

ORIGINAL ARTICLE

WILEY

Characterization of a robust probabilistic framework for brain magnetic resonance image data distributions

Abhirup Banerjee^{1,2}  | Sujay Mukhoti³ 

¹Institute of Biomedical Engineering,
Department of Engineering Science, University
of Oxford, Oxfordshire, OX3 7DQ, UK

²Division of Cardiovascular Medicine, Radcliffe
Department of Medicine, University of Oxford,
Oxfordshire, OX3 9DU, UK

³Operations Management and Quantitative
Techniques Area, Indian Institute of
Management Indore, Indore, Madhya Pradesh,
453556, India

Correspondence

Abhirup Banerjee, Institute of Biomedical
Engineering (IBME), Old Road Campus
Research Building, Oxford OX3 7DQ, UK.
Email: abhirup.banerjee@eng.ox.ac.uk

Funding information

A. Banerjee is a Royal Society University
Research Fellow and is supported by the Royal
Society Grant No. URF\R1\221314.

Probabilistic characterisation of image data for accurate prognosis and treatment planning remains a long-standing problem in medical research, especially when the data distribution depicts flat-top and high-order contact. Such flat-top distributions are quite common in brain magnetic resonance (MR) image data, where the density drops sharply beyond the flat interval. Intuitively, it would indicate a bipartition of data into positive region containing observations definitely belonging to the image class and boundary region with observations possibly belonging to it. The flat peak would also imply that multiple values are equally most likely to belong to that class. However, the popular probability distributions used in such cases are unimodal, creating ambiguity about the positive region. In this work, we study the statistical properties and develop likelihood-based iterative estimation method for the parameters of a novel class of platykurtic probability distributions containing normal, called the stomped normal distribution, that provides more accurate modelling to the flat-top data distributions. The robustness of the proposed stomped normal model has been illustrated with six simulated and nine real brain MR volumes. Our analysis shows substantial improvement in explaining a variety of shapes of data distributions using the proposed probability model.

KEYWORDS

Brain MRI, Data generating model, Likelihood-based estimation, Medical image analysis, Stomped normal distribution, Volumetric image analysis

1 | INTRODUCTION

A major problem in medical image data analysis is to devise an appropriate probability model for it. For example, the brain magnetic resonance (MR) image would contain four major parts, namely, cerebro-spinal fluids (CSF), grey matter (GM) tissues, white matter (WM) tissues, and background including skull, scalp, and other substances. The intensity of the signal emitted from these matters during MR imaging are plotted in grey scale to identify different types of objects (or *image classes*). In a three-dimensional image (also called *volumetric image* or simply *volume*), any voxel (3D analogue of pixel) is believed to be equally less likely to belong to a particular image class as it moves away from the corresponding grey-level intensity value on either side. Due to this symmetry, unimodal distributions such as Gaussian and Student's *t* have most popularly been used to model intensity distributions of different classes in image data. These models posit one-way decision structure with an object attaining maximum likelihood of belonging to a class at a single point. In reality though, an object belonging to a class may be depicted by multiple grey-level intensity values. It is, therefore, more appropriate to assume a two-way decision structure with the *positive region* consisting of the data-points that

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Stat* published by John Wiley & Sons Ltd.

definitely belong to the class and the *boundary region* consisting of the data-points that possibly belong to the class. In this paper, we discuss a novel platykurtic family of probability distributions, which contains corresponding symmetric distribution as special cases, for incorporating the two-way decision structure into the image data analysis. It should be noted here that the term 'region' refers to a subset of the support of the feature (e.g., voxel intensity) distribution and not the spatial distribution.

Currently, probabilistic image analysis plays a crucial role in detection and treatment of several medical conditions and diseases. For example, many diagnostic studies require the data generating models for different brain tissue classes, namely, CSF, GM, and WM, for automated segmentation (see Banerjee & Maji, 2019, 2020; Wang et al., 2014). In multiple sclerosis diseases, accurate quantification of WM lesions is widely used for drug treatment assessment (Altman & Petkau, 2005; García-Lorenzo et al., 2013), while in schizophrenia and epilepsy, volumetric analysis of CSF, GM, and WM tissues has been useful to distinguish the morphological differences between subjects (Beheshti et al., 2018; Kim et al., 2017). In cancer treatment, precise delineation of tumour, oedema, and necrotic tissues from MR images enables the radiation therapist to optimize maximal dose to the tumour with minimal radiation to surrounding normal tissues (Fraley et al., 2005; Zou et al., 2004). Modelling of the data generation for image features has also been used for automated cell delineation in histopathological image analysis for the diagnosis of connective tissue diseases (Banerjee & Maji, 2015b, 2016). In a set of related works, Li and Luo (2017) and Frost et al. (2004) used serial brain MR image data to model the progression of Alzheimer's disease.

Popular data generating models for image intensity data are majorly unimodal. Liang et al. (1994) assumed Gaussian models for different image classes and developed an approach for simultaneous estimation of the parameters in each image class and classification of image pixels into one of the classes. As a heavy-tailed alternative to Gaussian model, Banerjee and Maji (2018) considered Student's t distribution along with the hidden Markov random field model of Besag (1986) for modelling spatial distribution of pixels in an image.

Recent literature on medical imaging, on the other hand, reports multi-modality as a key feature of intensity distribution as opposed to the uni-modality (Kawaguchi & Yamashita, 2017). For example, Greenspan et al. (2006) modelled each data cluster by mixture of large number of Gaussian distributions to incorporate multimodality of an image class, while Nguyen et al. (2010) extended the standard Gaussian mixture model where the probability of each class varies over different pixels and is dependent on the neighbouring pixels. Ozenne et al. (2015) used an unsupervised multivariate segmentation algorithm that incorporates spatial information based on four-class Gaussian mixture model for classification of white matter diseased lesion in elderly patients.

A major concern about the above probability models for image classes is that the positive region is limited to the set of modes only, in a strict sense. Any extension of the positive region would, therefore, be subjective and ambiguous. For example, one may consider an interval around the mode as an extended positive region for a unimodal distribution. But the choice of the size of the interval (or coverage) would be subjective and not data driven. In reality, every object lying within the positive region of an image class, specified as a set or interval around the mode, should have the maximum likelihood of belonging to that class. This set or interval of values can all be considered as the modes of the distribution. The likelihood of an object belonging to a class would reduce as the observation deviates away from the positive region of the class.

