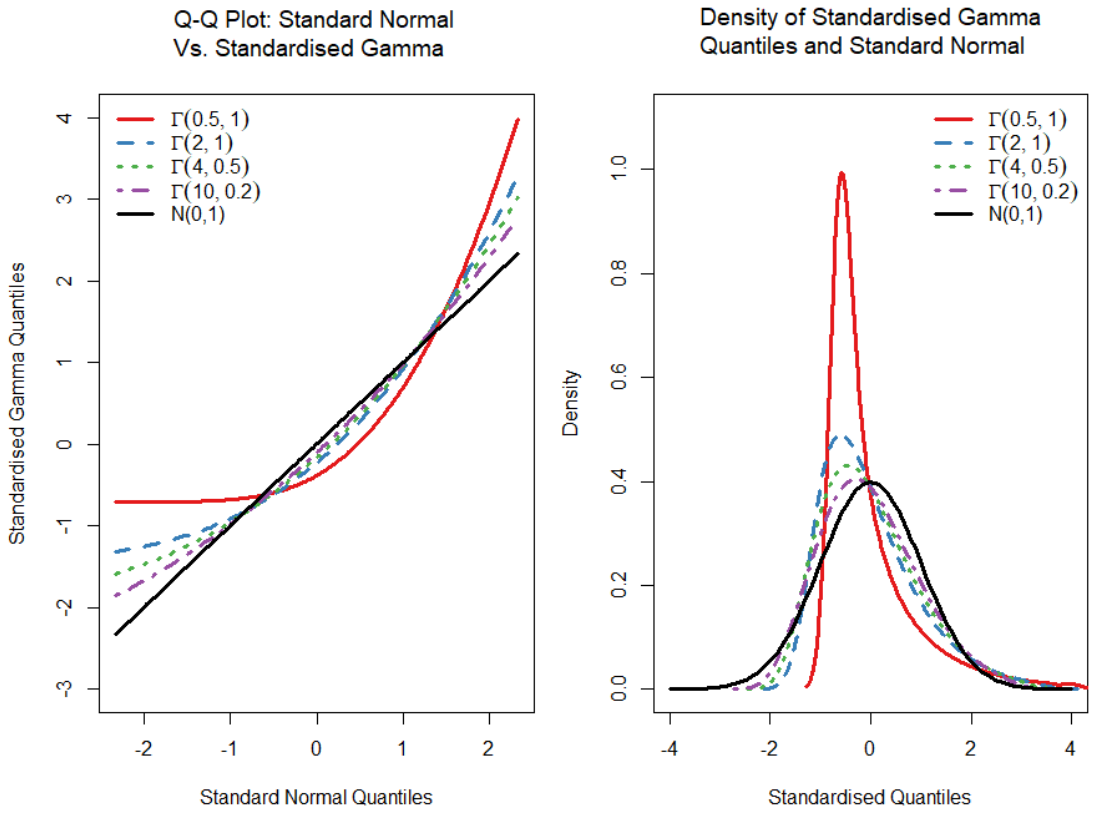

Figure 2.1: Left: Q–Q plots comparing standard normal quantiles to standardised Gamma quantiles for various shape and scale parameters. Right: Corresponding probability density functions.

2.3 Function Approximation with B-Splines

The quantile mapping introduced in Section 2.2 requires estimating a transformation function h such that $Y = h(X)$. This is not merely a curve-fitting task but also underpins the broader goal of recovering the data density from a reference distribution via the change-of-variable formula. This can be formulated as a function estimation problem where

$$y = h(x) + \epsilon$$

with ϵ representing the inherent variability in the quantile estimates. The smooth, monotonic nature of the relationships observed in Figure 2.1 suggests that B-splines are particularly suitable for this estimation.

B-splines provide a flexible and computationally efficient framework for function approximation (de Boor, 1978). A B-spline of degree k consists of piecewise polynomials joined at knot points, with $(k-1)$ continuous derivatives at these joins. This smoothness, combined with their flexibility, makes B-splines well-suited for modelling the non-linear characteristic of the quantile mapping. Figure 2.2 illustrates a fitted B-spline capturing the transformation between a standard normal and a sample of quantiles from a Gamma distribution

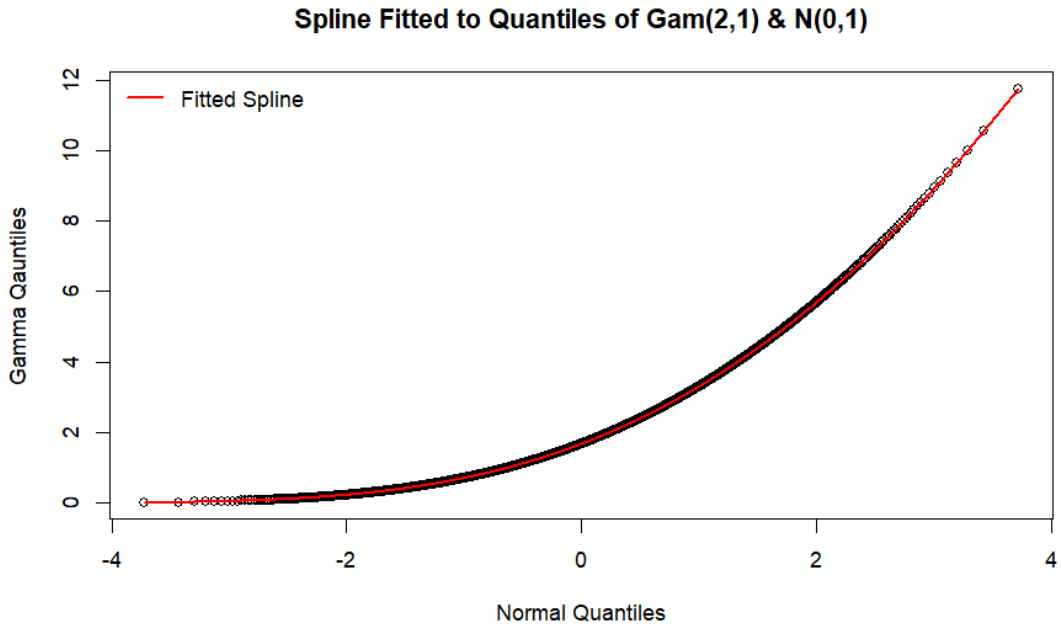

Figure 2.2: Estimation of the transformation function h between $N(0,1)$ and $\Gamma(2,1)$ quantiles using cubic B-splines. The smooth curve captures the non-linear relationship while maintaining the required monotonicity.

Following Wahba's framework for smoothing splines (Wahba, 1990), the estimation of the transformation function h is posed as a regularised variational problem. The objective is to balance fidelity to the observed quantile pairs (x_i, Y_i) with a smoothness constraint on the estimated function. This leads to the minimisation of the penalised weighted least squares functional:

$$\hat{h} = \arg\min_{h \in \mathcal{H}} \left\{ \sum_{i=1}^{n} w_i \left(Y_i - h(x_i)\right)^2 + \lambda \int \left(h''(x)\right)^2 \, dx \right\},$$

where $\mathcal{H}$ is a Sobolev space of twice-differentiable functions, w_i are weights informed by the reference distribution, and $\lambda > 0$ is a smoothing parameter that regulates the roughness penalty via the integrated squared second derivative.

This formulation ensures that the estimator $\hat{h}$ lies in the class of natural cubic splines with knots at the observed quantiles. The first term in the objective function encourages agreement between the estimated transformation and the data, while the second penalises excessive curvature to enforce smoothness and prevent overfitting. In practice, the spline is efficiently computed using R's `smooth.spline` function, with smoothness controlled via the degrees of freedom parameter, which can be tuned through cross-validation or similar criteria. This spline-based quantile transformation plays a central role in the method, as it enables density recovery via inversion and the change-of-variable formula.

The smoothing operation can be expressed in matrix form as:

$$\mathbf{S} = \mathbf{N}(\mathbf{N}^T\mathbf{W}\mathbf{N} + \lambda\mathbf{K})^{-1}\mathbf{N}^T\mathbf{W}$$

where $\mathbf{N}$ is the incidence matrix the B-spline basis, $\mathbf{W}$ is a diagonal weight matrix,

and $\mathbf{K}$ is the kernel of the roughness penalty for cubic spline smoothing. This matrix $\mathbf{S}$ defines a linear smoother mapping the observed quantile data to the fitted values. Its structure depends on both the B-spline basis $\mathbf{N}$ and the penalty matrix $\mathbf{K}$, with λ modulating the balance between fidelity and smoothness. In the implementation, predictor values are rescaled to the interval $[0, 1]$ and the weights are normalised to sum to n. In these rescaled units, λ in the original scale relates to a factor α via

$$s = \left((\max(x_i) - \min(x_i))^3 \, n^{-1} \, \sum w_i\right) \alpha,$$

so that $\lambda = \alpha s$ (Green and Silverman, 1993).

One key advantage of the quantile-based formulation is that explicit monotonicity constraints are often unnecessary, since the underlying quantile function is inherently non-decreasing. He and Shi, 1998 show that even if monotonicity is enforced, the asymptotic convergence rate remains essentially the same, lending additional theoretical support for an unconstrained spline fitting approach.

Under standard regularity assumptions (e.g., $h \in C^r$), spline estimators achieve an L^2 error rate of $O(n^{-2r/(2r+1)})$ (Silverman, 1986). In the case of cubic splines with $r = 2$, this yields $O(n^{-4/5})$. Though this rate is slower than what might be achieved under stronger parametric assumptions, it reflects the added flexibility of capturing arbitrary distributional shapes. Theoretical results indicate that for cubic splines, the effective degrees of freedom (or equivalently, the number of knots) should grow on the order of $n^{1/5}$ to optimally balance bias and variance (Silverman, 1986; Wahba, 1990).

Moreover, modern software typically allows practitioners to specify the degrees of freedom rather than λ directly, providing a more intuitive control over model complexity. This aligns with standard smoothing spline implementations in R and other statistical environments, making the approach both theoretically sound and practically accessible. While the basic framework here treats the weights as fixed, Section 3.4 will generalise these ideas by iteratively refining the weights as the density estimate evolves. Selection of the smoothing parameter is explored in Section 3.3.

2.4 Variability of the Quantile Function

The estimation of quantiles introduces inherent variability that must be accounted for in the spline fitting procedure. The error term ϵ in the model $y = h(x) + \epsilon$ has a non-constant variance structure which is evident in the asymptotic distribution of sample quantiles:

$$\sqrt{n}(\hat{Q}_p - Q_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{n f(Q_p)^2}\right)$$

where $\hat{Q}_p$ is the sample quantile, Q_p is the true quantile at probability p, and $f(Q_p)$ is the density at that quantile (Arnold et al., 1992). This result provides the theoretical foundation for incorporating the uncertainty of quantile estimates into the fitting procedure.

The asymptotic variance structure suggests a natural weighting scheme for the

spline estimation. Since optimal weights in weighted least squares are inversely proportional to the variance, this yields weights of the form:

$$w(p) = \frac{n f_X(Q_p)^2}{p(1-p)} \tag{2.2}$$

where f_X is the density of the reference distribution and $Q_p = F_X^{-1}(p)$ are its quantiles. This leads to a weighted least squares approach that accounts for the heteroscedastic nature of quantile estimates. These weights ensure that regions of the distribution where quantile estimates have lower variance typically near the centre of the distribution receive greater emphasis in the fitting procedure.

This weighting scheme not only is theoretically optimal under the asymptotic theory, but also helps protect against overfitting by reducing the influence of more variable tail quantiles on the fitted spline.

A practical challenge in implementing this weighting scheme arises because the weights depend on the unknown density $f_Y(Q_p)$. This circular dependency is addressed through an Iterative procedure detailed in Section 3.4. The effectiveness of this weighting approach and its impact on estimation accuracy are examined through simulation studies in Section 4.2.

Chapter 3

The Quantile Density Estimator (QDE) Algorithm

3.1 Reference Distribution

As mentioned in the introduction, this method differs from traditional density estimation approaches by incorporating a reference distribution into the estimation procedure. The reference distribution can be thought of as a prior belief about the underlying distribution of the data. This choice should reflect characteristics similar to those expected in the data, as it directly influences the estimated density through the transformation function (2.1).

The normal distribution is often a natural choice as reference distribution due to its well-studied properties and ubiquitous appearance across many fields. However, other distributions may be more appropriate depending on the application context. For example, in financial applications, heavy-tailed distributions like the skewed generalised t-distribution have been shown to better capture the characteristics of asset returns (Theodossiou, 1998). In insurance contexts, distributions such as the lognormal, Pareto, Burr, and Gamma distributions are commonly used for modelling property insurance losses (Burnecki et al., 2000).

The flexibility of this method allows practitioners to incorporate such domain knowledge through their choice of reference distribution.

3.2 Main QDE Procedure

This section describes the main steps of the Quantile Density Estimator (QDE).

1. **Determine Quantiles:** Calculate the sample quantiles $Q_Y(p_i)$ of the data and the theoretical quantiles $Q_X(p_i)$ of the chosen reference distribution F_X for a sequence of probabilities p_i for $i = 1, \ldots, n$.

2. **Assign Weights:** Calculate weights according to (2.2):

$$w_i(p_i) = \frac{n f_X(Q_X(p_i))^2}{p_i(1-p_i)}$$

   where $f_X(\cdot)$ is the density of the reference distribution evaluated at its quantiles.

3. **Fit a Spline:** Using R's `smooth.spline` function with an appropriate degrees of freedom selected through cross-validation or by optimising a suitable loss function, fit a smoothing spline $\hat{h}(\cdot)$ to the pairs $(Q_X(p_i), Q_Y(p_i))$ using the weights w_i. This relationship is demonstrated in the top-right panel of Figure 3.1, where the sample quantiles (circles) are connected by the fitted spline (red line).

4. **Transform to Obtain Density:** The density estimate $\hat{f}(y)$ is obtained through the transformation formula (2.1):

$$\hat{f}(y) = \frac{f_X(\hat{h}^{-1}(y))}{\hat{h}'(\hat{h}^{-1}(y))}$$

   This procedure requires evaluating:

   - $\hat{h}'(\cdot)$, the derivative of the fitted spline
   - $f_X(\cdot)$, the reference density
   - $h^{-1}(y)$ the inverse of the spline at y

   The bottom-left panel of Figure 3.1 compares this estimated density (dashed red line) with the true $\Gamma(1.2, 2)$ density (solid black line).

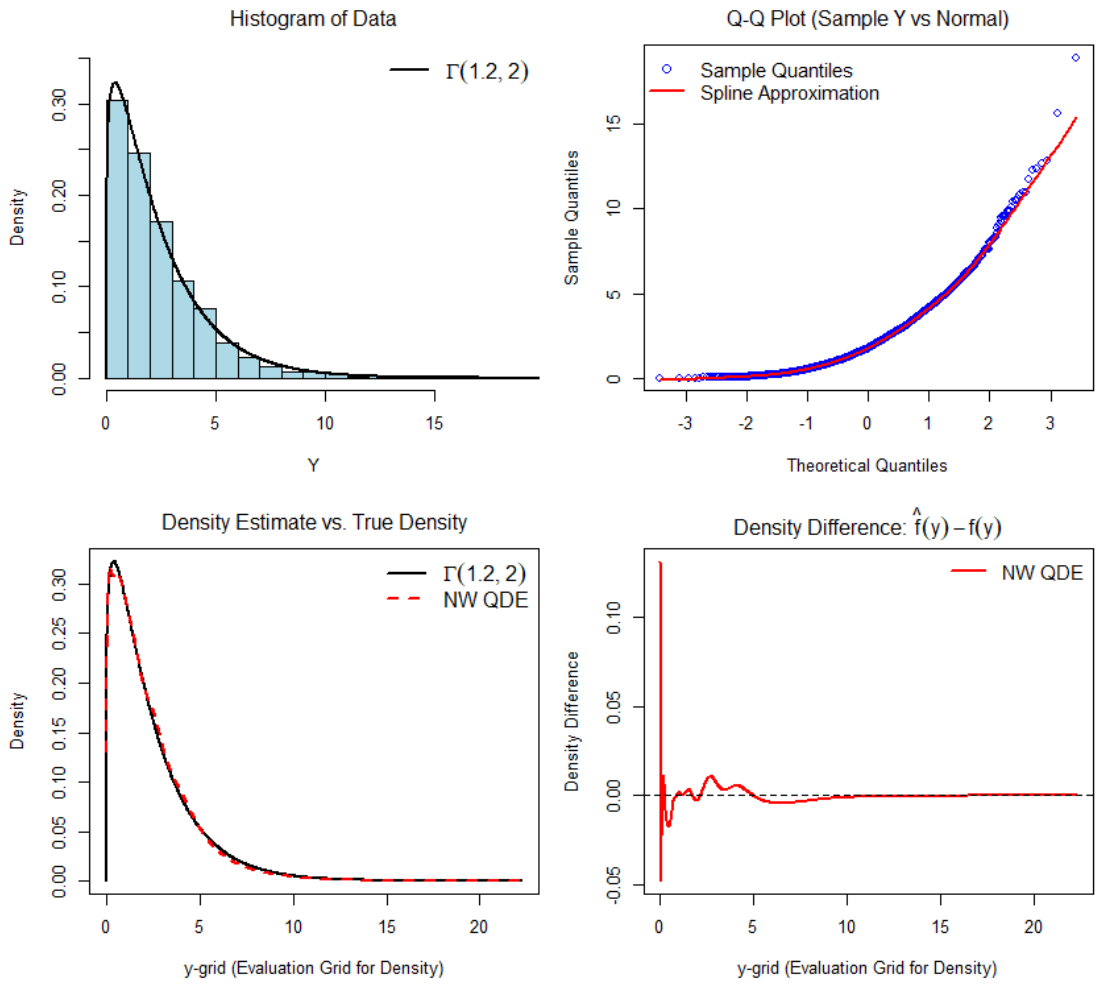

Figure 3.1: Illustration of the density estimation process using a $\Gamma(1.2, 2)$ sample of size 1600. Top left: histogram of the sample data with true density overlaid. Top right: Q-Q plot showing the relationship between sample and reference distribution quantiles with fitted spline. Bottom left: comparison of the estimated density NW QDE, (using normal distribution as reference) with the true Gamma density. Bottom right: pointwise difference between estimated and true densities, showing the accuracy of the fit.

3.3 Parameter Tuning

The QDE method requires careful selection of the degrees of freedom parameter to balance the bias-variance trade-off inherent in non-parametric density estimation. This parameter, which controls the smoothness of the fitted spline, significantly impacts the quality of the density estimate.

Figure 3.2 demonstrates how the integrated squared error (ISE), a measure of estimation accuracy, varies with different choices of degrees of freedom. At low degrees of freedom ($df < 5$), the spline is too rigid to capture the underlying density structure, resulting in high ISE values. The optimal performance occurs at $df = 10.85$, where the spline has sufficient flexibility to capture the true density without overfitting. Beyond this optimal point, increasing degrees of freedom leads to gradually deteriorating performance as the spline begins to fit noise in the data.

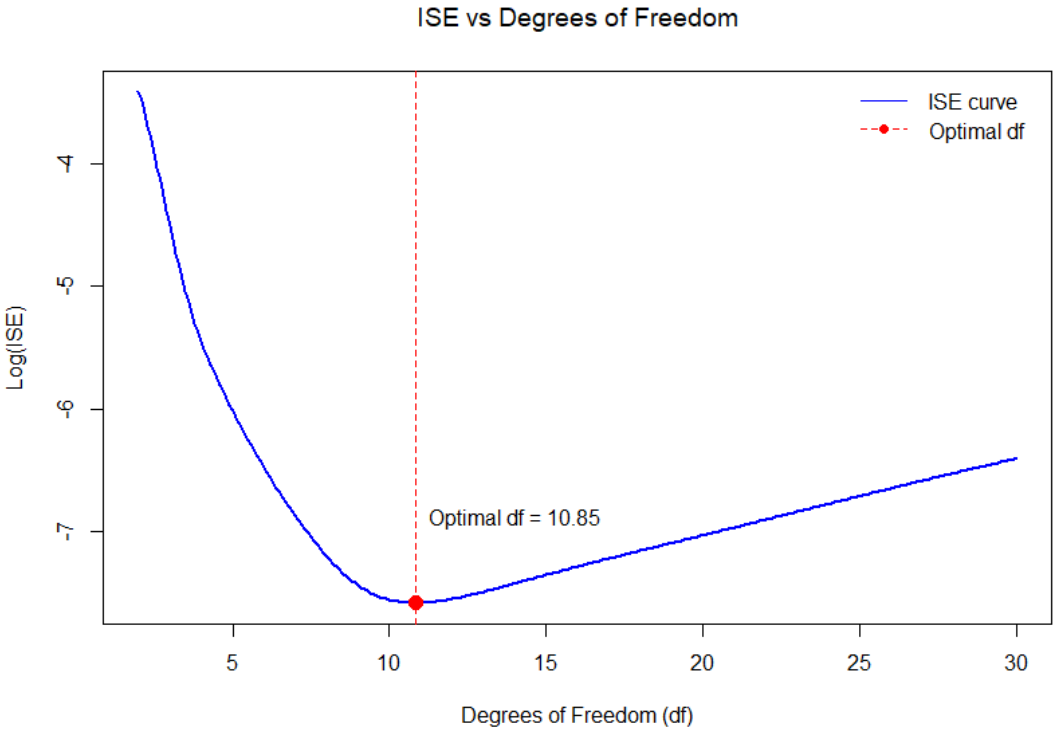

Figure 3.2: Log of the integrated squared error between the estimated density and the true density plotted against choice of degrees of freedom. Data is generated from a $\Gamma(1.2, 2)$ distribution with 1600 observations and set.seed = 9, using the Normal Weights method.

The impact of different degrees of freedom choices is illustrated in Figure 3.3. With $df = 4$, the spline is under-smoothed, resulting in a rigid fit that fails to capture the true shape of the density. The optimal choice ($df = 10.85$) provides a balanced fit that captures the essential features of the density while avoiding excessive noise. At $df = 40$, the spline is over-smoothed, leading to fluctuations in the density estimate, particularly visible in the tail regions.

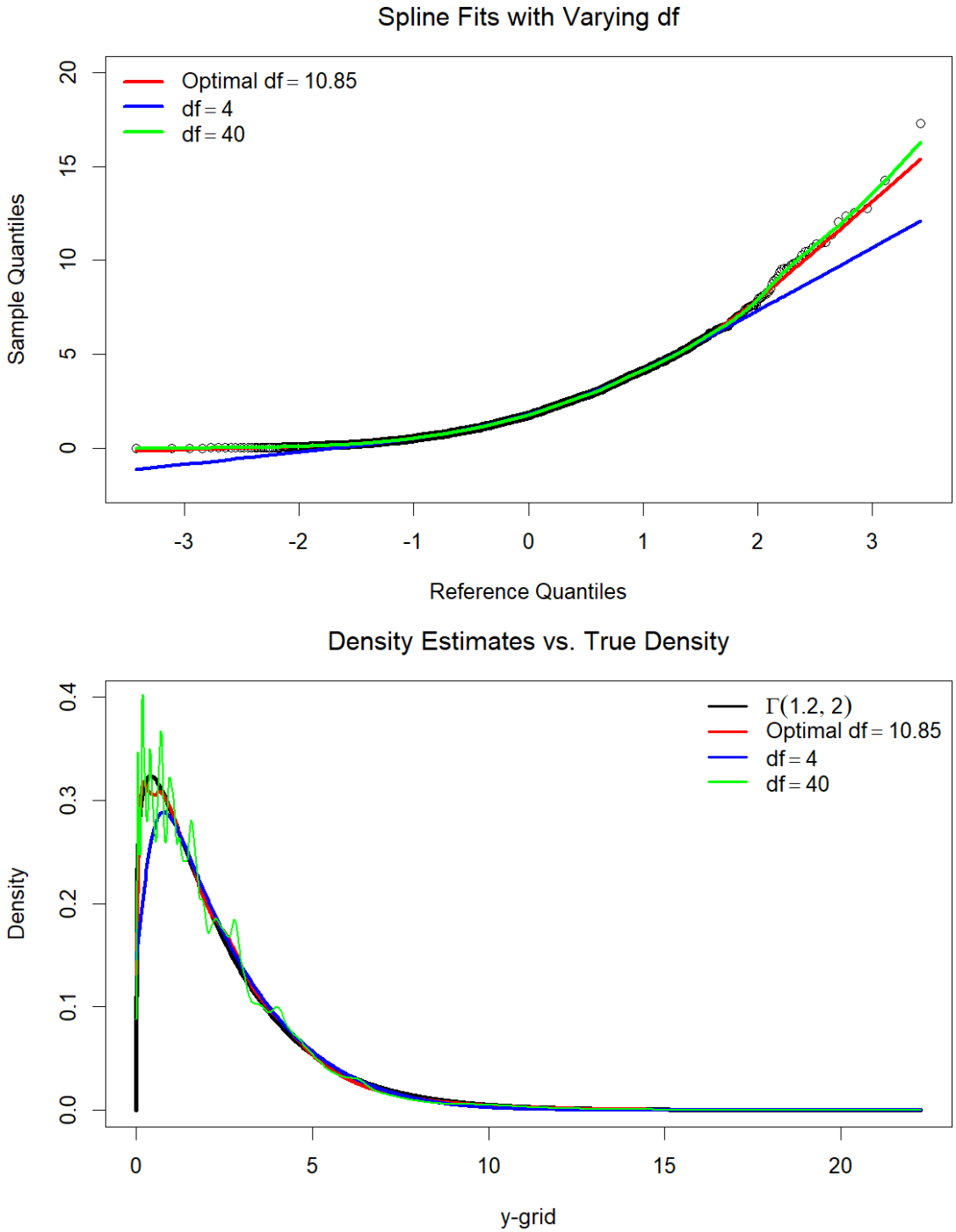

Figure 3.3: Impact of degrees of freedom on QDE performance. Top: Q-Q plots showing spline fits with varying degrees of freedom. Bottom: Corresponding density estimates compared to the true $\Gamma(1.2, 2)$ density. The optimal choice ($df = 10.85$) balances smoothness and fidelity to the data.

3.4 Iterative Weight Selection

While the procedure in Section 3.2 provides a solid foundation for density estimation, it relies on weights derived exclusively from the reference distribution. These weights remain fixed throughout the estimation process, regardless of how well the

reference distribution approximates the true underlying distribution of the data.

A more sophisticated approach would be to iteratively refine these weights, similar to weighted least squares, allowing them to adapt to the underlying structure of the data. The method proposed in this section extends the basic transformation process by introducing an iterative weight updating scheme that starts with the reference distribution but then evolves towards more optimal weights through successive refinements.

3.4.1 Motivation and Convergence

In an ideal scenario, weights should reflect the structure of the true underlying density. While the initial weights provide a reasonable starting point, they may be suboptimal if the reference distribution differs significantly from the true distribution. The iterative method proposed here allows weights to evolve from the initial reference-based values towards values that better reflect the underlying data structure.

Figure 3.4 demonstrates how the weights evolve through iterations. While these weights might appear to diverge from the true density shape, this is because the plot shows the actual weights, which are calculated as

$$\frac{n f_X(Q_X(p))^2}{p(1-p)}$$

not the density estimates themselves. The squared density term and scaling by $p(1-p)$ naturally produce a different shape from the underlying density.

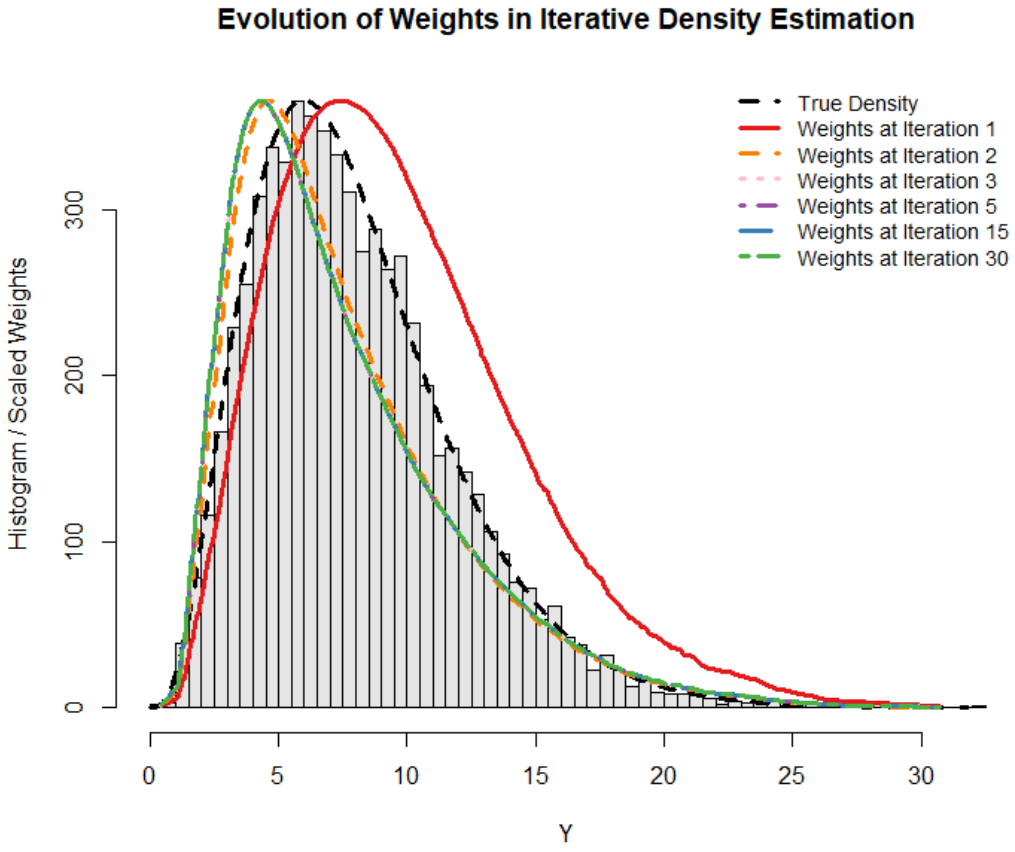

Figure 3.4: Evolution of weights through iterations for $\Gamma(4, 2)$ distribution with $n = 6400$. The weights show how the estimation process adapts to the underlying distribution structure.

Despite this visual difference, these evolving weights lead to improved density estimates, as evidenced by the decreasing ISE shown in Figure 3.5. The rapid decrease in ISE during early iterations indicates that this process efficiently improves the density estimate, with convergence typically achieved within 5-10 iterations.

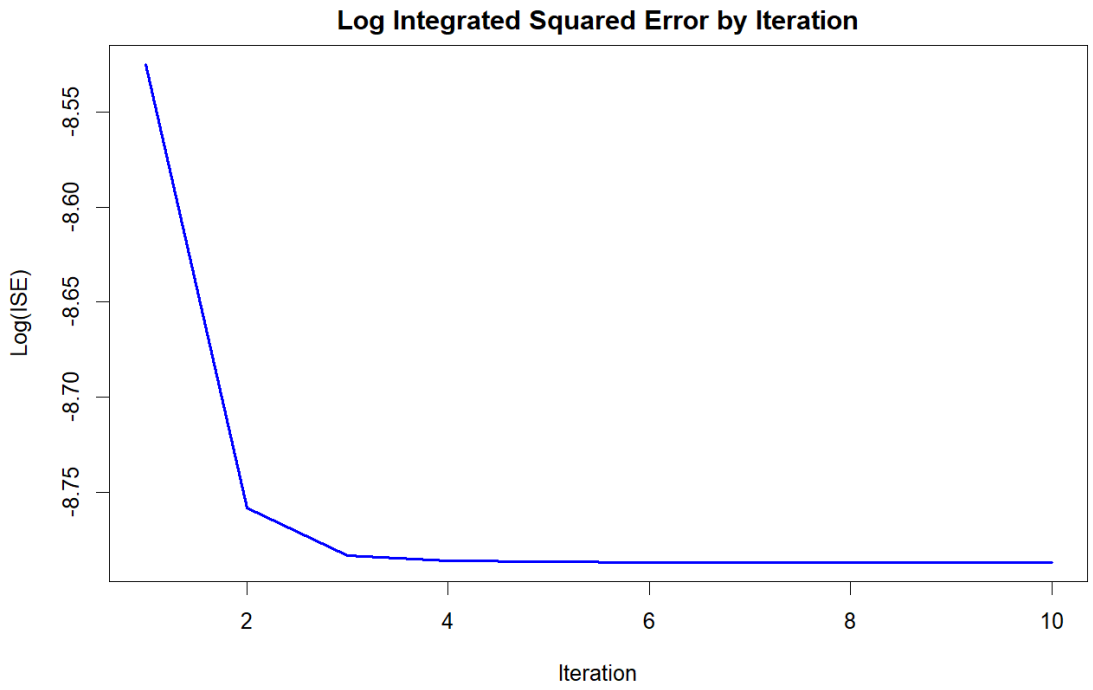

Figure 3.5: Convergence of the Integrated Squared Error (ISE) through iterations for $\Gamma(1.2, 2)$ distribution.

3.4.2 Iterative Weighting Algorithm

The procedure extends the transformation approach by introducing a weighting scheme that is iteratively updated based on successive density estimates. Let p denote a vector of probabilities (e.g., $p = (p_1, \ldots, p_n)$) at which quantiles are evaluated. The corresponding vectors of quantiles $Q_X(p)$, $Q_Y(p)$, and weights $w(p)$ are treated as functions evaluated element wise on this grid.

1. Initialise weights using the reference density:

$$w^{(0)}(p) = \frac{n f_X(Q_p)^2}{p(1-p)}$$

   where f_X is the density of the reference distribution and $Q_p = F_X^{-1}(p)$ are its quantiles.

2. For $j = 1, \ldots, k$ iterations:

   (a) Fit spline with current weights using a fixed degrees of freedom:

$$\texttt{smooth.spline}(\texttt{x} = F_X^{-1}(p), \texttt{y} = F_Y^{-1}(p), \texttt{weights} = w^{(j-1)}, \texttt{df} = df)$$

   (b) Compute transformed density $\hat{f}_Y^{(j)}(y)$ using the reference distribution by (2.1)

   (c) Update weights using:

$$w^{(j)}(p) = \frac{(n \hat{f}_Y^{(j)}(y))^2}{p(1-p)}$$

3. Use final weights $w^{(k)}$ in the optimisation process

Once the optimal weights are found, they can be directly incorporated into the quantile density estimation procedure outlined in Section 3.2.

A quick heuristic is to set df = 4 initially to get a moderately flexible fit. While different df choices might work better for different distributions or parameter settings, this provides a reasonable starting point. The choice is discussed next.

3.4.3 Choice of Smoothing Parameter

In step 2 of 3.4.2 a natural question arises being 'What does one choose the smoothing parameter as?'. While it was previously suggested to set $df = 4$ as a quick heuristic, this decision affects the final density estimate as seen in figures 3.6 and 3.7, so it deserves careful consideration. A moderate choice here is $df \in [4, 10]$. But justifications could be made to veer outside this interval depending on sample size or the underlying distribution. This comes back to the bias-variance trade-off idea that if the smoothing parameter choice is too high the weights may be capturing too much noise in the data while if it's too low it may not capture enough of the underlying structure.

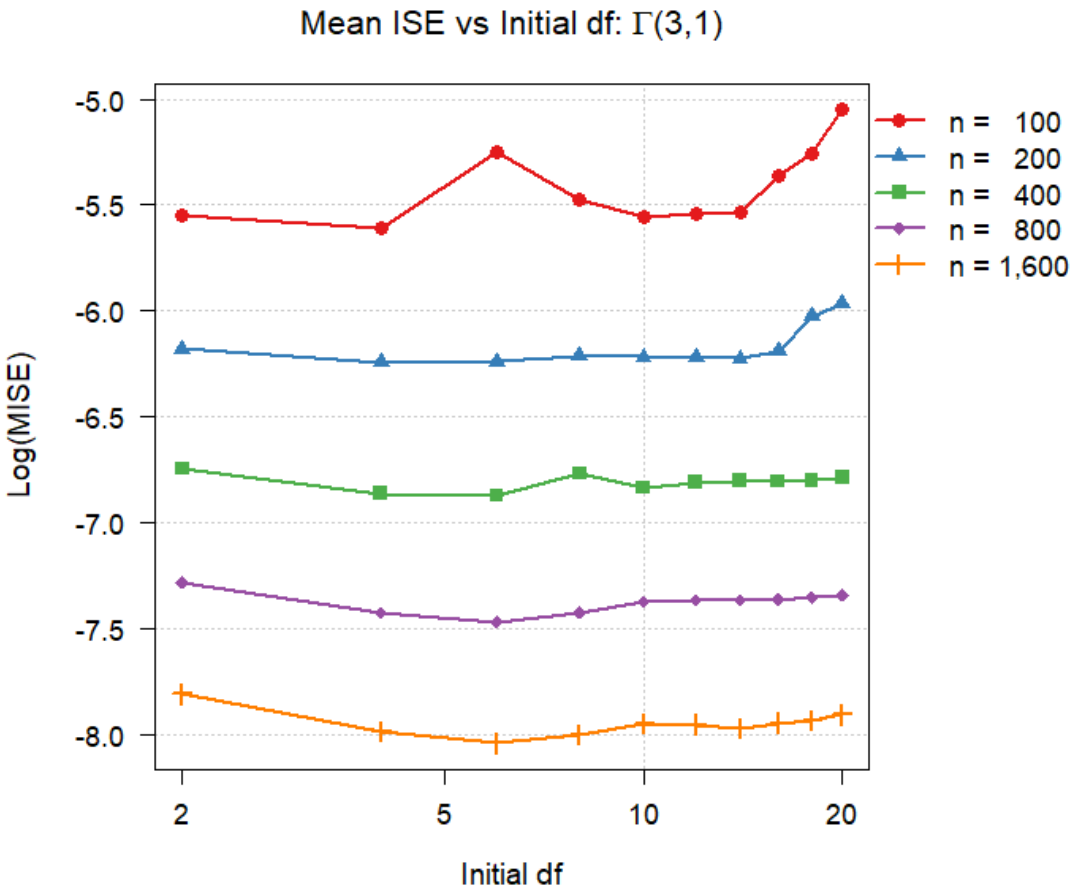

Figure 3.6: Effect of smoothing parameter choice in the Iterative weight calculation on final density estimation accuracy for $\Gamma(3,1)$ data, measuring Log(MISE) across different sample sizes (n).

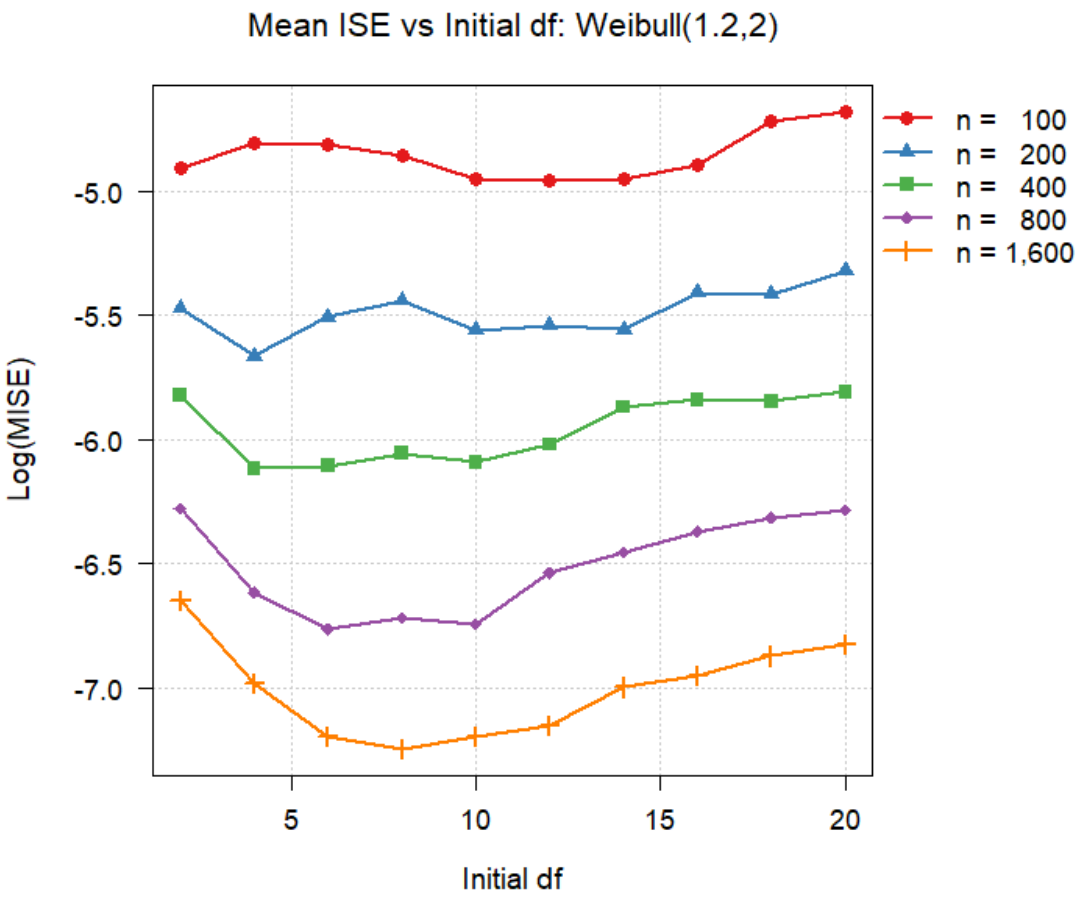

Figure 3.7: Effect of smoothing parameter choice in the Iterative weight calculation on final density estimation accuracy for $Weibull(1.2, 2)$ data, measuring Log(MISE) across different sample sizes (n).