In this paper, we characterize a probability model of an image class, which incorporates the two-way decision structure. Intuitively, a distribution with a modal set can be constructed from a unimodal distribution by replacing the density of an interval containing the mode with that of a suitable uniform distribution, so that continuity of the probability density function (*pdf*) is maintained. Since the central interval does not have a peak and rather is flat, the new distribution could be thought of as a *stomped* family of distributions. Clearly, stomped distributions with a modal set provide robust alternatives for data distributions ranging from platykurtic to mesokurtic shapes. As an example, the stomped normal distribution could be constructed by replacing the *pdf* of the underlying normal distribution with a suitable uniform *pdf* over an interval, assuming the mode of the normal distribution as its mid-point, and then normalizing the new function. Henceforth, we shall use the notation StN to indicate a stomped normal distribution. This particular family of probability distributions includes the normal distribution as a special case. The highest probability density of an StN distribution extends over an interval around its median, and it decreases while moving away from the interval. The interval of uniform highest probability density around median is used to model the positive region, while the thin tails are used to model the boundary region. Recently, Banerjee and Maji (2015a) explained the effectiveness of such probabilistic bipartition of image class intensities for segmenting medical images. McCormick et al. (2012) also reported a similar phenomenon in the posterior distribution of the probability of an association rule, connecting high cholesterol and hypertension to myocardial infarction. In this paper, we study extensively theoretical properties related to simulation and moment calculation of the stomped normal distribution.

A serious obstacle in using the stomped class of distributions is the inexistence of maximum likelihood estimators (MLEs) due to lack of unimodality. We develop a likelihood-based iterative (LI) method for parameter estimation and describe its asymptotic optimality using simulation in this work. We further compare the LI estimators with its method of moments (MoMs) counter parts using simulation. Finally, we demonstrate the effectiveness of stomped normal distribution in describing image classes taking example from real brain MR volumes, as well as over artificially created brain MR volumes using an MR simulator.

2 | STOMPED NORMAL DISTRIBUTION: CHARACTERIZATION AND PROPERTIES

The aforementioned two-way decision structure for a data distribution closely resembles the definitions of rough set theory. In rough set theory, an approximation space is represented as $\langle U, A \rangle$, where $U \neq \emptyset$ is the universe of discourse and A is an equivalence relation on U . $U/A = \{E_1, \dots, E_m\}$ denotes the quotient set of U by relation A , where E_i is an equivalence class of A . The equivalence classes of A and the empty set \emptyset are the elementary sets in the approximation space $\langle U, A \rangle$. Now, given any arbitrary set $X \in 2^U$, it may in general not be possible to describe X precisely in $\langle U, A \rangle$. Hence, according to the definitions of rough-set theory, X can be characterized by the lower approximation $\underline{A}(X)$, defined as the union of all elementary sets that are subsets of X , and the upper approximation $\bar{A}(X)$, defined as the union of all elementary sets that have a non-empty intersection with X . $B(X) = \bar{A}(X) - \underline{A}(X)$ is called the boundary region of set X .

In analogy to the definitions of rough set theory, in the proposed stomped normal distribution, we define the interval of uniform highest probability density as the lower approximation region or positive region and the interval with gradually decreasing density as the boundary region. Similar to the rough sets, we define both the *lower approximation* and *boundary regions* along the feature (e.g., voxel intensity) distribution and not along the spatial locations. Next, we provide the definition of the stomped normal distribution and present its important properties.

Definition 1. Let $X: \Omega \rightarrow \mathbb{R}$ be a random variable defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X is said to follow one-parameter stomped normal distribution (StN) iff its *pdf* is given as

$$f(x) = \begin{cases} \frac{1}{D} \phi(k), & \text{if } |x| < k \\ \frac{1}{D} \phi(x), & \text{otherwise} \end{cases}$$

where $k > 0$, $D = 2(\bar{\Phi}(k) + k\phi(k))$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the *pdf* and cumulative distribution function (CDF) of standard normal distribution, and $\bar{\Phi}(\cdot)$ denotes the reliability or survival function of standard normal distribution.

We would use $\text{StN}(k)$ as a notation for one-parameter stomped normal distribution. This class of distributions strictly includes the standard normal distribution at $k=0$ (see Azzalini, 1985, for strict inclusion).

It can be observed that the distribution has a modal set $(-k, k)$, in which the density is maximum. In general, the modal set of a stomped distribution would indicate the positive region for the corresponding image class. If an observed data point belongs to the modal set of an image class following stomped distribution (here StN), then it could be assigned to the corresponding image class. In addition, all the data points in the modal set would have the same probability to belong to the image class, measured by their densities. On the other hand, with a unimodal feature distribution, maximizing the data likelihood would point to a single value as a member to the class and all other nearby points will have decreased probability to belong to it. In such a case, defining a positive region would be subjective in nature.

One-parameter stomped normal distribution can be easily generalized to a class of three-parameter probability distribution, which strictly includes normal distribution at a specific parameter setting. The general StN distribution with three parameters is defined as follows:

Definition 2. Let $X: \Omega \rightarrow \mathbb{R}$ be a random variable defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X is said to follow three-parameter StN distribution iff its *pdf* is given as

$$f(x) = \begin{cases} \frac{1}{D\sigma} \phi(k) & \text{if } \left| \frac{x-\mu}{\sigma} \right| < k \\ \frac{1}{D\sigma} \phi\left(\frac{x-\mu}{\sigma}\right), & \text{otherwise} \end{cases}$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $k > 0$ and $D = 2(\bar{\Phi}(k) + k\phi(k))$.

The modal set of the distribution is $(\mu - k\sigma, \mu + k\sigma)$. As in the case of single-parameter stomped normal family, the three-parameter distribution is also strictly inclusive of $N(\mu, \sigma^2)$ at $k=0$. The following theorem provides the expression for the CDF of $\text{StN}(\mu, \sigma, k)$.

Theorem 1. Let X be a real valued random variable defined over $(\Omega, \mathcal{F}, \mathbb{P})$, following the $\text{StN}(\mu, \sigma, k)$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, $k > 0$. The CDF of X is given by

$$F(x) = \begin{cases} \frac{1}{D} \Phi\left(\frac{x-\mu}{\sigma}\right) & \text{if } x \leq \mu - k\sigma \\ \frac{1}{2} + \frac{\phi(k)}{D} \left(\frac{x-\mu}{\sigma}\right) & \text{if } \mu - k\sigma < x < \mu + k\sigma \\ 1 - \frac{1}{D} \Phi\left(\frac{\mu-x}{\sigma}\right) & \text{if } x \geq \mu + k\sigma \end{cases} \quad (1)$$

Proof of the above theorem is provided in Section 1.1 of the Supporting Information. For single-parameter StN distribution, $StN(k)$, the CDF expression reduces to

$$F(x) = \begin{cases} \frac{1}{D}\Phi(x), & \text{if } x \leq -k \\ \frac{1}{2} + \frac{\phi(k)}{D}x, & \text{if } -k < x < k \\ 1 - \frac{1}{D}\Phi(-x), & \text{if } x \geq k \end{cases} \quad (2)$$

It can be easily observed that at $k=0$, the CDF of three-parameter StN distribution reduces to a Gaussian one with mean μ and variance σ^2 . In what follows, we present an important result for the StN family. We show that this family has class preservation property with respect to linear transform; that is, linear transform of an StN random variable yields an StN random variable only.

Theorem 2. If $X \sim StN(\mu, \sigma, k)$, then $a + bX \sim StN(a + b\mu, |b|\sigma, k)$, where $b \neq 0$.

For the proof of the above theorem, see Section 1.2 of the Supporting Information.

2.1 | Simulating from stomped normal distribution

Using the result in Theorem 2, we can simulate from stomped normal distribution using the following inverse transformation. Let u be a random number drawn from $Unif(0,1)$. Then, a single-parameter stomped normal variate x is generated by the inverse of F on u as below:

$$x = F^{-1}(u) = \begin{cases} \Phi^{-1}(Du) & \text{if } u \leq \frac{1}{D}\Phi(-k) \\ \frac{D}{\phi(k)}\left(u - \frac{1}{2}\right) & \text{if } \frac{1}{D}\Phi(-k) < u < 1 - \frac{1}{D}\Phi(-k) \\ \Phi^{-1}(D(1-u)) & \text{if } u \geq 1 - \frac{1}{D}\Phi(-k) \end{cases} \quad (3)$$

To draw from $StN(\mu, \sigma, k)$, we note from Theorem 2 that $Y = \mu + \sigma X \sim StN(\mu, \sigma, k)$. Therefore, by linear transformation of X , a random sample from the three-parameter stomped normal distribution can be obtained.

2.2 | Quantiles

Let us denote the p th quantile of a three-parameter stomped normal distribution by z_p , that is, $z_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}$. Further, we define p^* as $p^* = F(\mu - k\sigma) = \frac{1}{D}\Phi(-k)$. Then it follows that $1 - p^* = F(\mu + k\sigma) = 1 - \frac{1}{D}\Phi(-k)$. Since $p^* < 0.5$, we arrive at the following results.

$$\begin{aligned} \text{For } p \leq p^* : F(z_p) = p &\Rightarrow \frac{1}{D}\Phi\left(\frac{z_p - \mu}{\sigma}\right) = p \Rightarrow z_p = \mu + \sigma\Phi^{-1}(Dp). \\ \text{For } p \geq 1 - p^* : F(z_p) = p &\Rightarrow 1 - \frac{1}{D}\Phi\left(\frac{\mu - z_p}{\sigma}\right) = p \Rightarrow z_p = \mu - \sigma\Phi^{-1}(D(1-p)). \\ \text{For } p^* < p < 1 - p^* : F(z_p) = p &\Rightarrow \frac{1}{2} + \frac{\phi(k)}{D}\left(\frac{z_p - \mu}{\sigma}\right) = p \Rightarrow z_p = \mu + \sigma\frac{D}{\phi(k)}\left(p - \frac{1}{2}\right). \end{aligned} \quad (4)$$

$$\text{Hence, quantile : } z_p = \begin{cases} \mu + \sigma\Phi^{-1}(Dp) & \text{if } p \leq p^* \\ \mu + \sigma\frac{D}{\phi(k)}\left(p - \frac{1}{2}\right) & \text{if } p^* < p < 1 - p^* \\ \mu - \sigma\Phi^{-1}(D(1-p)) & \text{if } p \geq 1 - p^* \end{cases}$$

Special Case: Median = $Q_2 = z_{\frac{1}{2}} = \mu$. Other two quartiles are as follows:

$$Q_1 = z_{\frac{1}{4}} = \begin{cases} \mu + \sigma\Phi^{-1}\left(\frac{D}{4}\right) & \text{if } p^* \geq \frac{1}{4} \\ \mu - \sigma\frac{D}{4\phi(k)} & \text{if } p^* < \frac{1}{4} \end{cases} \quad Q_3 = z_{\frac{3}{4}} = \begin{cases} \mu - \sigma\Phi^{-1}\left(\frac{D}{4}\right) & \text{if } p^* \geq \frac{1}{4} \\ \mu + \sigma\frac{D}{4\phi(k)} & \text{if } p^* < \frac{1}{4} \end{cases} \quad (5)$$

2.3 | Moment generating function (MGF) and moments

Existence of the MGF of stumped normal distribution follows from the fact that it is dominated by a normal distribution. In other words, if $X \sim \text{StN}(\mu, \sigma, k)$ and $Y \sim N(\mu, \sigma^2)$, then $X \leq Y$ a.e. Ω . As the MGF of Y exists, so would the MGF of X and hence its all moments as well. In the following theorem, we provide the expression of its MGF.

Theorem 3. Let X be a real valued random variable defined over $(\Omega, \mathcal{F}, \mathbb{P})$ following stumped normal distribution $\text{StN}(\mu, \sigma, k)$. Then, for any $t \in \mathbb{R}$, the MGF of X is given by

$$M_X(t) = \frac{e^{t\mu}}{D} \left[e^{\frac{t^2\sigma^2}{2}} (\Phi(-k - t\sigma) + \Phi(-k + t\sigma)) + 2\phi(k) \sum_{i=0}^{\infty} \frac{(t\sigma)^{2i} k^{2i+1}}{(2i+1)!} \right]. \quad (6)$$

For proof of the above theorem, see Section 1.3 of the Supporting Information. Differentiating the MGF, we can obtain different raw moments of the stumped normal distribution. Thus,

$$\begin{aligned} E[X] = \frac{\partial}{\partial t} M_X(t) \Big|_{t=0} &= \mu + \frac{1}{\sqrt{2\pi}D} e^{t\mu} \left[e^{\frac{t^2\sigma^2}{2}} t\sigma^2 \sqrt{2\pi} (\Phi(-k - t\sigma) + \Phi(-k + t\sigma)) \right. \\ &\quad \left. + e^{\frac{t^2\sigma^2}{2}} \sqrt{2\pi} (-\sigma\phi(-k - t\sigma) + \sigma\phi(-k + t\sigma)) + 2e^{-\frac{t^2}{2}} \sum_{i=1}^{\infty} \frac{(t\sigma)^{2i-1} k^{2i+1} 2i\sigma}{(2i+1)!} \right] \Big|_{t=0} = \mu. \end{aligned} \quad (7)$$

However, it seems difficult to get higher order moments, especially the central ones, by differentiating the MGF. In what follows, we provide a recursion relation for the central moments of the StN distribution.