Chapter 4

Numerical Experiments

4.1 Simulation Study Design

4.1.1 Aims

The aim of this simulation study is to evaluate the performance of the Quantile Density Estimator (QDE) for density estimation. The simulation conditions are designed to resemble typical scenarios encountered in statistical modelling. In particular, the study evaluates the performance of the QDE conditional on:

- Increasing sample size
- Varying distribution types and parameter combinations
- Different methods (Iterative vs Non-Iterative)

The QDE is compared to other established density estimation methods from the literature, namely:

- Kernel Density Estimation (KDE): the most widely used non-parametric density estimation method
- Parametric estimation via `fitdistr`(from the `MASS` library in R): a maximum likelihood approach assuming the correct distributional form

These methods represent distinct approaches to density estimation with varying flexibility and computational requirements, serving as appropriate benchmarks for the QDE.

4.1.2 Simulation Conditions

Data-generating process

In each simulation, a dataset is generated consisting of n independent and identically distributed observations from a specified probability distribution. The specific distributions and their parameter values are varied across simulation conditions (see below), leading to a total of 60,000 scenarios (5 parameter sets $\times$ 3 distributions $\times$ 8 sample sizes $\times$ 500 replications).

Sample size

The sample size used in density estimation applications varies widely in practice. The values $n \in \{100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ span typical values from small to very large samples. This range allows thorough investigation of the convergence properties of each method as the sample size increases.

Distribution types and parameters

Three families of distributions with five different parameter configurations each are considered:

1. **Gamma distribution**: Parameter sets $\{(1.2, 2), (2, 1.5), (3, 1), (4, 2), (5, 0.8)\}$, where each pair corresponds to (shape, scale)
2. **t-distribution**: Degrees of freedom $\in \{2.01, 3, 6, 12, 20\}$, covering very heavy-tailed to nearly normal distributions
3. **Weibull distribution**: Parameter sets $\{(1.2, 2), (1.5, 3), (2.0, 1.5), (2.5, 1.2), (3, 1.8)\}$, where each pair corresponds to (shape, scale)

These distributions were selected to represent a diverse range of shapes, including symmetric, asymmetric, skewed, light-tailed, and heavy-tailed distributions, which are commonly encountered in practice.

Test data

For evaluating the performance of the density estimators, the Integrated Squared Error (ISE) is calculated using a grid of 1000 equally spaced points spanning from the 0.0001 to the 0.9999 quantiles of each specific distribution. This ensures that the evaluation is focused on the relevant range of each distribution while maintaining a consistent approach across all methods.

4.1.3 Performance Metrics

Primary Performance Metric

Mean Integrated Squared Error (MISE). For each combination of distribution type, parameter set, sample size, and estimation method, the average ISE is computed across 500 independent replications:

$$\text{MISE} = \frac{1}{500} \sum_{r=1}^{500} \text{ISE}_r \tag{4.1}$$

where ISE_r is the Integrated Squared Error for the r-th replication:

$$\text{ISE}_r = \int (\hat{f}_r(x) - f(x))^2 \, dx \tag{4.2}$$

Here, $\hat{f}_r(x)$ is the estimated density from the r-th replication and $f(x)$ is the true density. Lower MISE values indicate better average estimation performance

across replications. The MISE provides a robust measure of each method's performance by averaging out the randomness in individual samples.

Since closed-form expressions for the ISE are rarely available when using non-parametric estimators on arbitrary target densities, the integral is approximated numerically using the trapezoidal rule for each replication (Stewart, 2012, Section 7.7):

$$\text{ISE}_r \approx \sum_{i=1}^{n-1} \frac{(x_{i+1} - x_i)}{2} \cdot \left((\hat{f}_r(x_i) - f(x_i))^2 + (\hat{f}_r(x_{i+1}) - f(x_{i+1}))^2 \right) \tag{4.3}$$

Secondary Performance Metrics

In addition to MISE, the following are reported:

- **Convergence rate**: Estimated by regressing log(MISE) on $\log(n)$ for each method and distribution.
- **Standard error of MISE**: To quantify the precision of the estimates.
- **Relative efficiency**: Ratios of MISE values between methods to measure comparative performance.

4.1.4 Methods

Quantile Density Estimator (QDE)

Two variants of the Quantile Density Estimator are implemented:

1. **Iterative QDE**: This approach implements the Iterative weighting algorithm described in Section 3.4.2, which progressively refines the density estimate through multiple iterations of weight updates.
2. **Non-Iterative QDE**: This simpler variant implements the direct transformation approach outlined in Section 3.2, without the Iterative refinement process. The normal distribution is used as the reference and so it is referred to throughout the remainder of the text as the Normal Weights approach.

For both variants, the smoothing parameter (degrees of freedom) is optimised to minimise the ISE using the Brent optimisation method within the `optim` function in R. The search considers a reasonable range of degrees of freedom between 1 and 30.

Benchmark methods

Kernel Density Estimation (KDE). KDE is implemented using a Gaussian kernel, with the bandwidth parameter optimised to minimise the ISE. The optimisation is performed using the Brent method with a search interval of [0.01, 4].

Parametric Estimation (`fitdistr`). Parametric density estimation is implemented using the `fitdistr` function from R's MASS package, which estimates

distribution parameters via maximum likelihood. Since the true distribution is known in this simulation, this method represents an ideal benchmark that would not be available in practical applications with unknown distributions.

4.1.5 Computational Implementation

The extensive nature of this simulation study, involving 60,000 scenarios, required efficient computational approaches. Parallel processing techniques were employed using R's parallel computing capabilities. Specifically, the simulation utilised the `parallel`, `doParallel`, and `foreach` packages to distribute the computational workload across multiple CPU cores.

The implementation employed a cluster-based approach where:

- The number of available cores was detected automatically, reserving one core for system processes
- A cluster was created and registered as the parallel backend
- Necessary functions and libraries were exported to each worker node
- Simulations were distributed across cores using a parallel `foreach` loop

This parallel computing approach improved computational efficiency by allowing simultaneous processing of multiple simulation scenarios, reducing the overall execution time compared to sequential processing.

4.2 Results

In this section, results are presented to answer the questions presented in 4.1. The focus is on empirical consistency of the estimators, relative efficiency, relative performance of the two Quantile Density Estimator(QDE) methods vs the Kernel Density Estimator(KDE), and whether optimal weights provide an improvement to the estimator. Results of the `fitdistr`(FD) method are also included as a benchmark. Detailed numerical evidence supporting the findings is provided in the Appendix, including complete tables of MISE values, standard deviations, confidence intervals, efficiency ratios, and convergence analyses for all distribution types and sample sizes considered in this study. In the main text, key findings and illustrative examples from the analysis is presented.

4.2.1 Empirical Consistency and Convergence Rates

The empirical convergence rates were quantified by fitting regression models of the form

$$\log(\text{MISE}) = \alpha + \beta \log(n)$$

where β represents the empirical convergence rate, α represents the value of $\log(MISE)$ when $n = 1$, and log denotes the natural logarithm. A more negative $\hat{\beta}$ indicates faster decay of MISE with sample size. For spline-based methods with twice differentiable functions, one might expect a theoretical rate of $-4/5$. All estimators demonstrate decreasing MISE as sample size increases, providing

empirical evidence of consistency. While this does not constitute a formal proof, it suggests that the estimators converge toward the true density with increasing data, consistent with the notion of asymptotic consistency. Confirmation of this can be found in the results section of the Appendix. The parametric FD method consistently demonstrates the fastest convergence across all distribution types, as expected for correctly specified parametric methods. For the non-parametric methods, it varies by distribution type. A complete tabulation of convergence rates, including confidence intervals and theoretical targets for all t , $Weibull$, and $Gamma$ distributions are provided in the Appendix.

t-Distribution

The t-distribution presents an informative case study in estimator performance, particularly in the context of heavy-tailed data. As shown in Table A.4 the Normal Weights method demonstrates the strongest convergence properties, followed by the Iterative approach and then KDE. This performance hierarchy between the Iterative method and Normal Weights method makes sense, given the structural similarities between the t and normal distributions. While the Normal Weights method excels, the Iterative weights procedure, though theoretically promising, exhibits a subtle but consistent bias that slightly diminishes its performance. Although not severe, this bias is a relevant consideration in practical applications. Both QDE variants achieve convergence rates that closely approach the theoretical optimal rate of $O(n^{-4/5})$. KDE, however, shows noticeably poorer performance for heavy-tailed distributions, with convergence rates consistently below the theoretical benchmark. This performance gap becomes more pronounced as the degrees of freedom increase, as illustrated in Figures 4.1 and 4.2.

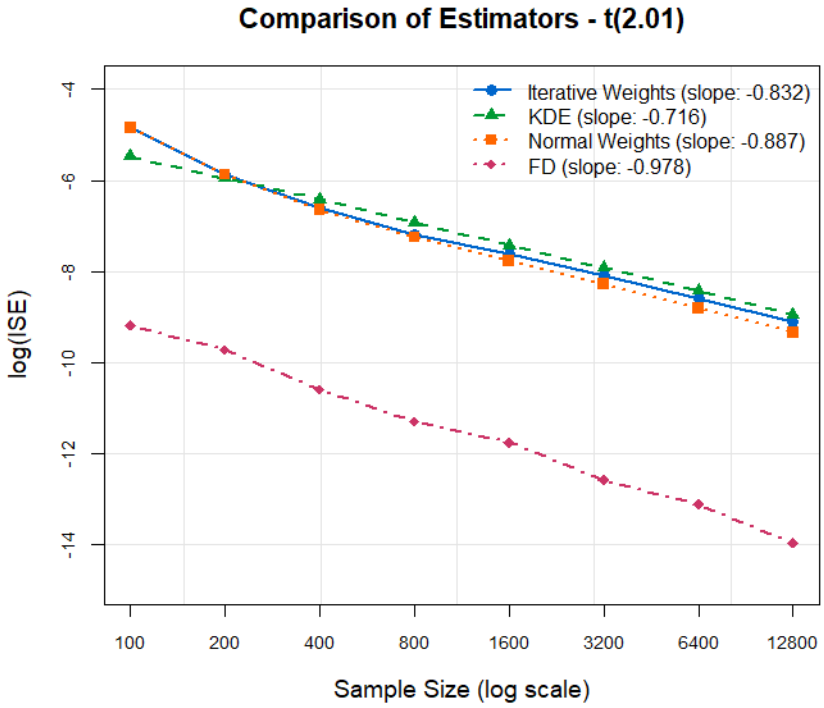

Figure 4.1: Comparison of estimator convergence rates for $t(2.01)$ distribution. The parametric FD method shows superior performance, while both QDE variants outperform KDE.

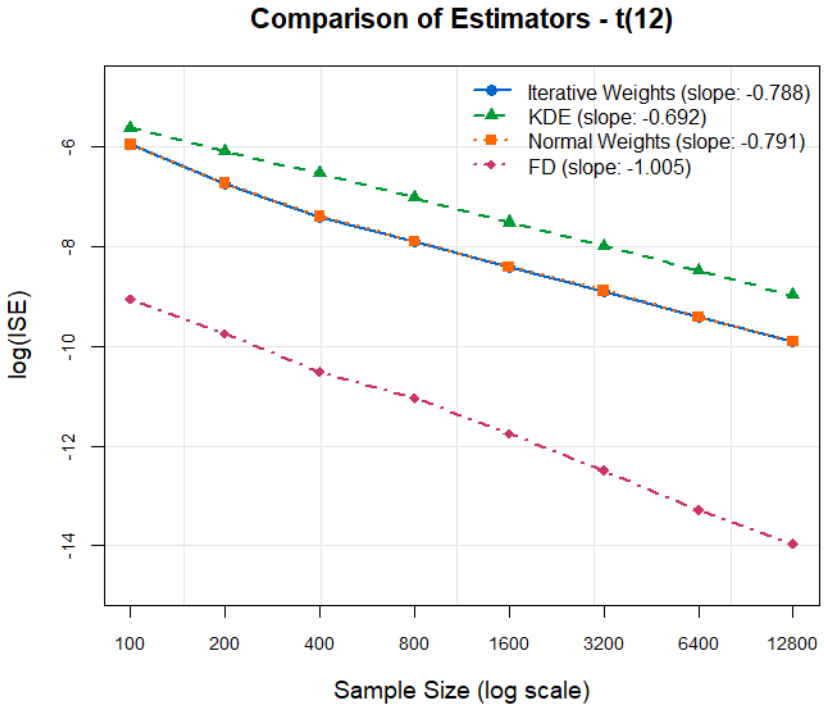

Figure 4.2: Comparison of estimator convergence rates for $t(12)$ distribution, showing maintained performance patterns at higher degrees of freedom.

The Normal Weights method maintains stable performance across all degrees of freedom, while the Iterative approach shows slightly greater variability. The KDE is particularly affected in the presence of heavy tails, where its convergence rate lags substantially behind both QDE methods. The parametric FD method demonstrates consistently excellent performance across all configurations, achieving convergence rates very close to the theoretical $O(n^{-1})$. This robust performance highlights the advantage of correctly specified parametric models.

Gamma Distribution

The Gamma distribution provides particularly insightful contrasts in estimator behaviour under varying degrees of skewness. The Iterative weighting procedure demonstrates the strongest overall performance, maintaining consistent convergence across all configurations considered. However, a clear relationship emerges between distribution shape and estimator effectiveness.

All methods exhibit diminished consistency when applied to highly skewed distributions (See Table A.8), with this effect most apparent in the case of $\Gamma(1.2, 2)$ compared to Gamma distributions with larger shape parameters. This sensitivity to skewness is illustrated in Figures 4.3 and 4.4, which contrast convergence behaviour in a highly skewed versus a more symmetric setting.

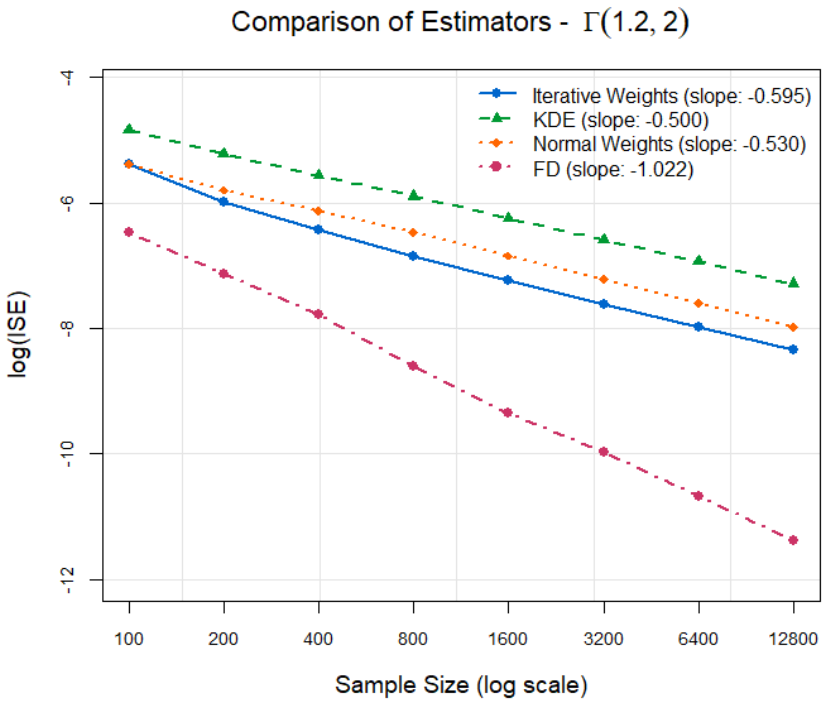

Figure 4.3: Comparison of estimator convergence rates for the $\Gamma(1.2, 2)$ distribution, demonstrating the impact of high skewness on estimator performance.

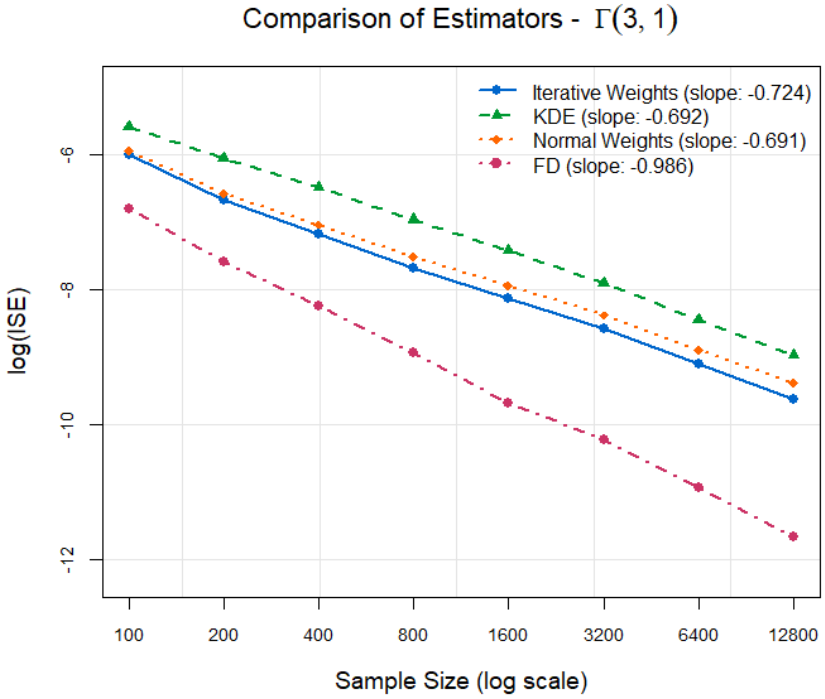

Figure 4.4: Comparison of estimator convergence rates for the $\Gamma(3, 1)$ distribution, showing improved performance with reduced skewness.

Empirical convergence rates reveal that all non-parametric methods fall short of the theoretical optimal rate of $O(n^{-4/5})$ in the Gamma setting. This shortfall is most pronounced in the highly skewed $\Gamma(1.2, 2)$ case, where even the best-performing non-parametric method, the Iterative QDE achieves considerably lower empirical rates than expected. As the shape parameter increases, estimator performance improves across the board, with the Iterative QDE retaining a consistent lead over KDE and Normal Weights approaches.

The parametric FD method again demonstrates strong and stable performance across all configurations, achieving convergence rates close to the theoretical $O(n^{-1})$ rate regardless of skewness.

Weibull Distribution

The Weibull distribution offers another valuable perspective on estimator behaviour in modelling skewed densities. All estimators demonstrate a positively correlated relationship between skew and consistency. The empirical convergence rates decrease as skewness decreases(i.e. as the shape parameter increases). For highly skewed configurations (shape parameter ≤ 2), the Iterative QDE retains a clear advantage over other non-parametric methods, even though none of the estimators achieve the theoretical convergence rate. This performance hierarchy is particularly evident in the Weibull(1.5,3) case, as illustrated in Figures 4.5 and 4.6.