Theorem 4. Central moments of an $\text{StN}(\mu, \sigma, k)$ random variable follows the recursion relation:

$$\mu_{p+1} = \begin{cases} p\sigma^2\mu_{p-1} + 2\frac{\phi(k)}{D}\frac{\sigma^{p+1}k^{p+2}}{(p+2)}, & \text{if } p = 2d-1 \\ 0, & \text{if } p = 2d \end{cases} \quad (8)$$

where d is any positive integer and $\mu_0 = 1$.

Detailed proof of the above theorem is provided in Section 1.4 of the Supporting Information. It follows from the above theorem that the variance of $\text{StN}(\mu, \sigma, k)$ is

$$\mu_2 = \sigma^2 + \frac{2\phi(k)}{D}\sigma^2\frac{k^3}{3} = \sigma^2 \left(1 + \frac{2\phi(k)}{D}\frac{k^3}{3} \right) \text{ [assuming } d = 1]. \quad (9)$$

Further, the fourth central moment is given by

$$\mu_4 = 3\sigma^4 + \frac{2\phi(k)}{D}\sigma^4 \left(k^3 + \frac{k^5}{5} \right) = \sigma^4 \left(3 + \frac{2\phi(k)}{D} \left(k^3 + \frac{k^5}{5} \right) \right). \quad (10)$$

As $k \rightarrow 0, \mu_4 \rightarrow 3\sigma^4$, indicating $\text{StN}(\mu, \sigma, k)$ resembles $N(\mu, \sigma^2)$ more and more closely as the truncation parameter approaches zero. In general, solving the recursion in (8), the even order moment can be expressed as

$$\begin{aligned} \mu_{2d} &= (2d-1)\sigma^2\mu_{2d-2} + 2\frac{\phi(k)}{D}\sigma^{2d}\frac{k^{2d+1}}{2d+1} = (2d-1)(2d-3)\sigma^4\mu_{2d-4} + 2\frac{\phi(k)}{D}\sigma^{2d}\left(k^{2d-1} + \frac{k^{2d+1}}{2d+1}\right) \\ &= (2d-1)(2d-3)(2d-5)\sigma^6\mu_{2d-6} + 2\frac{\phi(k)}{D}\sigma^{2d}\left((2d-1)k^{2d-3} + k^{2d-1} + \frac{k^{2d+1}}{2d+1}\right) \\ &= \frac{(2d-1)!}{2^{d-1}(d-1)!}\sigma^{2d}\left(1 + \frac{2\phi(k)}{D}\sum_{p=1}^d \frac{2^p p!}{(2p+1)!} k^{2p+1}\right). \end{aligned} \quad (11)$$

3 | PARAMETER ESTIMATION FOR STOMPED NORMAL DISTRIBUTION

In this section, we discuss the parameter estimation method of stomped normal distribution. The stomped normal density is uniform over the range $(\mu - k\sigma, \mu + k\sigma)$, and hence, there is no unique maxima in that interval, which in turn indicates maximum likelihood estimation of μ is not possible. We rather consider two other methods of estimation. First, we use the MoMs to estimate the parameters μ, σ , and k . The second approach exploits the symmetry to estimate μ iteratively, and the remaining parameters are estimated using an iterative optimization scheme. In both cases, we use simulation to show the accuracy of the estimators.

3.1 | Method of moments estimation

In the MoM estimation, the estimating equations are formed by equating the sample and population moments. Equating the first three non-zero moments, namely, 1^{st} , 2^{nd} , and 4^{th} , to their sample counter parts, we obtain the following three equations:

$$\bar{x} = \mu, \quad (12)$$

$$s^2 = \sigma^2 \left(1 + \frac{2\phi(k)}{D} \frac{k^3}{3} \right), \text{ and} \quad (13)$$

$$m_4 = \sigma^4 \left(3 + \frac{2\phi(k)}{D} \left(k^3 + \frac{k^5}{5} \right) \right), \quad (14)$$

where \bar{x}, s^2 , and m_4 are the sample mean, variance, and 4^{th} central moment, respectively. The parameter μ is estimated from (12), while k and σ are estimated from (13) and (14). After substituting σ^2 from (13) into (14), k is estimated numerically from the implicit equation. Using the estimated value of k in (13), we find the MoM estimator of σ .

The asymptotic unbiasedness of the estimators could be intuitively argued, noting that sample moments converge in L_2 to the population moments. Here, we inspect the bias and mean square errors (MSE) of the parameter estimates for increasing sample sizes to gauge the performance of the MoM estimators. A random sample of size n is simulated from an StN distribution with known μ, σ , and k , where $n \in \{20, 50, 100, 1000, 5000, 10000\}$. The MoM parameter estimates are obtained by solving (12)–(14). For each sample size, this process is repeated $M = 5000$ times to compute bias and MSE of the estimates. In this work, we have considered a fixed location parameter $\mu = 1$ and $\sigma \in \{0.1, 1, 5\}$, $k \in \{0.5, 1, 3\}$ for illustration purpose. Bias and MSE of the MoM estimators of μ, σ , and k are reported in Figure 1 corresponding to different sets of values.

It can be observed from the graphs that the MoM estimators are asymptotically unbiased, but the convergence happens at different rates for different parameters. For example, $\hat{\mu}$ is asymptotically unbiased for almost all combinations, with the most poor convergence occurring at $(\mu, \sigma, k) = (1, 5, 3)$ for both bias and MSE plots, that is, when σ and k are both high in this set up. In a similar manner, the MoM estimators of σ and k converge fast, except when the corresponding true values are high, that is, $\sigma = 5$ and $k = 3$.

3.2 | Likelihood-based iterative parameter estimation

As the MLE of the parameters of the stomped normal distribution does not exist, we suggest here a likelihood-based iterative (LI) procedure for estimating the location parameter μ , scale parameter σ , and truncation parameter k using ordered data, as follows:

Likelihood-based iterative (LI) procedure for parameter estimation

- (I) Initiate the values of μ , σ , and k at $\mu^{(0)}$, $\sigma^{(0)}$, and $k^{(0)}$.
 - (II) At the j^{th} step, standardize sample data with respect to $\mu = \mu^{(j-1)}$ and $\sigma = \sigma^{(j-1)}$.
 - (III) Find $k^{(j)}$ as magnitude of the r^{th} largest standardized sample $|s_{(r)}|$ that maximize the log-likelihood with $\mu = \mu^{(j-1)}$ and $\sigma = \sigma^{(j-1)}$, that is,

$$k^{(j)} = \underset{|s_{(r)}|: r \in \{1, 2, \dots, n\}}{\operatorname{argmax}} \sum_i \log f(x_i; \mu^{(j-1)}, \sigma^{(j-1)}, |s_{(r)}|).$$
 - (IV) Maximize log-likelihood with respect to σ assuming $\mu = \mu^{(j-1)}$ and $k = k^{(j)}$ to obtain $\sigma^{(j)}$.
 - (V) Estimate $\mu^{(j)}$ as the mean of all x_i that lies in the range of $(\mu^{(j-1)} - k^{(j)}\sigma^{(j)}, \mu^{(j-1)} + k^{(j)}\sigma^{(j)})$.
 - (VI) Repeat steps (II)–(V) until consecutive estimates of μ , σ , and k are of not different up to a said level of accuracy.
-

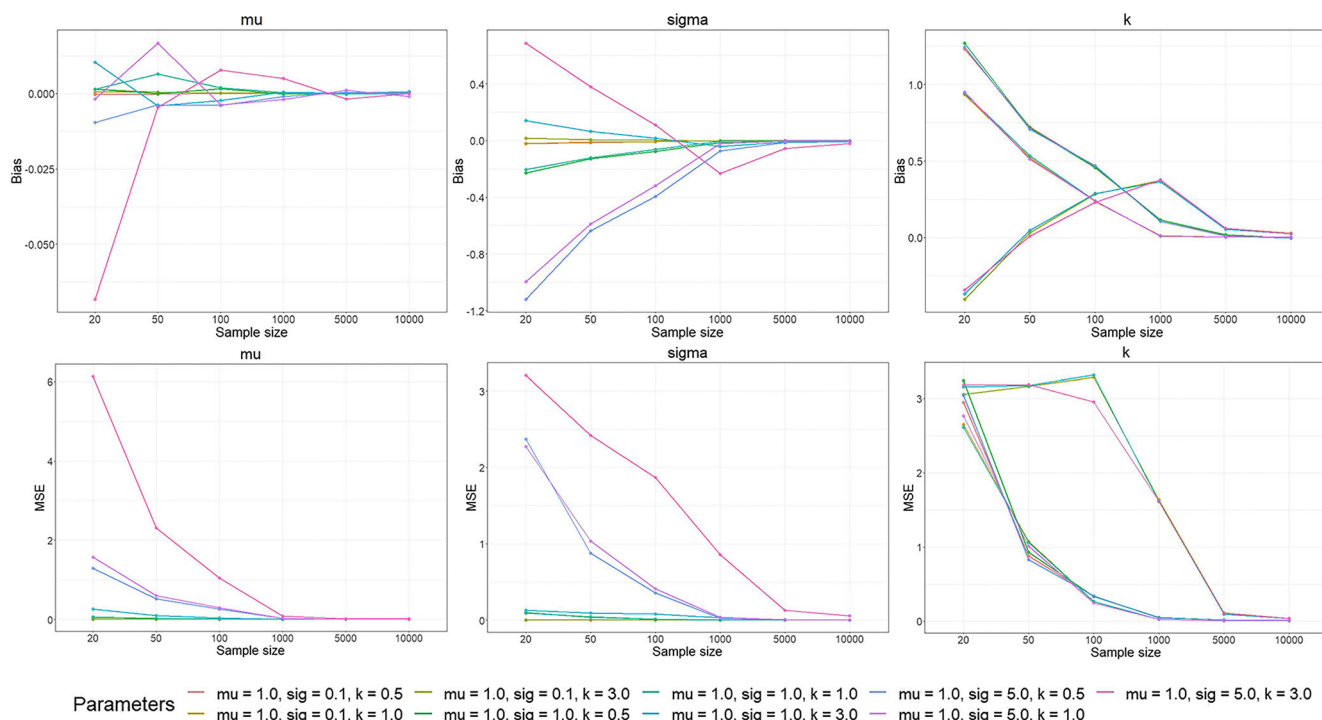


FIGURE 1 Estimation bias and mean-square error (MSE) plots of the MoM estimators of the parameters μ , σ , and k

Henceforth, we shall refer to these estimators as LI estimators (LIE). We indicate the LIE of the parameter θ using $\hat{\theta}$. In order to estimate the truncation parameter k in the LI parameter estimation procedure, we consider the order statistics of the standardized samples that maximize the log-likelihood given the estimates of μ and σ from previous iteration. Since the set of order statistics is finite, the search for truncation parameter k is restricted to the finite number of steps. The LIE of k converges as the order statistic, being the sample quantile, converges in probability to the population quantile. The scale parameter σ , on the other hand, is estimated using the maximum likelihood principle for given k and μ and hence, converges in probability. Finally, since the StN distribution is symmetric about its location parameter μ , we exploit this property in step (V) for estimating μ . The LIE of μ ($\hat{\mu}$) is measured as the α -trimmed mean of the sample where $P(X \leq -k) = \frac{\alpha}{2} = P(X \geq k)$. Bickel (1965) had provided the convergence of $\hat{\mu}$ under some regularity conditions.

Although step (III) provides a good estimator of the truncation parameter k , it can be improved further. The improved estimator of k in this case will lie between $|s_{(r-1)}|$ and $|s_{(r+1)}|$, if $|s_{(r)}|$ is the estimator of k in step (III). Hence, we can estimate k in step (III) as the convex combination of $|s_{(r-1)}|$ and $|s_{(r+1)}|$. In general, the parameters μ , σ , and k are initialized as $\mu = \bar{x}$, $\sigma = \sqrt{s^2}$, and $k = 0$. However, for faster convergence of the iterative algorithm, the parameters can be initialized with their MoM estimators.

The performance of the estimators of (μ, σ, k) , however, is not very straight forward to comment. We rather inspect the unbiasedness and accuracy of the estimators via simulation as in the case of MoM.

The performance of the LIEs are usually better or similar to the corresponding MoM estimators. Although the LIEs of μ , that is, $\hat{\mu}$, are asymptotically unbiased for all simulation studies, both high values of σ and k , that is, $\sigma = 5$ and $k = 3$, cause high bias in small sample size of $n = 20$. For the scale parameter σ , LIEs show large bias in small sample when the true value of σ is high, that is, $\sigma = 5$. Similarly, the LIEs of k , that is, \hat{k} , present high bias for small sample size of $n = 20$ only when the true k is high.

To compare the performances of the LIEs with the MoM estimators, the estimates and standard errors of the parameters have been presented as *estimate \pm standard error* in Tables S1–S9 in Section 2 of the Supporting Information.