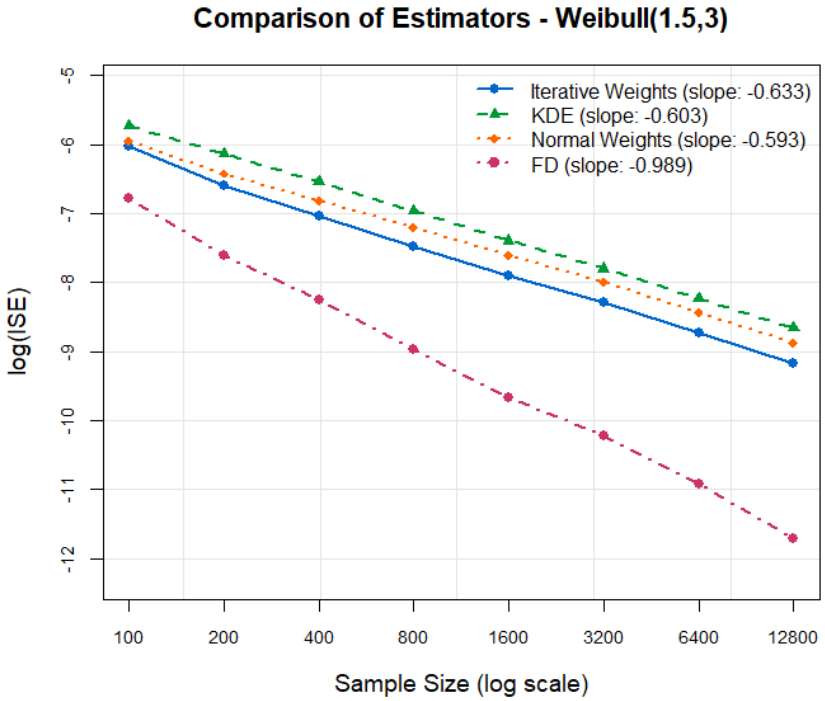

Figure 4.5: Comparison of estimator convergence rates for Weibull(1.5,3) distribution, demonstrating relative performance under moderate skewness.

As the shape parameter increases beyond 2, the empirical convergence rates of KDE and Iterative QDE become more similar in their point estimates, but the Iterative method shows significantly higher variability in its performance. The confidence intervals for the Iterative method are consistently wider than those of KDE, suggesting that while the average performance may be comparable, the Iterative method's performance is less stable across different samples. This increased variability in the Iterative method's performance becomes more pronounced as the distribution becomes more symmetric.

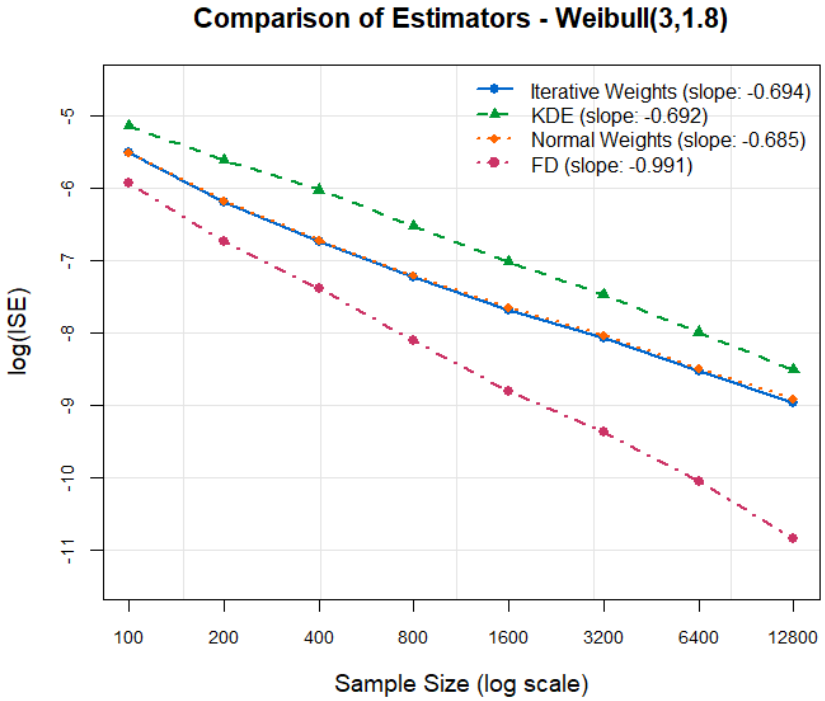

Figure 4.6: Comparison of estimator convergence rates for Weibull(3,1.8) distribution, showing convergence behaviour with reduced skewness.

The parametric FD method again demonstrates exemplary stability, achieving convergence rates closely aligned with its theoretical $O(n^{-1})$ target across all configurations. This consistency further emphasises the advantages of correctly specified parametric approaches when available.

A complete tabulation of convergence rates, including confidence intervals and theoretical targets for all Weibull distributions considered, is provided in Table A.12 in the Appendix.

Gap Between Theoretical and Empirical Rates: A notable finding across both Gamma and Weibull distributions is the consistent gap between theoretical and empirical convergence rates for all non-parametric methods. While the t-distribution results showed convergence rates close to the theoretical $O(n^{-4/5})$, both Gamma and Weibull distributions show substantially slower empirical convergence.

This discrepancy may be attributed to the boundary bias present at the origin, where these distributions exhibit rapid density changes that are particularly difficult to estimate accurately for shape parameters below 2. These challenges appear to affect all non-parametric methods, though to varying degrees depending on the specific distributional parameters.

The parametric FD approach achieves rates very close to the theoretical $O(n^{-1})$ rate across all distributions (ranging from -0.97 to -1.04), confirming the well-known advantage of correctly specified parametric methods.

4.2.2 Analysis of ISE Results

When comparing estimator performance across these distribution families, several patterns emerge that highlight methodological strengths and limitations:

t-distributions

The Integrated Squared Error (ISE) results across various t-distributions, summarised in Table A.1, reveal consistent patterns in both estimator accuracy and variability. Across all settings, the Iterative and Normal Weights (NW) estimators produce similar Mean ISE (MISE) values, while Kernel Density Estimation (KDE) generally exhibits higher error, particularly at larger sample sizes.

In the heavy-tailed case of $df = 2.01$, KDE performs competitively at small sample sizes but shows slower improvement as n increases. The Iterative and NW estimators demonstrate more stable reductions in MISE across the board. As the degrees of freedom increase to $df = 3$ and $df = 12$, all methods improve, but the spline-based approaches consistently outperform KDE. At $df = 12$, the Iterative and NW methods are nearly indistinguishable in performance, with KDE maintaining a persistent gap in MISE.

The corresponding standard deviations, also reported in Table A.1, mirror these patterns. Variability decreases with increasing sample size for all estimators, but remains notably higher for KDE, especially in the heavy-tailed setting. The Iterative and NW estimators show similar levels of dispersion throughout, suggesting comparable stability.

Confidence intervals for each method are presented in Table A.2. These intervals narrow with increasing n, as expected, and provide further evidence of the reliability of the spline-based approaches. The Iterative and NW methods exhibit nearly identical intervals across all scenarios, reflecting both low bias and consistent performance. KDE intervals are consistently wider, reinforcing its tendency toward greater variability.

While the parametric `fitdistr` method achieves the lowest MISE values overall, this result is expected under the correct model specification. The primary interest here lies in understanding the comparative behaviour of the non-parametric estimators under varying tail conditions.

Gamma Distributions

The ISE results for Gamma distributions, summarised in Table A.5, demonstrate how estimator performance varies with shape parameter. For highly skewed cases ($shape = 1.2$), the Normal Weights (NW) estimator initially outperforms both the Iterative and KDE approaches at $n = 100$, though this advantage diminishes as sample size increases. The Iterative method shows superior performance for $n \geq 400$, particularly in capturing the heavy right tail.

For moderate shape parameters ($shape = 2, 3$), both spline-based approaches demonstrate improved accuracy over KDE. The NW method shows particularly strong performance at smaller sample sizes, as evidenced by lower MISE values and narrower confidence intervals. As $shape$ increases ($shape \geq 4$), both QDE methods consistently outperform KDE across all sample sizes, with the gap widening at larger n.

The corresponding density difference plot (Figure 4.8) shows that while the Iterative method exhibits a sharper deviation near the mode, it stabilises quickly across the domain. KDE, on the other hand, demonstrates more oscillatory behaviour across the domain, particularly beyond the peak.

The standard deviations reported in Table A.5 show that while KDE exhibits lower variability in the highly skewed case where $shape = 1.2$ and $n = 100$, this advantage disappears as the both $shape$ and n increases.

As with the t-distributions, the parametric `fitdistr` method achieves the lowest MISE values. This reflects the benefit of correct model specification rather than general estimation capability.

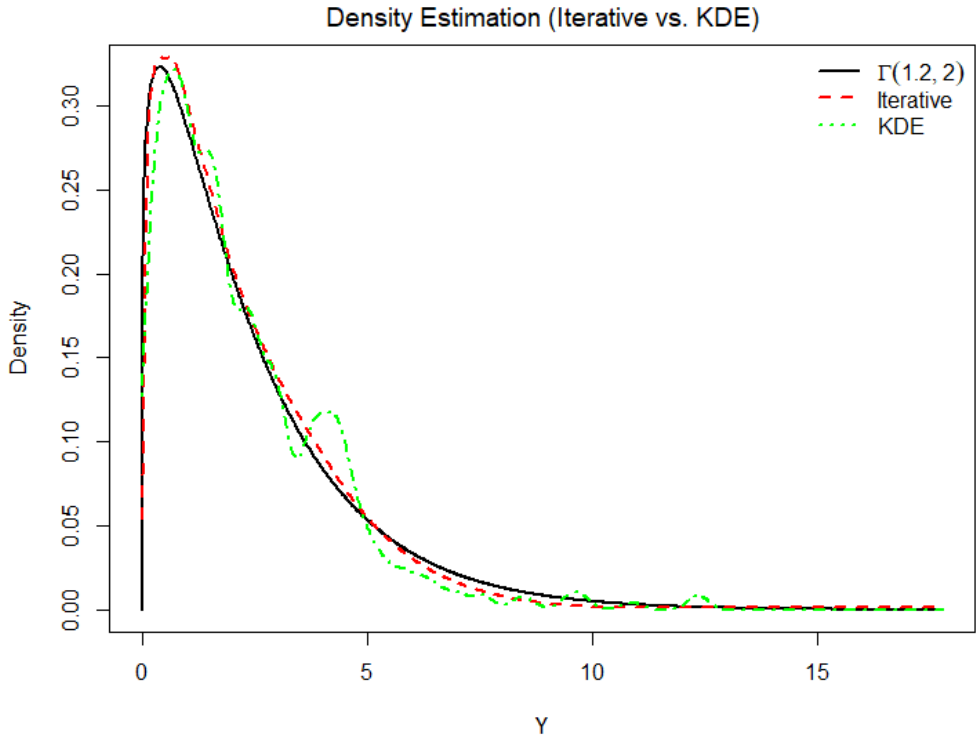

Figure 4.7: Density estimation of $\Gamma(shape = 1.2,\ scale = 2)$ at $n = 400$. The Iterative QDE (red dashed) better captures the true density (black solid) compared to KDE (green dotted), particularly in the tail.

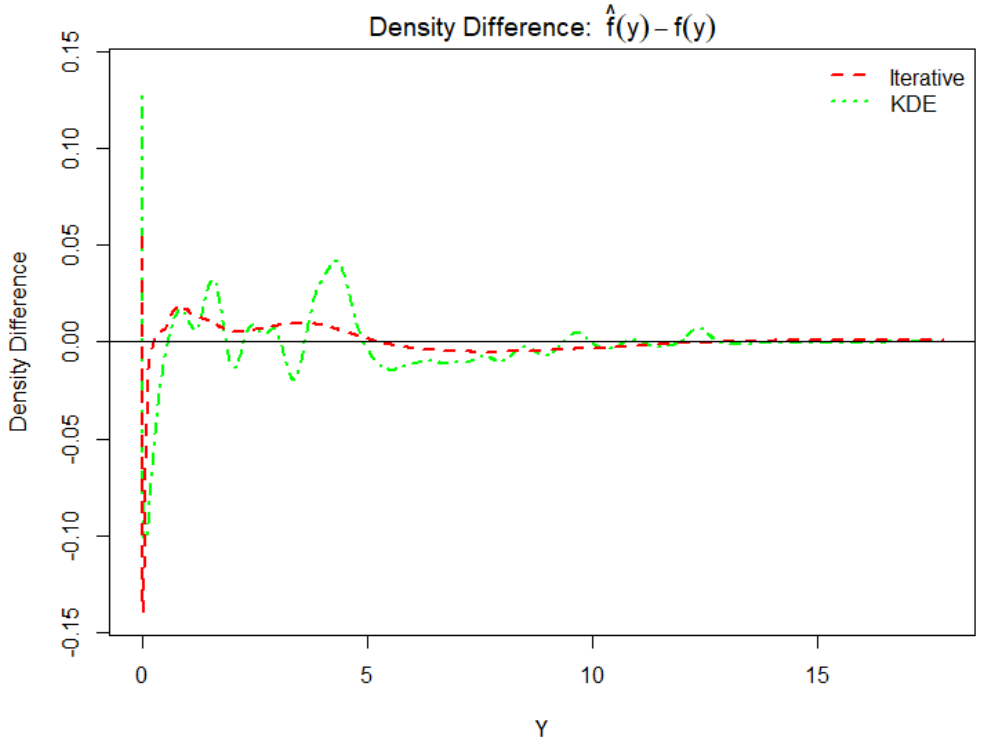

Figure 4.8: Pointwise density difference plot for $\Gamma(shape = 1.2,\ scale = 2)$ at $n = 400$. The Iterative QDE (red dashed) shows smaller and more consistent deviations from the true density compared to KDE (green dotted).

Weibull Distributions

Integrated Squared Error (ISE) results for various Weibull distributions are summarised in Appendix Table A.9 and A.10. For the heavily skewed Weibull$(1.2, 2)$ distribution, the Normal Weights (NW) estimator achieves the lowest MISE at small sample sizes. However, its advantage diminishes as n increases: the Iterative estimator surpasses both estimators by $n = 200$, maintaining a clear lead thereafter.

For moderate shapes (shape $= 2$), the spline-based methods show comparable performance, with both outperforming KDE across all sample sizes. As the shape parameter increases beyond 2.5, the Iterative and NW estimators remain nearly indistinguishable in performance, and both continue to outperform KDE. This pattern becomes especially clear at larger sample sizes, where the KDE method shows persistently higher MISE.

Standard deviations reported in the same appendix table show that the Iterative estimator exhibits greater variability at small n for highly skewed distributions. However, this variability decreases noticeably with sample size. For shape ≥ 2.5, all QDE methods display similar levels of stability.

As observed in previous settings, the parametric `fitdistr` method achieves the lowest MISE throughout. While this is expected under correct model specification, the primary focus here remains on the comparative behaviour of the non-parametric estimators under varying distributional forms.

4.2.3 Relative Efficiency Analysis

The relative efficiency (RE) between two estimators A and B is defined as

$$RE_{A/B} = \frac{MISE_A}{MISE_B},$$

where $MISE$ denotes the mean integrated squared error. Following Lehmann and Casella, 2006, this ratio indicates the asymptotic relative number of observations required by estimator A to achieve the same level of precision as estimator B. For example, a relative efficiency of 2 suggests that estimator A requires twice as many observations as estimator B to achieve the same precision.

Three key efficiency ratios are examined to compare estimator performance under varying distributional conditions:

- $RE_{\text{KDE/ITER}}$: Kernel Density Estimator (KDE) relative to the Iterative method
- $RE_{\text{KDE/NW}}$: KDE relative to the normal-weighted (NW) method
- $RE_{\text{NW/ITER}}$: NW relative to the Iterative method

These ratios provide a framework for assessing the comparative behaviour of the estimators under different levels of skewness and sample sizes. It should be noted that all results can be found in the Appendix.

t-Distribution Efficiency

Relative efficiency trends for t-distributions reveal notable dependence on the degrees of freedom, particularly in heavy-tailed settings. As shown in Appendix Table A.3 and Figure 4.9, for the heavy-tailed case $t(2.01)$, the QDE methods initially underperform compared to KDE. At $n = 100$, both $RE_{\text{KDE/ITER}}$ and $RE_{\text{KDE/NW}}$ are below 0.6, indicating KDE's advantage at small sample sizes. However, both ratios improve significantly with n, surpassing 1.2 by $n = 400$ and continuing to increase for $RE_{\text{KDE/NW}}$, as shown in Figure 4.9. The $RE_{\text{NW/ITER}}$ ratio decreases steadily with sample size, falling from approximately 1.01 to 0.80, suggesting that the Normal Weights method becomes more efficient in large sample regimes under heavy tails.

Moving to the t(3) case, at n=100, both efficiency ratios remain below 1 ($RE_{KDE/ITER} = 0.7638$, $RE_{KDE/NW} = 0.7611$), but they increase steadily with sample size, eventually surpassing 1.3 for large n. ($RE_{KDE/ITER}$ actually peaks at $n = 800(1.5677)$ while $RE_{KDE/NW}$ peaks at $n = 1600(1.6169)$ before both steadily decreasing.

For higher degrees of freedom $df \in \{6, 12, 20\}$, smoother and more consistent efficiency trends are observed. Both $RE_{KDE/ITER}$ and $RE_{KDE/NW}$ increase monotonically with n and df. In these moderately heavy to near-Gaussian settings, the QDE methods exhibit consistent efficiency advantages over KDE. Figure 4.10 illustrates this trend, with KDE becoming increasingly inefficient relative to the NW method as the distribution approaches normality.

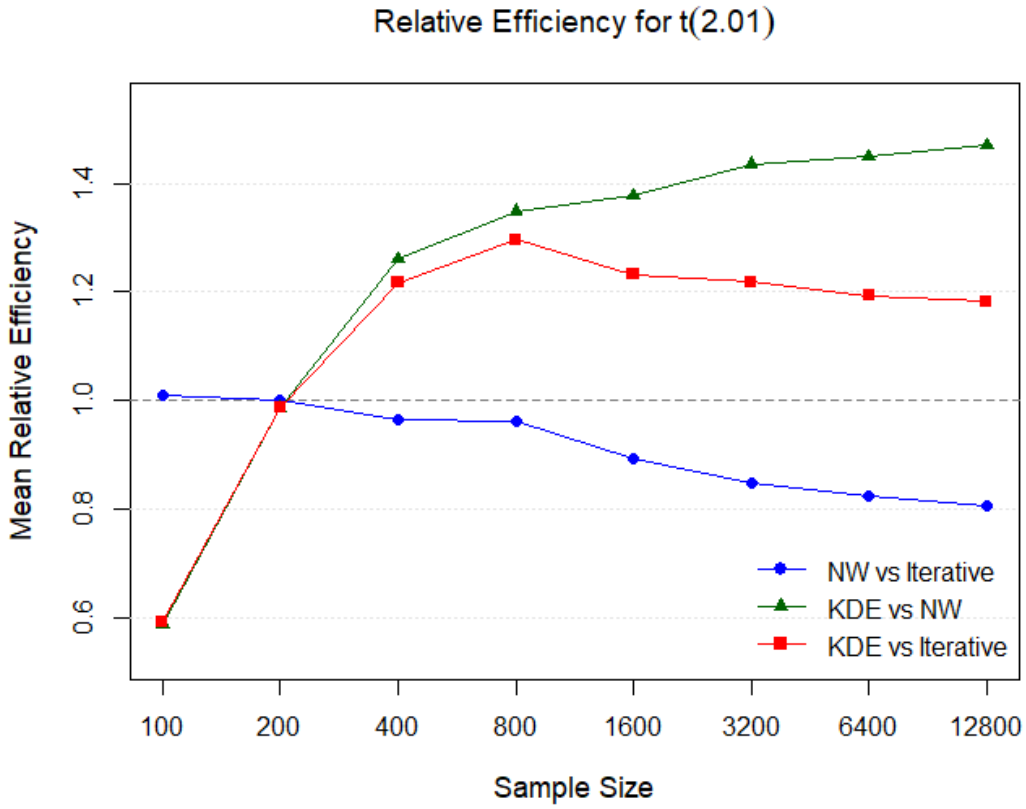

Figure 4.9: Relative efficiency comparisons for the $t(2.01)$ distribution across sample sizes ranging from $n = 100$ to $n = 12800$. The horizontal dashed line at $y = 1.0$ indicates equal efficiency between methods.