It can be observed from Figures 1 and 2 (and Tables S1–S9 of the Supporting Information) that both methods provide estimates of μ close to its true value almost always, except for small sample size of $n = 20$ and high values of both scale and truncation parameters, that is, $\sigma = 5$ and $k = 3$, in which case LIE is closer to the true value than the MoM estimate. In terms of the standard errors, there is not much difference among the two estimators of μ except for high truncation parameter value of $k = 3$. In that case, standard error of MoM estimator is larger than that of LIE in small sample of $n = 20$, although there is not much difference in the larger samples.

In case of estimation of the scale parameter σ , LIEs are much closer to the true value than the MoM estimates across all sample sizes when the truncation parameter is small, that is, $k = 0.5, 1$. For large value of the truncation parameter, that is, $k = 3$, LIEs are closer to the true σ in large samples, that is, $n \geq 1000$ in our example, than the MoM estimates, whereas the opposite happens in small samples where $n \leq 100$. In terms of

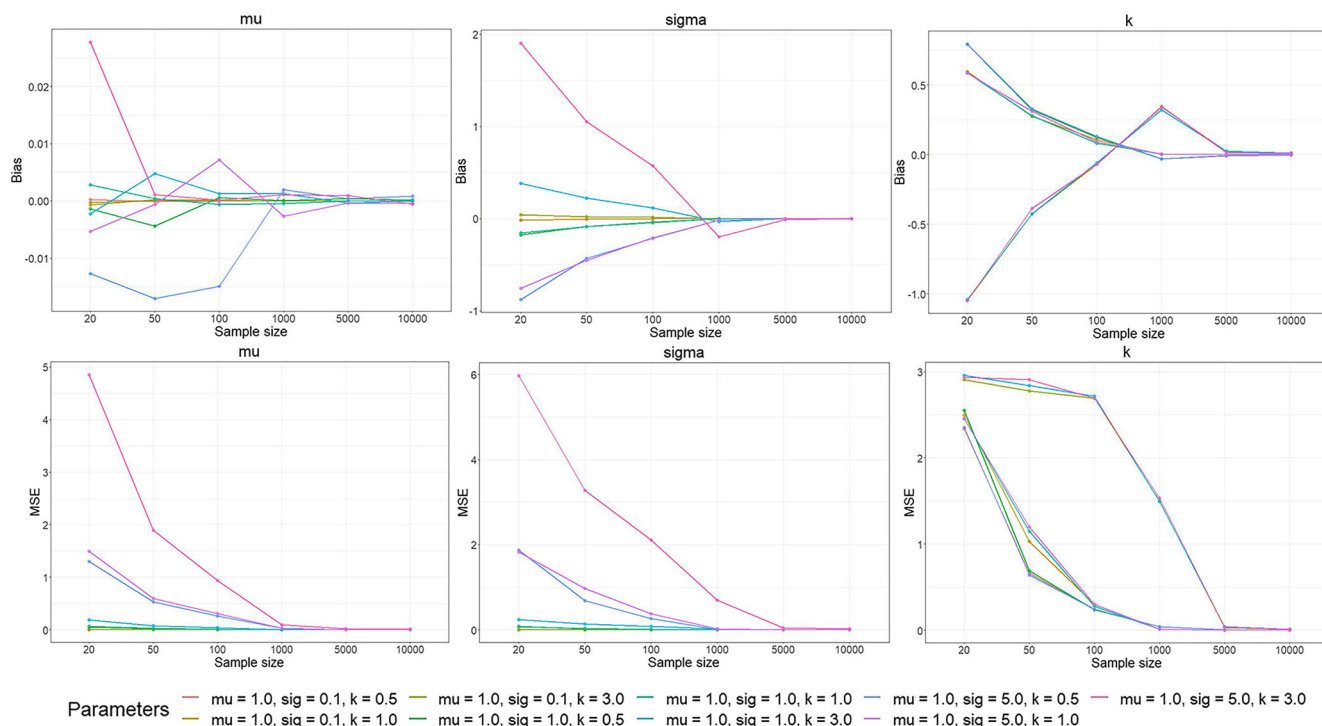


FIGURE 2 Estimation bias and mean-square error (MSE) plots of LIEs of μ , σ and k

standard errors, LI estimates of σ are better than the MoM estimates in large samples where $n \geq 1000$. In fact, for high truncation parameter of $k = 3$, LIE of σ outperforms the MoM estimator in small samples as well.

For estimating the truncation parameter k , the simulation study results show that the LIEs are closer to the true k than their MoM counterparts for all sample sizes when the true truncation parameter is small, that is, $k = 0.5, 1$. For large k , that is, $k = 3$ in our study, LIEs are closer to the true parameter value asymptotically, whereas MoM estimates outperform them in smaller samples where $n \leq 50$. The standard errors of the estimates of k show that the LIEs are asymptotically superior to the MoM estimates irrespective of the true choice of the parameters. Indeed, for large truncation parameter value of $k = 3$, LIEs outperform the MoM estimators for all sample sizes.

From the above discussion, it can be commented that LIEs outperform the MoM estimators at least asymptotically. Indeed, for large truncation parameter, LIE is better than MoM estimator both in terms of estimation bias and standard error. A crucial observation that can be made from here is that the value of the truncation parameter k governs the performance of the estimators.

4 | SIMULATION STUDY

In order to demonstrate that the StN distribution is a generalization of the Gaussian distribution and make a comparison between the two fits, we perform the following simulation study. We first draw a random sample of size $N = 50,000$ from three Gaussian distributions with parameters $(\mu, \sigma) = (10, 2)$, $(-10, 2)$, and $(10, 4)$. The generated data distributions are presented as histograms in the top row of Figure 3. We then estimate the parameters from the three simulated data distributions using both Gaussian and StN distributions. The estimated MLE parameters of the Gaussian and the LIEs of the StN are presented in the left panel of Table 1. From the results presented in Table 1, we can clearly see that the StN distribution can optimally estimate the truncation parameter k as 0 in each case, thus providing optimal modelling of the data generated from Gaussian distribution. Specifically for $N(10, 2)$ and $N(10, 4)$, the parameter k is estimated exactly as 0, while for $N(-10, 2)$ the k is estimated very close to 0 ($k = 0.03$). For parameters μ and σ , the MLEs and the LIEs for the Gaussian and StN respectively achieve the similar values for all three cases.