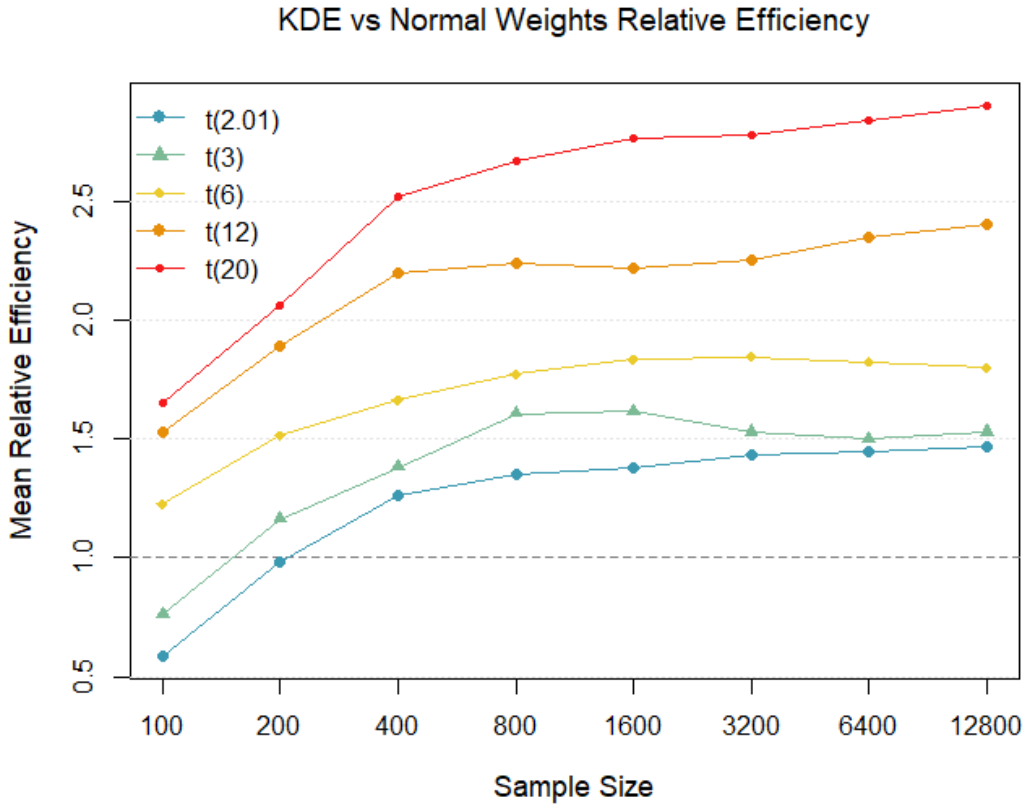

Figure 4.10: Relative efficiency comparisons between KDE and Normal Weights methods across different degrees of freedom in the t-distribution for sample sizes ranging from $n = 100$ to $n = 12800$.

The relative efficiency between the NW and Iterative methods remains close to unity for $df \geq 6$, with $RE_{\text{NW/ITER}}$ fluctuating around 1. For large n, the Normal Weights method retains a slight advantage in heavier-tailed settings that widens as n gets larger, while both methods perform nearly identically in more Gaussian-like cases.

Three key patterns emerge from the t-distribution analysis. First, KDE performs relatively well at very small sample sizes under heavy tails but loses ground as n increases. Second, both QDE methods exhibit increasing efficiency with lighter tails and larger sample sizes, particularly the NW method over the KDE. Third, the choice between NW and Iterative methods becomes less consequential for $df \geq 6$, with both methods exhibiting comparable performance.

Gamma Distribution Efficiency

Efficiency ratios were evaluated across a range of Gamma distributions, capturing varying degrees of skewness and shape parameters. As shown in Appendix Table A.7, for the highly skewed case $\Gamma(1.2,\ 2)$, both $RE_{\text{KDE/ITER}}$ and $RE_{\text{KDE/NW}}$ increase with sample size, with the former exhibiting a more pronounced growth. Figure 4.11 illustrates this growing relative efficiency advantage of the QDE methods over KDE.

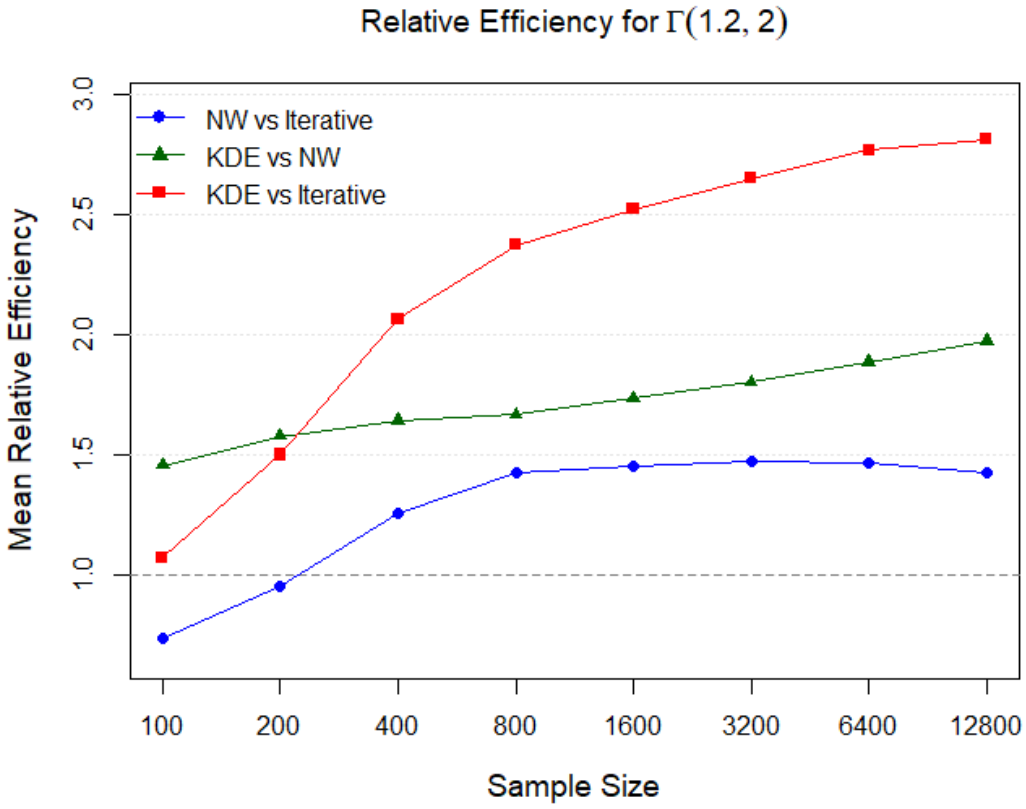

Figure 4.11: Relative efficiency comparisons for the $\Gamma(1.2, 2)$ distribution across sample sizes ranging from $n = 100$ to $n = 12800$. The horizontal dashed line at $y = 1.0$ indicates equal efficiency between methods.

For moderate shape parameters ($2 \leq$ shape ≤ 3), efficiency ratios increase more gradually. The $RE_{\text{KDE/ITER}}$ ratio stabilises as sample size increases, while $RE_{\text{KDE/NW}}$ remains relatively consistent across the range. The relationship between the two QDE methods, captured by $RE_{\text{NW/ITER}}$, follows a consistent trajectory across all Gamma configurations: it starts below unity at $n = 100$ for all bar $\Gamma(3, 1)$, generally increases with n, and eventually stabilises. As shown in Figure 4.12, the efficiency ratio decreases with increasing shape parameter, indicating that the advantage of the Iterative method becomes less pronounced in settings with lower skewness.

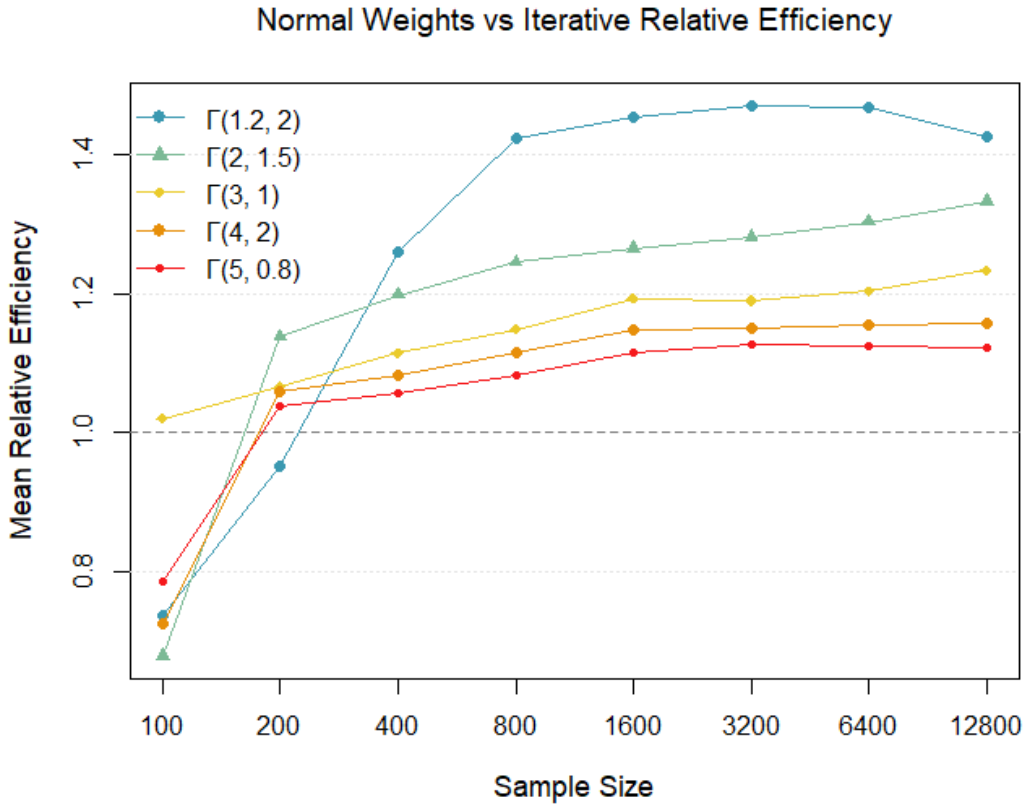

Figure 4.12: Relative efficiency comparisons of the Normal Weights relative to Iterative method across different shape parameters in the Gamma distribution. The plot illustrates how efficiency relationships evolve with increasing sample size.

Three key patterns emerge from this analysis. First, relative efficiency improves with increasing sample size. Second, the greatest relative efficiency gains over KDE are observed in more skewed distributions, particularly $\Gamma(1.2,\ 2)$. Third, estimator performance varies more significantly at small sample sizes, highlighting the importance of method selection in small-sample scenarios, a theme echoed across the other distribution families.

Weibull Distribution Efficiency

Efficiency behaviour for the Weibull distribution is notably influenced by the shape parameter, with more pronounced effects observed under skewed conditions. As shown in Appendix Table A.11, for the skewed Weibull(1.2, 2) case, both $RE_{\text{KDE/ITER}}$ and $RE_{\text{KDE/NW}}$ increase steadily with sample size, reaching 2.57 and 1.84, respectively, at $n = 12800$. This indicates a growing efficiency advantage of the QDE methods over KDE in highly skewed settings. Figure 4.13 visually reinforces this trend, showing how the relative efficiencies continue to rise with sample size in the most skewed scenario.

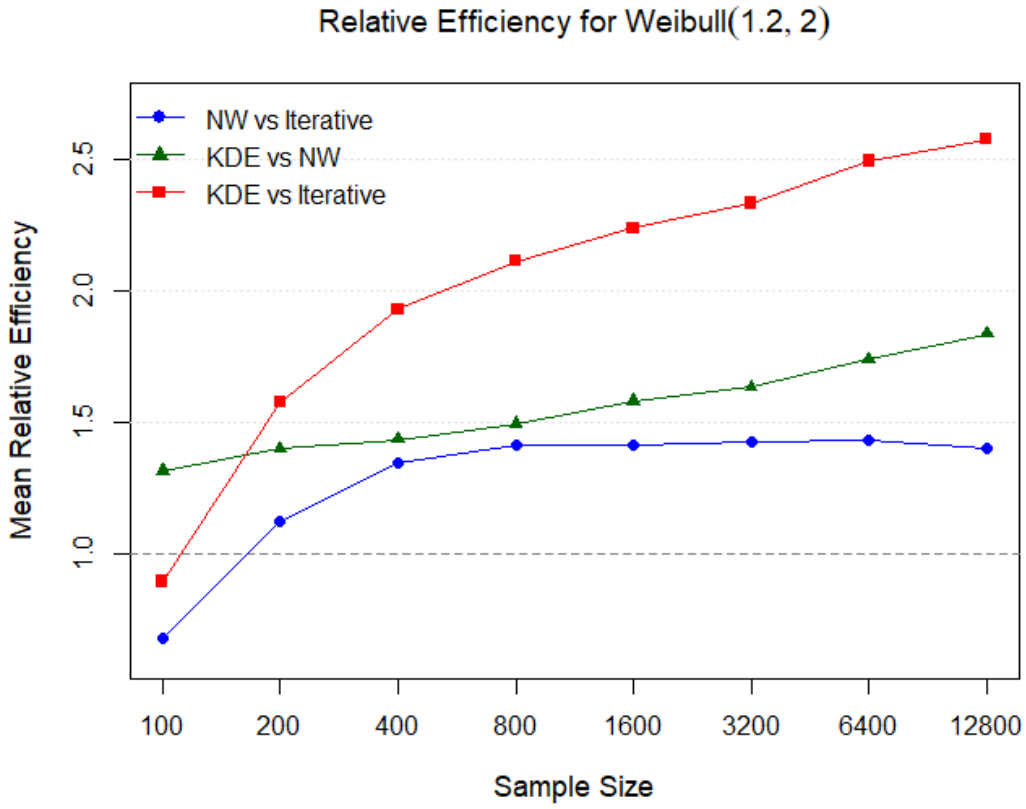

Figure 4.13: Relative efficiency comparisons for the Weibull(1.2, 2) distribution across sample sizes ranging from $n = 100$ to $n = 12800$. The horizontal dashed line at $y = 1.0$ indicates equal efficiency between methods.

As shape increases and the distribution becomes more symmetric, in general KDE converges closer to the QDE estimators than in heavily skewed scenarios. For example, while $RE_{\text{KDE/ITER}}$ is around 2.57 in the $shape = 1.2$ case at $n = 12800$ (indicating a large efficiency gap), it falls into the 1.3–1.6 range when shape ≥ 2.0. Although the relationship isn't perfectly monotonic across every shape-scale combination, the broader trend remains clear: the QDE and KDE are more similar in near-symmetric regimes than under heavy skew.

$RE_{\text{NW/ITER}}$ follows a familiar pattern across all Weibull configurations: it starts below 1 for small n, increases steadily with sample size, and typically stabilises in the 1.0–1.4 range (see Figure 4.14). The degree of improvement diminishes as the distribution becomes more symmetric, further reinforcing the notion that the QDE methods are most effective under skew. As skew decreases (shape parameter increases), the rate of change of $RE_{\text{NW/ITER}}$ diminishes with sample size. For example, in the most skewed case (shape=1.2), the NW/ITER ratio increases by 0.72 from $n = 100$ to $n = 12800$, while for the least skewed case (shape=3.0), this increase is only 0.05.

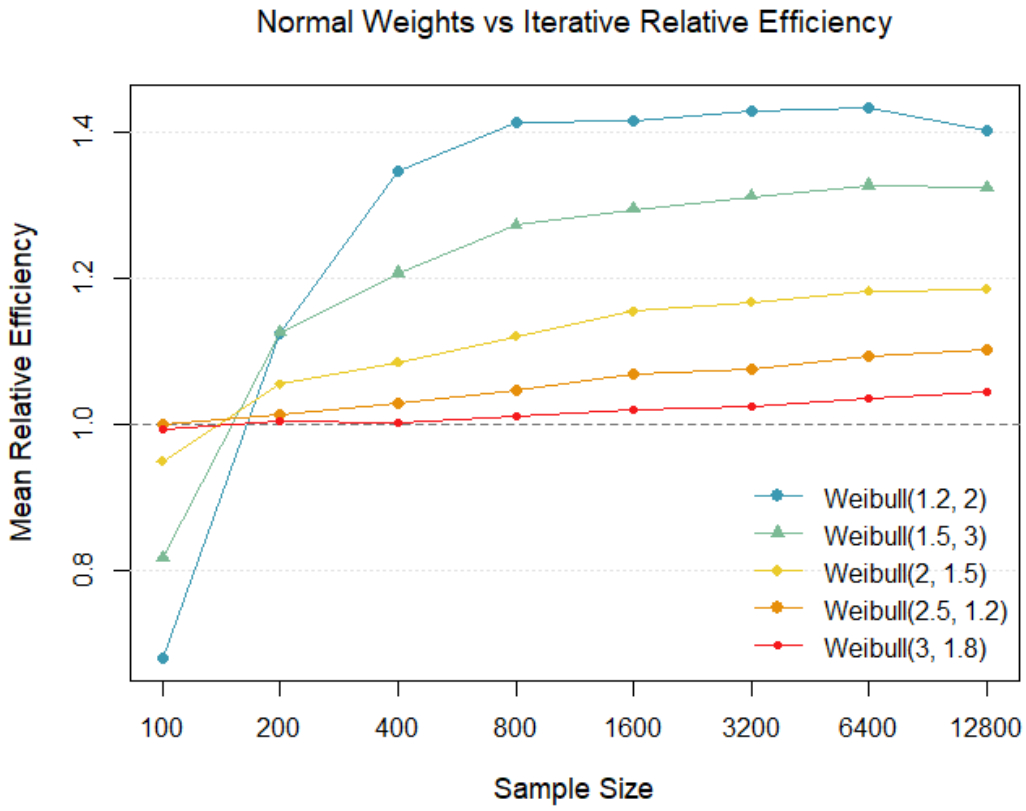

Figure 4.14: Relative efficiency comparisons between Normal Weights and Iterative methods across different shape parameters in the Weibull distribution for sample sizes ranging from $n = 100$ to $n = 12800$.

Three key patterns emerge from the Weibull analysis. First for highly skewed distributions, the relative efficiency ratios increase with sample size. Second, as shape increases, the relative efficiency difference between Normal Weights and Iterative methods becomes less sensitive to sample size, with the change in $RE_{\text{NW/ITER}}$ from $n = 100$ to $n = 12800$ decreasing from 0.72 for $shape = 1.2$ to just 0.05 for $shape = 3.0$. Third, for the highly skewed Weibull case ($shape = 1.2$), the NW method initially outperforms the Iterative method at very small n as indicated by $RE_{\text{NW/ITER}} < 1$ but is ultimately overtaken as n grows, with this ratio exceeding 1 beyond roughly $n = 200$. At $n = 12800$, the Iterative estimator is about 40% more efficient than NW and more than twice as efficient as Kernel Density Estimation.

Summary of Efficiency Analysis

These findings across all three distribution families suggest that the QDE methods offer particular advantages in handling skewed and heavy-tailed distributions, with efficiency gains typically increasing with sample size. The choice between NW and Iterative methods becomes less critical for more symmetric distributions or larger sample sizes.

4.2.4 Iterative or Normal Weights?

Simulation results reveal consistent patterns in the relative performance of the Iterative and Normal Weights (NW) estimators across sample sizes and distributional families:

Sample Size Dependence The NW estimator performs more reliably at small sample sizes ($n \leq 200$) for symmetric distributions with moderate to light tails, showing both lower mean integrated squared error (MISE) and reduced variance, particularly for the $\Gamma(2,\ 1.5)$. However, the Iterative method shows advantages even at small n for heavily skewed or very heavy-tailed distributions. The Iterative method improves with increasing sample size ($n \geq 400$), often surpassing NW in efficiency under conditions of skewness or heavy tails. This transition indicates that while NW provides greater stability in low-data regimes, the Iterative method is better suited to large-sample scenarios where its regularisation benefits are fully realised.

Distribution-Specific Characteristics For t-distributions with heavy tails ($df \leq 3$), the Iterative estimator maintains a slight edge in stability and performance at small to moderate n, though this advantage narrows as df increases and the distribution becomes more Gaussian. The NW estimator actually beats the Iterative estimator at $n = 100$ when $df \in \{6, 12, 20\}$ all while maintaining lower variability.

In skewed settings, particularly Gamma and Weibull distributions with shape parameters < 2 and $n \geq 400$ the Iterative method better accommodates boundary behaviour and captures sharp density gradients. While the NW estimator shows initial stability, its relative efficiency improves less substantially with n in these cases, indicating limited long-run gains compared to the Iterative approach.

Consistency Considerations The convergence behaviour of these estimators varies by distribution type. For t-distributions, particularly those with heavy tails, the NW estimator achieves better convergence rates than the Iterative method. However, this pattern reverses for skewed distributions, where the Iterative method demonstrates superior convergence rates for both Gamma and Weibull distributions, particularly under high skew. Notably, the NW estimator exhibits narrower confidence intervals for its convergence rates in skewed settings, suggesting more stable convergence behaviour despite potentially slower rates. This pattern is particularly evident in highly skewed Gamma and Weibull cases, where the width of NW confidence intervals is often less than half that of the Iterative method. This suggests that while NW might be preferred for symmetric heavy-tailed data, the Iterative method's convergence advantages become more pronounced in skewed settings, albeit with greater variability in its convergence behaviour.