Next, in order to demonstrate the performance over small sample sizes, we draw random samples of sizes $N = 20, 50$, and 100 from the Gaussian distribution with parameters $(\mu, \sigma) = (10, 2)$. As visible in Figure 4, the empirical distribution in small samples appears to be flat-top due to lack of information. In all these cases, the stumped normal distribution provides better fit to the data with lower Kolmogorov–Smirnov (KS) and χ^2 distances as compared to the Gaussian distribution (see Table 2). We can also observe that the optimally fitted Gaussian and StN distributions

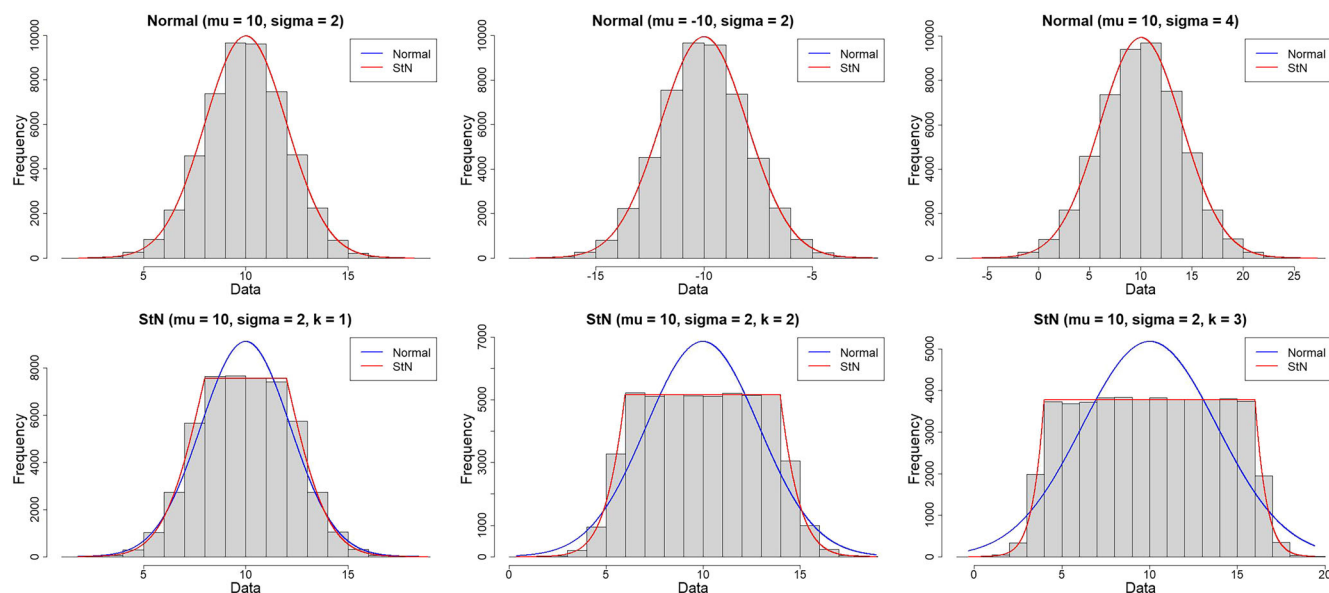


FIGURE 3 The data generating model fitting of simulated data from normal (top) and stomped normal (bottom) distributions. In each case, the optimally fitted normal and StN (based on LIE) distributions are presented in blue and red colors, respectively, over the data distribution

TABLE 1 Parameter estimates of the fitted Gaussian and stomped normal distributions for random samples drawn from the Gaussian and StN distributions with varying parameters

Gaussian distribution			Stomped normal distribution		
Parameters (μ, σ)	Gaussian MLE ^a	StN LIE ^b	Parameters (μ, σ, k)	Gaussian MLE ^a	StN LIE ^b
(10, 2)	(10.003, 2.002)	(10.003, 2.002, 0)	(10, 2, 1)	(9.994, 2.185)	(9.994, 1.993, 1.002)
(-10, 2)	(-10.008, 2.005)	(-10.007, 2.004, 0.032)	(10, 2, 2)	(9.986, 2.909)	(9.986, 2.010, 1.992)
(10, 4)	(10.028, 4.019)	(10.032, 4.019, 0)	(10, 2, 3)	(10.008, 3.853)	(10.008, 1.991, 3.019)

^aValues represent estimates of (μ, σ).

^bvalues represent estimates of (μ, σ, k).

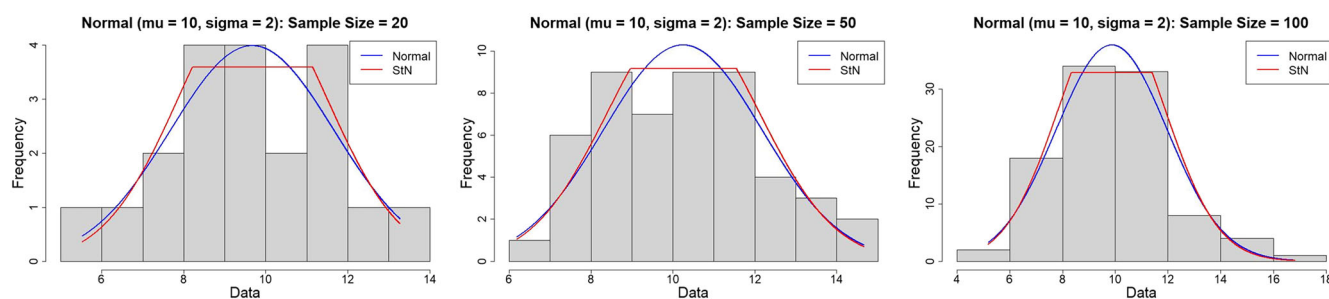


FIGURE 4 The data generating model fitting of simulated data of sizes 20, 50 and 100 (left to right) from normal ($\mu = 10, \sigma = 2$) distribution. In each case, the optimally fitted normal and StN distributions are presented in blue and red colors, respectively

are closely overlapped on each other for larger sample size ($N = 100$) and the estimates of μ and σ remain very close as shown in Table 2. We can, therefore, argue that the general family of StN distributions adopts to the data in a more robust manner even in the smaller samples.