Implementation Considerations The empirical results motivate the following practical guidance:

- **Small samples ($n \leq 200$):** Use NW estimator for symmetric distributions with moderate to light tails; prefer Iterative for heavy-tailed data.

- **Moderate to large samples ($n \geq 400$):** Prefer the Iterative method, particularly under skew or heavy tails.
- **Stability vs Efficiency:** Consider the NW estimator when stability of convergence is paramount when dealing with low sample size, as evidenced by its consistently narrower confidence intervals, particularly in skewed settings.

This relationship between sample size, distributional shape, and estimator performance offers a practical framework for method selection in applied settings.

Chapter 5

Discussion

5.1 Interpreting Performance Patterns

The performance patterns observed across distribution families suggest certain insights about density estimation approaches. The rather favourable performance of the QDE for right-skewed distributions appears to stem from its direct modelling of quantile relationships, which seems to handle asymmetric shapes rather nicely. Unlike kernel methods that place symmetric kernels around observations, which might occasionally struggle with asymmetric tails, the quantile-based approach maps reference tails to data tails in a somewhat more direct fashion. This tail-respecting transformation might explain why the the QDE approach shows a certain steadiness in regions where traditional kernel approaches display a touch more volatility. As illustrated in the density difference plot, the kernel density estimator seems to exhibit more pronounced oscillations around the true density, particularly in the right tail region, while the QDE method appears to maintain a rather more consistent relationship with the true density, showing smaller deviations that on the whole are more well-behaved.

The cross-over pattern between Normal Weights and Iterative QDE methods as sample size increases may reflect an interesting property of the weight refinement process. Initially, standard normal-based weights provide a reasonable starting point, but as sample size grows, the data increasingly supports estimation at the tails, allowing the Iterative procedure to refine these weights in a manner that appears beneficial.

5.2 Limitations and Opportunities for Refinement

The QDE method relies on the selection of a reference distribution to anchor the quantile mapping process. This choice implicitly encodes prior assumptions about the data's shape and tail behaviour. While the method is non-parametric in its transformation, the reference distribution still plays a guiding role, particularly in shaping the initial weight structure and influencing tail estimation. In practice, selecting a reference that aligns with broad features of the data, such as skewness or tail heaviness, can improve performance and stability.

However, this flexibility also introduces a potential vulnerability: if the reference distribution diverges too sharply from the true underlying distribution,

estimation accuracy may suffer. Such misspecification could lead to suboptimal weight assignments or distortions in tail behaviour, particularly when sample sizes are small. Incorporating diagnostics such as Q-Q plots or summary statistics (e.g. skewness, kurtosis) into the reference selection process may provide a pragmatic safeguard. Further research could explore automated strategies for selecting reference families based on pre-estimation criteria.

The current implementation may face certain computational limitations for very large datasets. The Iterative procedure, while seemingly improving estimation quality, introduces computational overhead that might be somewhat challenging for real-time applications or particularly large datasets. Applications involving high-frequency financial data or extensive simulations might benefit from algorithmic refinements or computing adaptations.

Additionally, boundary effects may occasionally present challenges for highly skewed distributions with hard bounds such as strictly positive data. While the quantile-based approach appears to mitigate boundary bias to some extent by preserving support limitations, additional boundary correction techniques might prove helpful for distributions with sharp cut-offs or point masses at boundaries.

Refinements in the smoothing parameter selection process might enhance the QDE approach. Cross-validation methods for selecting this parameter could possibly provide more consistent performance across different distribution types than the current ISE minimisation approach. In particular, k-fold cross-validation might better account for sample variability in smaller datasets and perhaps reduce overfitting risks. Although computational efficiency should be considered here.

Another improvement could be made in pinning down the optimal choice of fixed df in the Iterative procedure discussed in Section 3.4.2. In the current implementation $df = 4$ was chosen as a conservative guess. This could be refined and maybe optimised as already mentioned in 3.4.3.

Future work might examine these improvements across a range of distribution families and parameter configurations to develop implementation strategies for different data scenarios.

5.3 Extension to Discrete Distributions

A limitation of the current approach is its focus on continuous distributions. Discrete distributions present certain challenges for quantile-based methods due to the discontinuous nature of their distribution functions. Unlike continuous distributions where quantiles form a smooth mapping, discrete quantiles introduce jumps that complicate the transformation approach.

One potential adaptation might address this challenge through jittering (see Akinshin, 2024), whereby modest continuous noise is added to discrete observations to approximate continuity. Density estimation is then performed on the jittered data and subsequently mapped back to the original discrete setting. The addition of small-scale continuous noise could smooth out the discrete jumps in a way that makes the quantile-based transformation more tractable, though one would need to be rather careful about the choice of jittering distribution and its parameters.

The QDE framework might retain some conceptual value for discrete data if appropriate continuity assumptions are introduced, though this represents a rather challenging direction for future methodological development.

5.4 Applications in Financial Modelling

The QDE method's apparent strength with heavy-tailed distributions suggests potential applications in finance. Traditional financial models often assume normal or log-normal return distributions, but empirical evidence tends to show heavier tails that lead to modest underestimation of extreme risks. The quantile-based approach might provide more accurate Value-at-Risk (VaR) and Expected Shortfall estimates by better capturing these tail behaviours.

For option pricing models that rely on specific distributional assumptions, the QDE approach offers a diagnostic tool to identify when these assumptions appear at odds with observed data and a correction mechanism through density estimation. This dual functionality might help address the volatility smile phenomenon in options markets, where traditional Black-Scholes assumptions sometimes seem less than entirely satisfactory.

Credit risk modelling, which often deals with skewed loss distributions, represents another potential application area. The QDE method's ability to adapt to varying shapes of skewed distributions might improve loss given default estimates compared to both parametric approaches that perhaps simplify the distribution and standard kernel methods that occasionally struggle with boundary effects near zero.

For challenging financial data, combining the QDE with heavy-tailed reference distributions like skewed-t could improve tail mapping for equity or foreign exchange returns, offering a balanced approach between parametric and non-parametric risk modelling.

5.5 Integration into Model-Based Diagnostic Workflows

The `fitdistr` function in R uses a maximum likelihood approach to estimate the best-fitting parameters for a user-specified parametric distribution. But what if the user assumes the wrong distribution? At present, there is no built-in diagnostic to warn against a poor distributional choice. The function simply returns the estimated parameters, even if the model is a poor fit.

This is where the QDE method could offer added value. By taking the fitted distribution from `fitdistr` as the reference, the QDE could be used to generate a smoothed density estimate. Comparing this estimate to the original fitted parametric density would provide a useful diagnostic: if the QDE estimate diverges significantly from the parametric fit, it would indicate that the assumed distribution may not be appropriate for the data. In this way, the QDE serves as a semi-parametric diagnostic tool, highlighting potential misfits in distributional assumptions especially in the tails or under skewness that might not be obvious from standard output.

Chapter 6

Conclusions

This paper has developed and evaluated a quantile-based approach to density estimation that unifies distribution checking and non-parametric density estimation. By mapping between a reference distribution and the data distribution through their quantiles, the approach serves both as a distributional diagnostic tool and as an effective density estimator.

The primary methodological contribution lies in establishing a connection between quantile relationships and formal density estimation. Using a reference distribution as a starting point for density estimation, then transforming this reference through quantile relationships to match the target data, represents an approach distinct from many standard methods. By leveraging this connection that is familiar to practitioners through Q-Q plots, the method transforms a common visual diagnostic tool into an estimation procedure with theoretical foundations.

From a practical perspective, the method offers value for statistical workflows. It integrates distributional diagnostics and density estimation in contexts where Q-Q plots are already standard practice, allowing practitioners to move from identifying departures from assumed distributions to obtaining corrected density estimates. This integration may streamline statistical analysis and provide a more consistent treatment of distributional assumptions.

The simulation studies across multiple distribution families have demonstrated that the QDE method provides competitive or superior performance compared to kernel density estimation in many scenarios. The Iterative weighting procedure generally yields the best results for moderate to large sample sizes, particularly when the target distribution is skewed or heavy-tailed. In small samples, however, the normal-weight (NW) estimator may offer more stable performance, especially under conditions of heavy skewness, where the Iterative method can be prone to higher variability. This makes NW a useful alternative when convergence stability is prioritised over absolute efficiency.

The performance advantages for heavy-tailed distributions suggest potential applications in fields such as finance, economics, and actuarial science, where such distributions frequently arise and accurate tail estimation is rather important (Burnecki et al., 2000; Theodossiou, 1998). The method's ability to adapt to varying distributional shapes without requiring parametric specification offers a certain flexibility while maintaining estimation quality.

Beyond the current implementation, future research could explore discrete data extensions, improved reference distribution selection under model misspecification,

and computational optimisations for large-scale use. Additionally, integrating the approach into regression or generalised linear model frameworks could enhance model diagnostics and inference, particularly when distributional assumptions are suspect.

In conclusion, the quantile-based density estimation approach offers a theoretically grounded method that bridges diagnostic checking and density estimation. By examining quantile relationships between reference and data distributions, a technique not commonly employed in other density estimation methods, it provides information about distributional properties and creates density estimates for diverse scenarios. As distributional assumptions remain important to valid statistical inference, this unified approach may represent a useful addition to the statistician's toolkit.

Appendix A

Appendix

A.1 t-Distribution

A.1.1 Density Estimation Results

Table A.1: Results from distribution estimation simulations. Each section shows the means and standard deviations (SDs) of the Integrated Squared Error (ISE) for the Iterative, Normal Weights (NW), KDE, and FD methods at different sample sizes, for various distributions defined by their degrees of freedom (df).

n	**Means**				**SDs**			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
df = 2.01								
100	0.009095	0.009191	0.005390	0.0003948	0.005443	0.005642	0.004244	0.0005754
200	0.003242	0.003243	0.003198	0.0001868	0.001941	0.001949	0.002094	0.0002460
400	0.001597	0.001543	0.001945	0.00009432	0.001041	0.001042	0.001173	0.0001341
800	0.0008871	0.0008526	0.001150	0.00004881	0.0005654	0.0005800	0.0006763	0.00007266
1600	0.0005593	0.0005003	0.0006892	0.00002438	0.0003087	0.0002995	0.0003748	0.00003402
3200	0.0003437	0.0002918	0.0004189	0.00001130	0.0001773	0.0001664	0.0002199	0.00001600
6400	0.0002038	0.0001677	0.0002431	0.000006305	0.00009544	0.00008719	0.0001171	0.000008433
12800	0.0001231	0.00009902	0.0001455	0.000002978	0.00005825	0.00005127	0.00007824	0.000004080
df = 3								
100	0.006762	0.006786	0.005165	0.0004173	0.003665	0.003724	0.004125	0.0006147
200	0.002668	0.002728	0.003175	0.0002213	0.001595	0.001714	0.002170	0.0003224
400	0.001421	0.001426	0.001971	0.0001042	0.001051	0.001091	0.001307	0.0001488
800	0.0007457	0.0007268	0.001169	0.00004270	0.0004950	0.0004878	0.0006904	0.00005950
1600	0.0004507	0.0004270	0.0006904	0.00002147	0.0002635	0.0002560	0.0003779	0.00002876
3200	0.0002895	0.0002687	0.0004110	0.00001095	0.0001619	0.0001592	0.0002285	0.00001545
6400	0.0001743	0.0001584	0.0002378	0.000005683	0.00009356	0.00008952	0.0001181	0.000007873
12800	0.0001035	0.00009114	0.0001394	0.000002779	0.00004990	0.00004644	0.00006603	0.000003867
df = 6								

Continued on next page

Table A.1 (continued)

n	Means				SDs			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
100	0.004328	0.004310	0.005287	0.0005097	0.002686	0.002626	0.003940	0.0005559
200	0.002049	0.002062	0.003122	0.0002601	0.001381	0.001384	0.002095	0.0003479
400	0.001119	0.001138	0.001895	0.0001188	0.0007605	0.0007783	0.001263	0.0001817
800	0.0006334	0.0006366	0.001130	0.00005460	0.0004337	0.0004405	0.0006801	0.00007740
1600	0.0003703	0.0003690	0.0006779	0.00002563	0.0002658	0.0002664	0.0004006	0.00003960
3200	0.0002299	0.0002266	0.0004184	0.00001361	0.0001567	0.0001562	0.0002487	0.00002076
6400	0.0001336	0.0001307	0.0002385	0.000006360	0.00007846	0.00007782	0.0001258	0.000008282
12800	0.00008035	0.00007789	0.0001403	0.000003127	0.00004026	0.00003961	0.00006594	0.000004120
df = 12								
100	0.003298	0.003293	0.005039	0.0003153	0.002681	0.002618	0.003864	0.0003695
200	0.001571	0.001579	0.002988	0.0001770	0.001302	0.001294	0.002205	0.0001913
400	0.0008148	0.0008245	0.001816	0.00009295	0.0006616	0.0006693	0.001254	0.0001162
800	0.0004792	0.0004844	0.001087	0.00004984	0.0003512	0.0003526	0.0006677	0.00006716
1600	0.0002954	0.0002973	0.0006606	0.00002578	0.0002339	0.0002352	0.0004110	0.00003586
3200	0.0001772	0.0001777	0.0004008	0.00001235	0.0001337	0.0001344	0.0002313	0.00001687
6400	0.0001014	0.0001013	0.0002384	0.000006056	0.00007195	0.00007221	0.0001292	0.000008537
12800	0.00005862	0.00005823	0.0001399	0.000002988	0.00003345	0.00003333	0.00006333	0.000004089
df = 20								
100	0.003127	0.003109	0.005139	0.0002323	0.002875	0.002826	0.004060	0.0003715
200	0.001485	0.001489	0.003073	0.0001313	0.001289	0.001278	0.002243	0.0001755
400	0.0007201	0.0007204	0.001818	0.00007524	0.0006748	0.0006705	0.001301	0.00008589
800	0.0004059	0.0004077	0.001088	0.00004574	0.0003357	0.0003343	0.0006803	0.00005351
1600	0.0002381	0.0002394	0.0006626	0.00002308	0.0002008	0.0002006	0.0004155	0.00002923
3200	0.0001430	0.0001437	0.0003997	0.00001206	0.0001049	0.0001056	0.0002337	0.00001833
6400	0.00008202	0.00008220	0.0002337	0.000005979	0.00006038	0.00006050	0.0001242	0.000008442
12800	0.00004780	0.00004779	0.0001388	0.000002877	0.00003199	0.00003201	0.00006226	0.000004034

A.1.2 Confidence Intervals For Density Estimation Results

Table A.2: Results showing confidence intervals for different estimators across t distributions. Each section shows the lower and upper bounds of confidence intervals for the Iterative, NW, KDE, and FD methods at different sample sizes.

n	Confidence Intervals			
	Iterative	NW	KDE	FD
df = 2.01				
100	[0.0080282, 0.0101618]	[0.0080852, 0.0102968]	[0.0045582, 0.0062218]	[0.0002820, 0.0005076]

Continued on next page

Table A.2 (continued)

n	Confidence Intervals			
	Iterative	NW	KDE	FD
200	[0.0029730, 0.0035110]	[0.0029729, 0.0035131]	[0.0029078, 0.0034882]	[0.0001527, 0.0002209]
400	[0.0014950, 0.0016990]	[0.0014409, 0.0016451]	[0.0018300, 0.0020600]	[0.0000812, 0.0001075]
800	[0.0008479, 0.0009263]	[0.0008124, 0.0008928]	[0.0011031, 0.0011969]	[0.0000438, 0.0000538]
1600	[0.0005442, 0.0005744]	[0.0004856, 0.0005150]	[0.0006708, 0.0007076]	[0.0000227, 0.0000260]
3200	[0.0003376, 0.0003498]	[0.0002860, 0.0002976]	[0.0004113, 0.0004265]	[0.0000107, 0.0000119]
6400	[0.0002015, 0.0002061]	[0.0001656, 0.0001698]	[0.0002402, 0.0002460]	[0.0000061, 0.0000065]
12800	[0.0001221, 0.0001241]	[0.0000981, 0.0000999]	[0.0001441, 0.0001469]	[0.0000029, 0.0000030]
df = 3				
100	[0.0060437, 0.0074803]	[0.0060561, 0.0075159]	[0.0043565, 0.0059735]	[0.0002968, 0.0005378]
200	[0.0024469, 0.0028891]	[0.0024905, 0.0029655]	[0.0028743, 0.0034757]	[0.0001766, 0.0002660]
400	[0.0013180, 0.0015240]	[0.0013191, 0.0015329]	[0.0018429, 0.0020991]	[0.0000896, 0.0001188]
800	[0.0007114, 0.0007800]	[0.0006930, 0.0007606]	[0.0011212, 0.0012168]	[0.0000386, 0.0000468]
1600	[0.0004378, 0.0004636]	[0.0004145, 0.0004395]	[0.0006719, 0.0007089]	[0.0000201, 0.0000229]
3200	[0.0002839, 0.0002951]	[0.0002632, 0.0002742]	[0.0004031, 0.0004189]	[0.0000104, 0.0000115]
6400	[0.0001720, 0.0001766]	[0.0001562, 0.0001606]	[0.0002349, 0.0002407]	[0.0000055, 0.0000059]
12800	[0.0001026, 0.0001044]	[0.0000903, 0.0000919]	[0.0001383, 0.0001405]	[0.0000027, 0.0000028]
df = 6				
100	[0.0038015, 0.0048545]	[0.0037953, 0.0048247]	[0.0045148, 0.0060592]	[0.0004007, 0.0006187]
200	[0.0018576, 0.0022404]	[0.0018702, 0.0022538]	[0.0028316, 0.0034124]	[0.0002119, 0.0003083]
400	[0.0010445, 0.0011935]	[0.0010617, 0.0012143]	[0.0017712, 0.0020188]	[0.0001010, 0.0001366]
800	[0.0006033, 0.0006635]	[0.0006061, 0.0006671]	[0.0010829, 0.0011771]	[0.0000492, 0.0000600]
1600	[0.0003573, 0.0003833]	[0.0003559, 0.0003821]	[0.0006583, 0.0006975]	[0.0000237, 0.0000276]
3200	[0.0002245, 0.0002353]	[0.0002212, 0.0002320]	[0.0004098, 0.0004270]	[0.0000129, 0.0000143]
6400	[0.0001317, 0.0001355]	[0.0001288, 0.0001326]	[0.0002354, 0.0002416]	[0.0000062, 0.0000066]
12800	[0.0000797, 0.0000810]	[0.0000772, 0.0000786]	[0.0001392, 0.0001414]	[0.0000031, 0.0000032]
df = 12				
100	[0.0027725, 0.0038235]	[0.0027799, 0.0038061]	[0.0042817, 0.0057963]	[0.0002429, 0.0003877]
200	[0.0013906, 0.0017514]	[0.0013997, 0.0017583]	[0.0026824, 0.0032936]	[0.0001505, 0.0002035]
400	[0.0007500, 0.0008796]	[0.0007589, 0.0008901]	[0.0016931, 0.0019389]	[0.0000816, 0.0001043]
800	[0.0004549, 0.0005035]	[0.0004600, 0.0005088]	[0.0010407, 0.0011333]	[0.0000452, 0.0000545]
1600	[0.0002839, 0.0003069]	[0.0002858, 0.0003088]	[0.0006405, 0.0006807]	[0.0000240, 0.0000275]
3200	[0.0001726, 0.0001818]	[0.0001730, 0.0001824]	[0.0003928, 0.0004088]	[0.0000118, 0.0000129]
6400	[0.0000996, 0.0001032]	[0.0000995, 0.0001031]	[0.0002352, 0.0002416]	[0.0000058, 0.0000063]
12800	[0.0000580, 0.0000592]	[0.0000577, 0.0000588]	[0.0001388, 0.0001410]	[0.0000029, 0.0000031]
df = 20				
100	[0.0025635, 0.0036905]	[0.0025551, 0.0036629]	[0.0042432, 0.0058346]	[0.0001595, 0.0003051]
200	[0.0013064, 0.0016636]	[0.0013119, 0.0016661]	[0.0027621, 0.0033839]	[0.0001070, 0.0001556]

Continued on next page

Table A.2 (continued)

n	Confidence Intervals			
	Iterative	NW	KDE	FD
400	[0.0006540, 0.0007862]	[0.0006547, 0.0007861]	[0.0016905, 0.0019455]	[0.0000668, 0.0000837]
800	[0.0003826, 0.0004292]	[0.0003845, 0.0004309]	[0.0010409, 0.0011351]	[0.0000420, 0.0000494]
1600	[0.0002283, 0.0002479]	[0.0002296, 0.0002492]	[0.0006422, 0.0006830]	[0.0000216, 0.0000245]
3200	[0.0001394, 0.0001466]	[0.0001400, 0.0001474]	[0.0003916, 0.0004078]	[0.0000114, 0.0000127]
6400	[0.0000805, 0.0000835]	[0.0000807, 0.0000837]	[0.0002307, 0.0002367]	[0.0000058, 0.0000062]
12800	[0.0000472, 0.0000484]	[0.0000472, 0.0000483]	[0.0001377, 0.0001399]	[0.0000028, 0.0000029]

A.1.3 Efficiency Results

Table A.3: Efficiency ratios from t distribution simulations. Each section shows results for various t distributions defined by their degrees of freedom.

n	NW/ITER	KDE/ITER	KDE/NW
t (df = 2.01)			
100	1.0106	0.5926	0.5864
200	1.0003	0.9864	0.9861
400	0.9662	1.2179	1.2605
800	0.9611	1.2964	1.3488
1600	0.8945	1.2323	1.3776
3200	0.8490	1.2188	1.4356
6400	0.8229	1.1928	1.4496
12800	0.8044	1.1820	1.4694
t (df = 3)			
100	1.0035	0.7638	0.7611
200	1.0225	1.1900	1.1639
400	1.0035	1.3871	1.3822
800	0.9747	1.5677	1.6084
1600	0.9474	1.5318	1.6169
3200	0.9282	1.4197	1.5296
6400	0.9088	1.3643	1.5013
12800	0.8806	1.3469	1.5295
t (df = 6)			
100	0.9958	1.2216	1.2267
200	1.0063	1.5237	1.5141
400	1.0170	1.6935	1.6652
800	1.0051	1.7840	1.7751
1600	0.9965	1.8307	1.8371
3200	0.9856	1.8199	1.8464
6400	0.9783	1.7852	1.8248
12800	0.9694	1.7461	1.8013
t (df = 12)			
100	0.9985	1.5279	1.5302

Continued on next page

Table A.3 (continued)

n	NW/ITER	KDE/ITER	KDE/NW
200	1.0051	1.9020	1.8923
400	1.0119	2.2288	2.2025
800	1.0109	2.2684	2.2440
1600	1.0064	2.2363	2.2220
3200	1.0028	2.2619	2.2555
6400	0.9990	2.3511	2.3534
12800	0.9933	2.3866	2.4025
t (df = 20)			
100	0.9942	1.6434	1.6529
200	1.0027	2.0694	2.0638
400	1.0004	2.5246	2.5236
800	1.0044	2.6805	2.6686
1600	1.0055	2.7829	2.7678
3200	1.0049	2.7951	2.7815
6400	1.0022	2.8493	2.8431
12800	0.9998	2.9038	2.9044

A.1.4 Convergence Results

Table A.4: Convergence rates from t distribution simulations. Each section shows results for various t distributions defined by their df parameter.

Estimator	Empirical Rate	Theoretical Rate	95% CI
t(df = 2.01)			
Iterative Weights	-0.8317	-0.8000	[-0.9535, -0.7098]
KDE	-0.7163	-0.8000	[-0.7276, -0.7049]
Normal Weights	-0.8871	-0.8000	[-1.0026, -0.7715]
FD	-0.9779	-1.0000	[-1.0318, -0.9239]
t(df = 3)			
Iterative Weights	-0.8205	-0.8000	[-0.9306, -0.7103]
KDE	-0.7105	-0.8000	[-0.7370, -0.6840]
Normal Weights	-0.8522	-0.8000	[-0.9553, -0.7490]
FD	-1.0439	-1.0000	[-1.1065, -0.9814]
t(df = 6)			
Iterative Weights	-0.7951	-0.8000	[-0.8653, -0.7249]
KDE	-0.7088	-0.8000	[-0.7200, -0.6975]
Normal Weights	-0.8037	-0.8000	[-0.8692, -0.7383]
FD	-1.0632	-1.0000	[-1.1094, -1.0169]
t(df = 12)			
Iterative Weights	-0.7878	-0.8000	[-0.8555, -0.7200]
KDE	-0.6918	-0.8000	[-0.7019, -0.6817]
Normal Weights	-0.7908	-0.8000	[-0.8562, -0.7255]
FD	-1.0054	-1.0000	[-1.0412, -0.9697]
t(df = 20)			

Continued on next page

Table A.4 (continued)

Estimator	Empirical Rate	Theoretical Rate	95% CI
Iterative Weights	-0.8169	-0.8000	[-0.8951, -0.7386]
KDE	-0.6991	-0.8000	[-0.7058, -0.6925]
Normal Weights	-0.8183	-0.8000	[-0.8964, -0.7403]
FD	-0.9747	-1.0000	[-1.0288, -0.9207]

A.2 Gamma

A.2.1 Density Estimation Results

Table A.5: Results from Gamma distribution simulations. Each section shows the means and standard deviations (SDs) of the ISE for the Iterative, NW, KDE, and fitdistr (FD) methods at different sample sizes, for various Gamma distributions defined by their shape and scale parameters.

n	**Means**				**SDs**			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
$\mathbf{\Gamma}(\text{shape} = \mathbf{1.2}, \text{scale} = \mathbf{2})$								
100	0.008315	0.006125	0.008917	0.002920	0.01338	0.005129	0.004701	0.003220
200	0.004007	0.003810	0.006013	0.001521	0.01102	0.002694	0.002833	0.001751
400	0.002012	0.002533	0.004160	0.0007888	0.001871	0.001489	0.001688	0.0009398
800	0.001220	0.001738	0.002899	0.0003452	0.0006865	0.0008513	0.0009402	0.0003590
1600	0.0008010	0.001164	0.002022	0.0001611	0.0004090	0.0005095	0.0005936	0.0001760
3200	0.0005335	0.0007847	0.001415	0.00008593	0.0002252	0.0002826	0.0003326	0.00009257
6400	0.0003591	0.0005271	0.0009946	0.00004554	0.0001303	0.0001633	0.0001994	0.00005587
12800	0.0002462	0.0003510	0.0006928	0.00002235	0.00006888	0.00008677	0.0001077	0.00002631
$\mathbf{\Gamma}(\text{shape} = \mathbf{2}, \text{scale} = \mathbf{1.5})$								
100	0.005252	0.003560	0.004410	0.002081	0.01529	0.003140	0.002819	0.002230
200	0.001652	0.001879	0.002694	0.0009555	0.001353	0.001454	0.001530	0.0009504
400	0.001021	0.001223	0.001790	0.0004883	0.0008054	0.0009211	0.001008	0.0005067
800	0.0006112	0.0007611	0.001110	0.0002370	0.0004161	0.0004987	0.0005492	0.0002534
1600	0.0003680	0.0004656	0.0006991	0.0001076	0.0002314	0.0002699	0.0003045	0.0001146
3200	0.0002327	0.0002981	0.0004355	0.00005700	0.0001217	0.0001397	0.0001626	0.00005698
6400	0.0001384	0.0001803	0.0002653	0.00002797	0.00006958	0.00007808	0.00008634	0.00002998
12800	0.00008391	0.0001118	0.0001609	0.00001414	0.00003833	0.00004478	0.00004923	0.00001490
$\mathbf{\Gamma}(\text{shape} = \mathbf{3}, \text{scale} = \mathbf{1})$								
100	0.003454	0.003523	0.004438	0.002008	0.003558	0.002999	0.002848	0.001984
200	0.001735	0.001850	0.002735	0.0009644	0.001578	0.001628	0.001719	0.001065
400	0.0009950	0.001109	0.001723	0.0004787	0.0008196	0.0008677	0.0009546	0.0004882
800	0.0005899	0.0006775	0.001056	0.0002339	0.0004902	0.0005195	0.0005623	0.0002429

Continued on next page

Table A.5 (continued)

n	Means				SDs			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
1600	0.0003531	0.0004212	0.0006611	0.0001177	0.0002452	0.0002734	0.0003174	0.0001235
3200	0.0002173	0.0002585	0.0003961	0.00006297	0.0001246	0.0001381	0.0001748	0.00006100
6400	0.0001249	0.0001502	0.0002267	0.00003143	0.00006460	0.00007130	0.00008942	0.00003078
12800	0.00007423	0.00009158	0.0001339	0.00001541	0.00003585	0.00004039	0.00004804	0.00001537
$\mathbf{\Gamma}(\text{shape} = \mathbf{4}, \text{scale} = \mathbf{2})$								
100	0.001901	0.001377	0.001685	0.0008102	0.007614	0.001223	0.001145	0.0008178
200	0.0006541	0.0006929	0.001025	0.0003788	0.0006230	0.0006378	0.0006680	0.0003998
400	0.0003794	0.0004104	0.0006473	0.0001936	0.0003293	0.0003383	0.0003920	0.0001886
800	0.0002104	0.0002346	0.0003846	0.00008886	0.0001638	0.0001728	0.0002146	0.00008936
1600	0.0001223	0.0001403	0.0002320	0.00004501	0.00008407	0.00009063	0.0001134	0.00004734
3200	0.00007780	0.00008951	0.0001428	0.00002355	0.00004768	0.00005107	0.00006940	0.00002435
6400	0.00004705	0.00005438	0.00008463	0.00001216	0.00002751	0.00003018	0.00003629	0.00001291
12800	0.00002782	0.00003222	0.00004904	0.000006156	0.00001429	0.00001530	0.00001808	0.000006405
$\mathbf{\Gamma}(\text{shape} = \mathbf{5}, \text{scale} = \mathbf{0.8})$								
100	0.003770	0.002956	0.003558	0.001752	0.01314	0.002678	0.002481	0.001782
200	0.001384	0.001436	0.002138	0.0008146	0.001302	0.001309	0.001380	0.0008124
400	0.0007799	0.0008249	0.001328	0.0004294	0.0006861	0.0006990	0.0008217	0.0004240
800	0.0004275	0.0004624	0.0007999	0.0001944	0.0003358	0.0003466	0.0004557	0.0001917
1600	0.0002424	0.0002704	0.0004745	0.00009536	0.0001733	0.0001823	0.0002379	0.00009651
3200	0.0001536	0.0001732	0.0002945	0.00004924	0.00009762	0.0001042	0.0001514	0.00005020
6400	0.00009306	0.0001047	0.0001727	0.00002588	0.00005631	0.00006041	0.00007687	0.00002638
12800	0.00005520	0.00006199	0.00009977	0.00001285	0.00002915	0.00003119	0.00003935	0.00001274

A.2.2 Confidence Intervals For Density Estimation Results

Table A.6: Results showing confidence intervals for different estimators across Gamma distributions. Each section shows the lower and upper bounds of confidence intervals for the Iterative, NW, KDE, and FD methods at different sample sizes.

n	Confidence Intervals			
	Iterative	NW	KDE	FD
$\mathbf{\Gamma}(\text{shape} = \mathbf{1.2}, \text{scale} = \mathbf{2})$				
100	[0.0056925, 0.0109375]	[0.0051197, 0.0071303]	[0.0079956, 0.0098384]	[0.0022889, 0.0035511]
200	[0.0024797, 0.0055343]	[0.0034366, 0.0041834]	[0.0056204, 0.0064056]	[0.0012783, 0.0017637]
400	[0.0018286, 0.0021954]	[0.0023871, 0.0026789]	[0.0039946, 0.0043254]	[0.0006967, 0.0008809]
800	[0.0011724, 0.0012676]	[0.0016790, 0.0017970]	[0.0026338, 0.0027642]	[0.0003203, 0.0003701]
1600	[0.0007810, 0.0008210]	[0.0011390, 0.0011890]	[0.0019929, 0.0020511]	[0.0001525, 0.0001697]

Continued on next page

Table A.6 (continued)

n	**Confidence Intervals**			
	Iterative	NW	KDE	FD
3200	[0.0005257, 0.0005413]	[0.0007749, 0.0007945]	[0.0014035, 0.0014265]	[0.0000827, 0.0000891]
6400	[0.0003559, 0.0003623]	[0.0005231, 0.0005311]	[0.0009897, 0.0009995]	[0.0000442, 0.0000469]
12800	[0.0002450, 0.0002474]	[0.0003495, 0.0003525]	[0.0006909, 0.0006947]	[0.0000219, 0.0000228]
	$\mathbf{\Gamma}(\text{shape} = \mathbf{2}, \text{scale} = \mathbf{1.5})$			
100	[0.0022552, 0.0082488]	[0.0029446, 0.0041754]	[0.0038575, 0.0049625]	[0.0016439, 0.0025181]
200	[0.0014645, 0.0018395]	[0.0016775, 0.0020805]	[0.0024820, 0.0029060]	[0.0008238, 0.0010872]
400	[0.0009421, 0.0010999]	[0.0011327, 0.0013133]	[0.0016912, 0.0018888]	[0.0004386, 0.0005380]
800	[0.0005824, 0.0006400]	[0.0007265, 0.0007957]	[0.0010719, 0.0011481]	[0.0002194, 0.0002546]
1600	[0.0003567, 0.0003793]	[0.0004524, 0.0004788]	[0.0006842, 0.0007140]	[0.0001020, 0.0001132]
3200	[0.0002285, 0.0002369]	[0.0002933, 0.0003029]	[0.0004299, 0.0004411]	[0.0000550, 0.0000590]
6400	[0.0001367, 0.0001401]	[0.0001784, 0.0001822]	[0.0002632, 0.0002674]	[0.0000272, 0.0000287]
12800	[0.0000832, 0.0000846]	[0.0001110, 0.0001126]	[0.0001600, 0.0001618]	[0.0000139, 0.0000144]
	$\mathbf{\Gamma}(\text{shape} = \mathbf{3}, \text{scale} = \mathbf{1})$			
100	[0.0027566, 0.0041514]	[0.0023352, 0.0041108]	[0.0038798, 0.0053962]	[0.0016191, 0.0023969]
200	[0.0015163, 0.0019537]	[0.0016244, 0.0020756]	[0.0024968, 0.0029732]	[0.0008168, 0.0011120]
400	[0.0009147, 0.0010753]	[0.0010240, 0.0011940]	[0.0016294, 0.0018166]	[0.0004309, 0.0005265]
800	[0.0005559, 0.0006239]	[0.0006415, 0.0007135]	[0.0010170, 0.0010950]	[0.0002171, 0.0002507]
1600	[0.0003411, 0.0003651]	[0.0004078, 0.0004346]	[0.0006458, 0.0006767]	[0.0001116, 0.0001238]
3200	[0.0002130, 0.0002216]	[0.0002537, 0.0002633]	[0.0003900, 0.0004022]	[0.0000599, 0.0000651]
6400	[0.0001233, 0.0001265]	[0.0001485, 0.0001519]	[0.0002245, 0.0002289]	[0.0000307, 0.0000322]
12800	[0.0000736, 0.0000749]	[0.0000909, 0.0000923]	[0.0001331, 0.0001347]	[0.0000151, 0.0000157]
	$\mathbf{\Gamma}(\text{shape} = \mathbf{4}, \text{scale} = \mathbf{2})$			
100	[0.0004087, 0.0033933]	[0.0011373, 0.0016167]	[0.0014606, 0.0019094]	[0.0006499, 0.0009705]
200	[0.0005678, 0.0007404]	[0.0006045, 0.0007813]	[0.0009324, 0.0011176]	[0.0003234, 0.0004342]
400	[0.0003471, 0.0004117]	[0.0003772, 0.0004436]	[0.0006089, 0.0006857]	[0.0001751, 0.0002121]
800	[0.0001990, 0.0002218]	[0.0002226, 0.0002466]	[0.0003697, 0.0003995]	[0.0000827, 0.0000951]
1600	[0.0001182, 0.0001264]	[0.0001359, 0.0001447]	[0.0002264, 0.0002376]	[0.0000427, 0.0000473]
3200	[0.0000761, 0.0000795]	[0.0000877, 0.0000913]	[0.0001404, 0.0001452]	[0.0000227, 0.0000244]
6400	[0.0000464, 0.0000477]	[0.0000536, 0.0000551]	[0.0000837, 0.0000855]	[0.0000118, 0.0000125]
12800	[0.0000276, 0.0000281]	[0.0000320, 0.0000325]	[0.0000487, 0.0000494]	[0.0000060, 0.0000063]
	$\mathbf{\Gamma}(\text{shape} = \mathbf{5}, \text{scale} = \mathbf{0.8})$			
100	[0.0011946, 0.0035454]	[0.0024311, 0.0034809]	[0.0030717, 0.0040443]	[0.0014027, 0.0021013]
200	[0.0012036, 0.0015644]	[0.0012546, 0.0016174]	[0.0019467, 0.0023293]	[0.0007020, 0.0009272]
400	[0.0007127, 0.0008471]	[0.0007564, 0.0008934]	[0.0012475, 0.0014085]	[0.0003878, 0.0004710]
800	[0.0004042, 0.0004508]	[0.0004384, 0.0004864]	[0.0007683, 0.0008315]	[0.0001811, 0.0002077]
1600	[0.0002339, 0.0002509]	[0.0002615, 0.0002793]	[0.0004628, 0.0004862]	[0.0000906, 0.0001001]
3200	[0.0001502, 0.0001570]	[0.0001696, 0.0001768]	[0.0002893, 0.0002997]	[0.0000475, 0.0000510]

Continued on next page

Table A.6 (continued)

n	Confidence Intervals			
	Iterative	NW	KDE	FD
6400	[0.0000917, 0.0000944]	[0.0001032, 0.0001062]	[0.0001708, 0.0001746]	[0.0000252, 0.0000265]
12800	[0.0000547, 0.0000557]	[0.0000614, 0.0000625]	[0.0000991, 0.0001005]	[0.0000126, 0.0000131]

A.2.3 Efficiency Results

Table A.7: Efficiency ratios from Gamma distribution simulations. Each section shows results for various Gamma distributions defined by their shape and scale parameters.

n	NW/ITER	KDE/ITER	KDE/NW
$\mathbf{\Gamma}$ **(shape = 1.2, scale = 2)**			
100	0.7366	1.0724	1.4558
200	0.9508	1.5006	1.5782
400	1.2589	2.0676	1.6423
800	1.4246	2.3762	1.6680
1600	1.4532	2.5243	1.7371
3200	1.4709	2.6523	1.8032
6400	1.4678	2.7697	1.8869
12800	1.4257	2.8140	1.9738
$\mathbf{\Gamma}$ **(shape = 2, scale = 1.5)**			
100	0.6778	0.8397	1.2388
200	1.1374	1.6308	1.4337
400	1.1978	1.7532	1.4636
800	1.2453	1.8161	1.4584
1600	1.2652	1.8997	1.5015
3200	1.2810	1.8715	1.4609
6400	1.3027	1.9169	1.4714
12800	1.3324	1.9175	1.4392
$\mathbf{\Gamma}$ **(shape = 3, scale = 1)**			
100	1.0200	1.2849	1.2597
200	1.0663	1.5764	1.4784
400	1.1146	1.7317	1.5537
800	1.1485	1.7901	1.5587
1600	1.1929	1.8723	1.5696
3200	1.1896	1.8228	1.5323
6400	1.2026	1.8151	1.5093
12800	1.2337	1.8039	1.4621
$\mathbf{\Gamma}$ **(shape = 4, scale = 2)**			
100	0.7244	0.8864	1.2237
200	1.0593	1.5670	1.4793
400	1.0817	1.7061	1.5772
800	1.1150	1.8279	1.6394

Continued on next page

Table A.7 (continued)

n	NW/ITER	KDE/ITER	KDE/NW
1600	1.1472	1.8970	1.6536
3200	1.1505	1.8355	1.5954
6400	1.1558	1.7987	1.5563
12800	1.1582	1.7628	1.5220
Γ (shape = 5, scale = 0.8)			
100	0.7841	0.9438	1.2037
200	1.0376	1.5448	1.4889
400	1.0577	1.7028	1.6099
800	1.0816	1.8711	1.7299
1600	1.1155	1.9575	1.7548
3200	1.1276	1.9173	1.7003
6400	1.1251	1.8558	1.6495
12800	1.1230	1.8074	1.6095

A.2.4 Convergence Results

Table A.8: Convergence rates from Gamma distribution simulations. Each section shows results for various Gamma distributions defined by their shape and scale parameters.

Estimator	Empirical Rate	Theoretical Rate	95% CI
Γ (shape = 1.2, scale = 2)			
Iterative Weights	-0.5947	-0.8000	[-0.6416, -0.5477]
KDE	-0.4997	-0.8000	[-0.5046, -0.4948]
Normal Weights	-0.5297	-0.8000	[-0.5423, -0.5170]
FD	-1.0217	-1.0000	[-1.0472, -0.9962]
Γ (shape = 2, scale = 1.5)			
Iterative Weights	-0.7030	-0.8000	[-0.7545, -0.6514]
KDE	-0.6501	-0.8000	[-0.6676, -0.6326]
Normal Weights	-0.6485	-0.8000	[-0.6760, -0.6210]
FD	-1.0236	-1.0000	[-1.0507, -0.9964]
Γ (shape = 3, scale = 1)			
Iterative Weights	-0.7244	-0.8000	[-0.7597, -0.6892]
KDE	-0.6924	-0.8000	[-0.7174, -0.6674]
Normal Weights	-0.6906	-0.8000	[-0.7214, -0.6599]
FD	-0.9861	-1.0000	[-1.0148, -0.9574]
Γ (shape = 4, scale = 2)			
Iterative Weights	-0.7365	-0.8000	[-0.7891, -0.6839]
KDE	-0.6969	-0.8000	[-0.7136, -0.6803]
Normal Weights	-0.7078	-0.8000	[-0.7466, -0.6690]
FD	-0.9938	-1.0000	[-1.0305, -0.9571]
Γ (shape = 5, scale = 0.8)			
Iterative Weights	-0.7571	-0.8000	[-0.8176, -0.6966]
KDE	-0.7018	-0.8000	[-0.7165, -0.6870]

Continued on next page

Table A.8 (continued)

Estimator	Empirical Rate	Theoretical Rate	95% CI
Normal Weights	-0.7311	-0.8000	[-0.7794, -0.6829]
FD	-1.0074	-1.0000	[-1.0402, -0.9745]

A.3 Weibull

A.3.1 Density Estimation Results

Table A.9: Results from Weibull simulations. Each section shows the means and standard deviations (SDs) of the ISE for the Iterative, NW, KDE, and FD methods at different sample sizes, for various Weibull distributions defined by their shape and scale parameters.

n	**Means**				**SDs**			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
Weibull (shape = 1.2, scale = 2)								
100	0.01016	0.006912	0.009102	0.003519	0.02328	0.005691	0.005230	0.004257
200	0.004000	0.004501	0.006302	0.001582	0.007334	0.003392	0.003204	0.001853
400	0.002284	0.003073	0.004408	0.0007868	0.001708	0.001985	0.001972	0.0008281
800	0.001423	0.002010	0.003007	0.0003733	0.0009364	0.001129	0.001124	0.0004055
1600	0.0009263	0.001312	0.002074	0.0001876	0.0004795	0.0005996	0.0006526	0.0002015
3200	0.0006192	0.0008845	0.001444	0.0001066	0.0002814	0.0003417	0.0003857	0.0001155
6400	0.0004007	0.0005742	0.0009991	0.00005395	0.0001454	0.0001841	0.0002180	0.00006423
12800	0.0002724	0.0003820	0.0007014	0.00002543	0.00008072	0.0001019	0.0001228	0.00003012
Weibull (shape = 1.5, scale = 3)								
100	0.004518	0.003691	0.003954	0.002156	0.01164	0.003327	0.002674	0.002504
200	0.001952	0.002196	0.002573	0.0009768	0.001977	0.001871	0.001658	0.001106
400	0.001170	0.001412	0.001673	0.0004840	0.0009885	0.001091	0.0009907	0.0004949
800	0.0007068	0.0008992	0.001055	0.0002290	0.0005312	0.0005951	0.0005293	0.0002367
1600	0.0004432	0.0005735	0.0006736	0.0001146	0.0002711	0.0003145	0.0003035	0.0001190
3200	0.0002897	0.0003800	0.0004415	0.00006483	0.0001549	0.0001830	0.0001760	0.00006710
6400	0.0001775	0.0002355	0.0002804	0.00003293	0.00008059	0.00009567	0.00009173	0.00003737
12800	0.0001130	0.0001496	0.0001818	0.00001548	0.00004661	0.00005499	0.00005497	0.00001743
Weibull (shape = 2, scale = 1.5)								
100	0.007555	0.007174	0.007479	0.004565	0.01165	0.006673	0.005194	0.005131
200	0.003738	0.003943	0.004668	0.002087	0.003627	0.003644	0.003247	0.002303
400	0.002139	0.002320	0.002907	0.001032	0.001942	0.001984	0.001766	0.001037
800	0.001253	0.001403	0.001755	0.0004872	0.0009987	0.001016	0.0009665	0.0004844
1600	0.0007685	0.0008875	0.001065	0.0002424	0.0005069	0.0005417	0.0005139	0.0002459
3200	0.0005037	0.0005879	0.0006760	0.0001367	0.0002907	0.0003191	0.0003014	0.0001362

Continued on next page

Table A.9 (continued)

n	Means				SDs			
	Iterative	NW	KDE	FD	Iterative	NW	KDE	FD
6400	0.0003046	0.0003602	0.0004058	0.00006966	0.0001477	0.0001622	0.0001656	0.00007590
12800	0.0001911	0.0002264	0.0002480	0.00003268	0.00008661	0.00009549	0.00009878	0.00003521
Weibull (**shape** $= \mathbf{2.5}$, **scale** $= \mathbf{1.2}$)								
100	0.009188	0.009181	0.01049	0.006461	0.009019	0.008871	0.007851	0.007159
200	0.004739	0.004804	0.006419	0.002970	0.004633	0.004574	0.004717	0.003242
400	0.002684	0.002762	0.003958	0.001466	0.002525	0.002525	0.002505	0.001466
800	0.001530	0.001603	0.002355	0.0006916	0.001301	0.001287	0.001419	0.0006767
1600	0.0009180	0.0009819	0.001395	0.0003432	0.0006592	0.0006712	0.0007473	0.0003453
3200	0.0005978	0.0006434	0.0008767	0.0001932	0.0003733	0.0003832	0.0004429	0.0001892
6400	0.0003609	0.0003949	0.0005172	0.00009862	0.0001844	0.0001919	0.0002536	0.0001054
12800	0.0002289	0.0002523	0.0003082	0.00004623	0.0001109	0.0001183	0.0001438	0.00004875
Weibull (**shape** $= \mathbf{3}$, **scale** $= \mathbf{1.8}$)								
100	0.006344	0.006301	0.007992	0.004909	0.006649	0.006540	0.006234	0.005399
200	0.003094	0.003108	0.004868	0.002265	0.003156	0.003136	0.003656	0.002457
400	0.001721	0.001724	0.003001	0.001117	0.001632	0.001621	0.001943	0.001116
800	0.0009876	0.0009988	0.001779	0.0005268	0.0008541	0.0008407	0.001115	0.0005114
1600	0.0005933	0.0006052	0.001046	0.0002610	0.0004558	0.0004545	0.0005940	0.0002617
3200	0.0003816	0.0003912	0.0006555	0.0001467	0.0002621	0.0002628	0.0003524	0.0001424
6400	0.0002297	0.0002380	0.0003848	0.00007499	0.0001299	0.0001316	0.0002049	0.00007936
12800	0.0001454	0.0001518	0.0002276	0.00003513	0.00007596	0.00007762	0.0001133	0.00003663

A.3.2 Confidence Intervals For Density Estimation Results

Table A.10: Results showing confidence intervals for different estimators across Weibull distributions. Each section shows the lower and upper bounds of confidence intervals for the Iterative, NW, KDE, and FD methods at different sample sizes.

n	Confidence Intervals			
	Iterative	NW	KDE	FD
Weibull(shape $= \mathbf{1.2}$, scale $= \mathbf{2}$)				
100	[0.0055971, 0.0147229]	[0.0057966, 0.0080274]	[0.0080769, 0.0101271]	[0.0026846, 0.0043534]
200	[0.0029836, 0.0050164]	[0.0040309, 0.0049711]	[0.0058579, 0.0067461]	[0.0013252, 0.0018388]
400	[0.0021166, 0.0024514]	[0.0028785, 0.0032675]	[0.0042147, 0.0046013]	[0.0007056, 0.0009680]
800	[0.0013581, 0.0014879]	[0.0019318, 0.0020882]	[0.0029291, 0.0030849]	[0.0003452, 0.0004014]
1600	[0.0009028, 0.0009498]	[0.0012826, 0.0013414]	[0.0020420, 0.0021060]	[0.0001777, 0.0001975]
3200	[0.0006094, 0.0006290]	[0.0008727, 0.0008963]	[0.0014306, 0.0014574]	[0.0001026, 0.0001106]
6400	[0.0003971, 0.0004043]	[0.0005697, 0.0005787]	[0.0009938, 0.0010044]	[0.0000524, 0.0000555]

Continued on next page

Table A.10 (continued)

n	Confidence Intervals			
	Iterative	NW	KDE	FD
12800	[0.0002710, 0.0002738]	[0.0003802, 0.0003838]	[0.0006993, 0.0007035]	[0.0000249, 0.0000260]
	Weibull(shape = **1.5**, scale = **3**)			
100	[0.0022366, 0.0067994]	[0.0030389, 0.0043431]	[0.0034299, 0.0044781]	[0.0016652, 0.0025468]
200	[0.0016780, 0.0022260]	[0.0019367, 0.0024553]	[0.0023432, 0.0028028]	[0.0008235, 0.0011301]
400	[0.0010731, 0.0012669]	[0.0013051, 0.0015189]	[0.0015759, 0.0017701]	[0.0004355, 0.0005325]
800	[0.0006700, 0.0007436]	[0.0008580, 0.0009404]	[0.0010183, 0.0010917]	[0.0002126, 0.0002454]
1600	[0.0004299, 0.0004565]	[0.0005581, 0.0005889]	[0.0006587, 0.0006885]	[0.0001088, 0.0001204]
3200	[0.0002843, 0.0002951]	[0.0003737, 0.0003863]	[0.0004354, 0.0004476]	[0.0000625, 0.0000672]
6400	[0.0001755, 0.0001795]	[0.0002332, 0.0002378]	[0.0002782, 0.0002826]	[0.0000320, 0.0000338]
12800	[0.0001122, 0.0001138]	[0.0001486, 0.0001506]	[0.0001808, 0.0001828]	[0.0000152, 0.0000158]
	Weibull(shape = **2**, scale = **1.5**)			
100	[0.0052716, 0.0098384]	[0.0058661, 0.0084819]	[0.0064610, 0.0084970]	[0.0035593, 0.0055707]
200	[0.0032353, 0.0042407]	[0.0034380, 0.0044480]	[0.0042180, 0.0051180]	[0.0017678, 0.0024062]
400	[0.0019487, 0.0023293]	[0.0021256, 0.0025144]	[0.0027339, 0.0030801]	[0.0009304, 0.0011336]
800	[0.0011838, 0.0013222]	[0.0013326, 0.0014734]	[0.0016880, 0.0018220]	[0.0004536, 0.0005208]
1600	[0.0007437, 0.0007933]	[0.0008610, 0.0009140]	[0.0010398, 0.0010902]	[0.0002304, 0.0002544]
3200	[0.0004936, 0.0005138]	[0.0005768, 0.0005990]	[0.0006556, 0.0006864]	[0.0001320, 0.0001414]
6400	[0.0003010, 0.0003082]	[0.0003562, 0.0003642]	[0.0004017, 0.0004099]	[0.0000678, 0.0000715]
12800	[0.0001896, 0.0001926]	[0.0002247, 0.0002281]	[0.0002463, 0.0002497]	[0.0000321, 0.0000333]
	Weibull(shape = **2.5**, scale = **1.2**)			
100	[0.0074203, 0.0109557]	[0.0074423, 0.0109197]	[0.0089512, 0.0120288]	[0.0050578, 0.0078642]
200	[0.0040969, 0.0053811]	[0.0041701, 0.0054379]	[0.0057653, 0.0070727]	[0.0025207, 0.0034193]
400	[0.0024366, 0.0029315]	[0.0025146, 0.0030095]	[0.0037125, 0.0042035]	[0.0013223, 0.0016097]
800	[0.0014398, 0.0016202]	[0.0015138, 0.0016922]	[0.0022567, 0.0024533]	[0.0006447, 0.0007385]
1600	[0.0008857, 0.0009503]	[0.0009490, 0.0010148]	[0.0013584, 0.0014316]	[0.0003263, 0.0003601]
3200	[0.0005849, 0.0006107]	[0.0006301, 0.0006567]	[0.0008614, 0.0008920]	[0.0001866, 0.0001998]
6400	[0.0003564, 0.0003654]	[0.0003902, 0.0003996]	[0.0005110, 0.0005234]	[0.0000960, 0.0001012]
12800	[0.0002270, 0.0002308]	[0.0002503, 0.0002543]	[0.0003057, 0.0003107]	[0.0000454, 0.0000471]
	Weibull(shape = **3**, scale = **1.8**)			
100	[0.0050408, 0.0076472]	[0.0050192, 0.0075828]	[0.0067701, 0.0092139]	[0.0038508, 0.0056722]
200	[0.0026566, 0.0035314]	[0.0026734, 0.0035426]	[0.0043613, 0.0053747]	[0.0019245, 0.0026055]
400	[0.0015611, 0.0018809]	[0.0015651, 0.0018829]	[0.0028106, 0.0031914]	[0.0010076, 0.0012264]
800	[0.0009284, 0.0010468]	[0.0009405, 0.0010571]	[0.0017017, 0.0018563]	[0.0004914, 0.0005622]
1600	[0.0005710, 0.0006156]	[0.0005829, 0.0006275]	[0.0010169, 0.0010751]	[0.0002482, 0.0002738]
3200	[0.0003725, 0.0003907]	[0.0003821, 0.0004003]	[0.0006433, 0.0006677]	[0.0001418, 0.0001516]
6400	[0.0002265, 0.0002329]	[0.0002348, 0.0002412]	[0.0003798, 0.0003898]	[0.0000730, 0.0000769]
12800	[0.0001441, 0.0001467]	[0.0001505, 0.0001531]	[0.0002256, 0.0002296]	[0.0000345, 0.0000358]

A.3.3 Efficiency Results

Table A.11: Efficiency ratios from Weibull distribution simulations. Each section shows results for various Weibull distributions defined by their shape and scale parameters.

n	NW/ITER	KDE/ITER	KDE/NW
Weibull (shape = 1.2, scale = 2)			
100	0.6803	0.8959	1.3168
200	1.1253	1.5755	1.4001
400	1.3454	1.9299	1.4344
800	1.4125	2.1131	1.4960
1600	1.4164	2.2390	1.5808
3200	1.4285	2.3320	1.6326
6400	1.4330	2.4934	1.7400
12800	1.4023	2.5749	1.8361
Weibull (shape = 1.5, scale = 3)			
100	0.8170	0.8752	1.0713
200	1.1250	1.3181	1.1717
400	1.2068	1.4299	1.1848
800	1.2722	1.4926	1.1733
1600	1.2940	1.5199	1.1745
3200	1.3117	1.5240	1.1618
6400	1.3268	1.5797	1.1907
12800	1.3239	1.6088	1.2152
Weibull (shape = 2, scale = 1.5)			
100	0.9496	0.9899	1.0425
200	1.0548	1.2488	1.1839
400	1.0846	1.3590	1.2530
800	1.1197	1.4006	1.2509
1600	1.1548	1.3858	1.2000
3200	1.1672	1.3421	1.1499
6400	1.1825	1.3322	1.1266
12800	1.1847	1.2977	1.0954
Weibull (shape = 2.5, scale = 1.2)			
100	0.9992	1.1417	1.1426
200	1.0137	1.3545	1.3362
400	1.0291	1.4747	1.4330
800	1.0477	1.5392	1.4691
1600	1.0696	1.5196	1.4207
3200	1.0763	1.4665	1.3626
6400	1.0942	1.4331	1.3097
12800	1.1022	1.3464	1.2216
Weibull (shape = 3, scale = 1.8)			
100	0.9932	1.2598	1.2684
200	1.0045	1.5734	1.5663
400	1.0017	1.7438	1.7407

Continued on next page

Table A.11 (continued)

n	NW/ITER	KDE/ITER	KDE/NW
800	1.0113	1.8013	1.7811
1600	1.0201	1.7630	1.7284
3200	1.0252	1.7178	1.6756
6400	1.0361	1.6752	1.6168
12800	1.0440	1.5653	1.4993

A.3.4 Convergence Results

Table A.12: Convergence rates from Weibull distribution simulations. Each section shows results for various Weibull distributions defined by their shape and scale parameters.

Estimator	Empirical Rate	Theoretical Rate	95% CI
Weibull (shape = 1.2, scale = 2)			
Iterative Weights	-0.5917	-0.8000	[-0.6313, -0.5521]
KDE	-0.5033	-0.8000	[-0.5084, -0.4982]
Normal Weights	-0.5426	-0.8000	[-0.5566, -0.5286]
FD	-0.9880	-1.0000	[-1.0278, -0.9482]
Weibull (shape = 1.5, scale = 3)			
Iterative Weights	-0.6325	-0.8000	[-0.6613, -0.6038]
KDE	-0.6034	-0.8000	[-0.6093, -0.5976]
Normal Weights	-0.5925	-0.8000	[-0.6073, -0.5778]
FD	-0.9891	-1.0000	[-1.0265, -0.9517]
Weibull (shape = 2, scale = 1.5)			
Iterative Weights	-0.6643	-0.8000	[-0.7034, -0.6252]
KDE	-0.6676	-0.8000	[-0.6795, -0.6557]
Normal Weights	-0.6369	-0.8000	[-0.6676, -0.6062]
FD	-0.9900	-1.0000	[-1.0255, -0.9545]
Weibull (shape = 2.5, scale = 1.2)			
Iterative Weights	-0.6824	-0.8000	[-0.7310, -0.6338]
KDE	-0.6876	-0.8000	[-0.7037, -0.6715]
Normal Weights	-0.6646	-0.8000	[-0.7103, -0.6190]
FD	-0.9904	-1.0000	[-1.0250, -0.9559]
Weibull (shape = 3, scale = 1.8)			
Iterative Weights	-0.6944	-0.8000	[-0.7519, -0.6369]
KDE	-0.6920	-0.8000	[-0.7103, -0.6737]
Normal Weights	-0.6855	-0.8000	[-0.7426, -0.6284]
FD	-0.9908	-1.0000	[-1.0247, -0.9569]

Bibliography

Akinshin, A. (2024). Quantile-respectful density estimation based on the harrell–davis quantile estimator. *arXiv preprint arXiv:2404.03835*. https://arxiv.org/abs/2404.03835

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics*. Society for Industrial; Applied Mathematics.

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, *3*(1), 39–52. https://doi.org/10.2307/3001536

Bland, J. M., & Altman, D. G. (1996). Transformations, means, and confidence intervals. *BMJ*, *312*, 1079. https://doi.org/10.1136/bmj.312.7038.1079

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243.

Box, G. E. P., & Cox, D. R. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, *77*, 209–210.

Burnecki, K., Kukla, G., & Weron, R. (2000). Property insurance loss distributions. *Physica A: Statistical Mechanics and its Applications*, *287*(1-2), 269–278.

Carroll, R. J., & Ruppert, D. (1981). On prediction and the power transformation family. *Biometrika*, *68*, 609–615.

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

de Boor, C. (1978). *A practical guide to splines*. Springer-Verlag.

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: A roughness penalty approach* (Vol. 58). Chapman; Hall/CRC.

He, X., & Shi, P. (1998). Monotone b-spline smoothing. *Journal of the American Statistical Association*, *93*(442), 643–650.

Hinkley, D. V., & Runger, G. (1984). The analysis of transformed data. *Journal of the American Statistical Association*, *79*, 302–309.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, *4*, 83–91.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation* (2nd ed.). Springer Science & Business Media.

Mächler, M. B. (1996). Density estimation: New spline approaches and a partial review [Manuscript, Seminar für Statistik, ETH Zentrum, Zürich, Switzerland]. https://stat.ethz.ch/Manuscripts/maechler/wp-Density-CS96.pdf

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*(3), 832–837. https://doi.org/10.1214/aoms/1177728190

Sakia, R. M. (1992). The Box-Cox transformation technique: A review. *The Statistician*, *41*, 169–178.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611. https://doi.org/10.2307/2333709

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). Chapman; Hall/CRC. https://doi.org/10.1201/9781315140919

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*(2), 279–281. https://doi.org/10.1214/aoms/1177730256

Stewart, J. (2012). *Calculus: Early transcendentals* (7th). Brooks/Cole, Cengage Learning.

Taylor, J. M. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association*, *81*, 114–118.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, *44*, 1650–1661.

Wahba, G. (1990). *Spline models for observational data* (Vol. 59). Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9781611970128