Finally, we draw a random sample of size $N = 50,000$ from three StN distributions with the same location and scale parameters (μ, σ) = (10, 2), but varying truncation parameter $k = 1, 2, 3$. The generated data distributions are presented as histograms in the bottom row of Figure 3. Similar to the previous analysis, here also we estimate the parameters from the simulated data distributions using both Gaussian and StN distributions. From the results presented in the right panel of Table 1, we find that the LI estimators of the StN distribution can accurately estimate the

TABLE 2 Parameter estimates and the Kolmogorov–Smirnov (KS) and χ^2 -distances of the fitted Gaussian and stomped normal distributions for different random samples drawn from the Gaussian ($\mu = 10, \sigma = 2$) distribution

Parameters (μ, σ)	Sample size	Gaussian MLE			Stomped normal LIE		
		Parameters ^a	KS-dist	χ^2 -dist	Parameters ^b	KS-dist	χ^2 -dist
(10, 2)	20	(9.670, 2.001)	0.0858	1.952	(9.679, 1.812, 0.808)	0.0778	1.810
	50	(10.243, 1.941)	0.0627	2.069	(10.258, 1.853, 0.699)	0.0574	1.633
	100	(9.870, 2.118)	0.0788	7.382	(9.870, 2.017, 0.762)	0.0649	6.406

^aValues represent estimates of (μ, σ).^bvalues represent estimates of (μ, σ, k).

three parameters in each case. The performance of the LIE for truncation parameter k slightly decreases with increasing k ; the LIEs for μ and σ perform consistently well in all three cases. On the other hand, the Gaussian distribution aims to model the *positive regions* of the StN data distributions by increasing the variance parameter and thus overestimates σ . The estimation bias increases with increasing k of the original data distribution, while the estimation of μ remains consistently well in all three cases. From the results presented in the bottom row of Figure 3, we see that the optimally fitted StN distributions perfectly represent the underlying data histograms in each case, while the Gaussian distribution fails to model the flat-tops of the data distributions.

5 | BRAIN MR IMAGE ANALYSIS

In order to demonstrate the performance of stomped normal distribution for modelling voxel intensities from different image classes in a volume, specifically the brain MR image classes, we have incorporated two brain MR image databases in our study, namely, the ‘BrainWeb: Simulated Brain Database’ (<https://www.bic.mni.mcgill.ca/brainweb/>) (Collins et al., 1998; Kwan et al., 1999) and the ‘IBSR: Internet Brain Segmentation Repository’ (<https://www.nitrc.org/projects/ibsr/>) (Rohlfing, 2012). We first describe the image classes and the data extracted from these classes for model fitting in the following subsection.

5.1 | Data description

In this paper, we consider simulated three-dimensional (3D) T1 weighted volumes. In case of T1-weighted volume, cerebro-spinal fluid (CSF) is indicated as the darkest grey-level shade and white matter (WM) is shaded the lightest. The volumes from the Brainweb database have been generated by perturbing pre-classified or ground-truth 3D images. Apart from random signal noise, the artefact that poses serious problem in brain MR image analysis is the voxel inhomogeneity in the same tissue, primarily caused by the non-uniformity of magnetic field of the MRI coil. It is also known as the intensity nonuniformity (INU) artefact (Banerjee & Maji, 2013a, 2013b, 2013c). These two types of noises have been used for perturbation of the simulated volumes using an MRI simulator with different noise levels. The simulated images have been generated by adding Rayleigh noise to the background and Rician noise to the image classes or signal region in the ground-truth volume. The percent-noise denotes the percent ratio of the standard deviation of additive Gaussian white noise over the signal for WM tissue. For 100p% level of INU artefact, the multiplicative INU field varies within the range of $(1 - \frac{p}{2})$ to $(1 + \frac{p}{2})$ over brain region. The anatomical model and hence all simulated brain MR volumes are defined at 1-mm isotropic voxel grid in the atlas of the human brain in Talairach space, with dimensions $181 \times 217 \times 181$ ($X \times Y \times Z$). Two example simulated T1-weighted brain MR volumes, with noise 3% INU 40% and with noise 9% INU 20%, from the Brainweb database are presented in the two rows of Figure 5. A common practice in analysing image data is to consider the logarithmic transformation so that the heavy-tails can be tackled easily. In our case, the logarithmic transformation creates serious asymmetry in the data distribution (see Section 3 of the Supporting Information), and hence, we have analysed the unchanged intensity data.

Images in columns 1, 3, and 5 in Figure 5 represent the simulated brain MR volumes without any image class identification and those in columns 2, 4, and 6 represent the ground-truth segmentations with white areas indicating WM tissues, light grey areas indicating grey matter (GM) tissues, dark grey shade indicating CSF, and the black shade indicating the background. In the 1st row, the brain MR volume contains signal noise of 3% and INU 40%, whereas the volume in the 2nd row contains 9% signal noise and 20% INU. From left to right in each example, the first two images represent the top or mid-axial, next two the frontal or mid-coronal, and the last two represent the side-wise or mid-sagittal slices for the same brain MR volume.

The IBSR database consists of T1-weighted brain MR volumes from 20 different subjects. For each subject, the volumetric images are positionally normalized into the Talairach orientation with rotation only. All volumes are of size $256 \times 128 \times 256$ ($X \times Y \times Z$). The manual

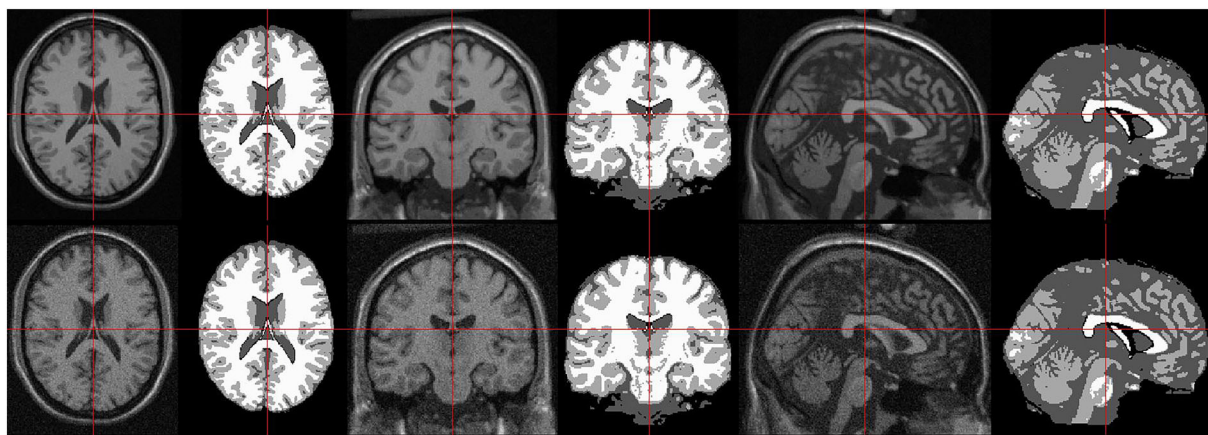


FIGURE 5 Two simulated T1-weighted brain MR volumes from the Brainweb database, with noise 3% INU 40% (top) and with noise 9% INU 20% (bottom). Left to right: mid-axial, mid-coronal, and mid-sagittal slices in columns 1, 3, and 5; corresponding ground-truth segmentations in columns 2, 4, and 6

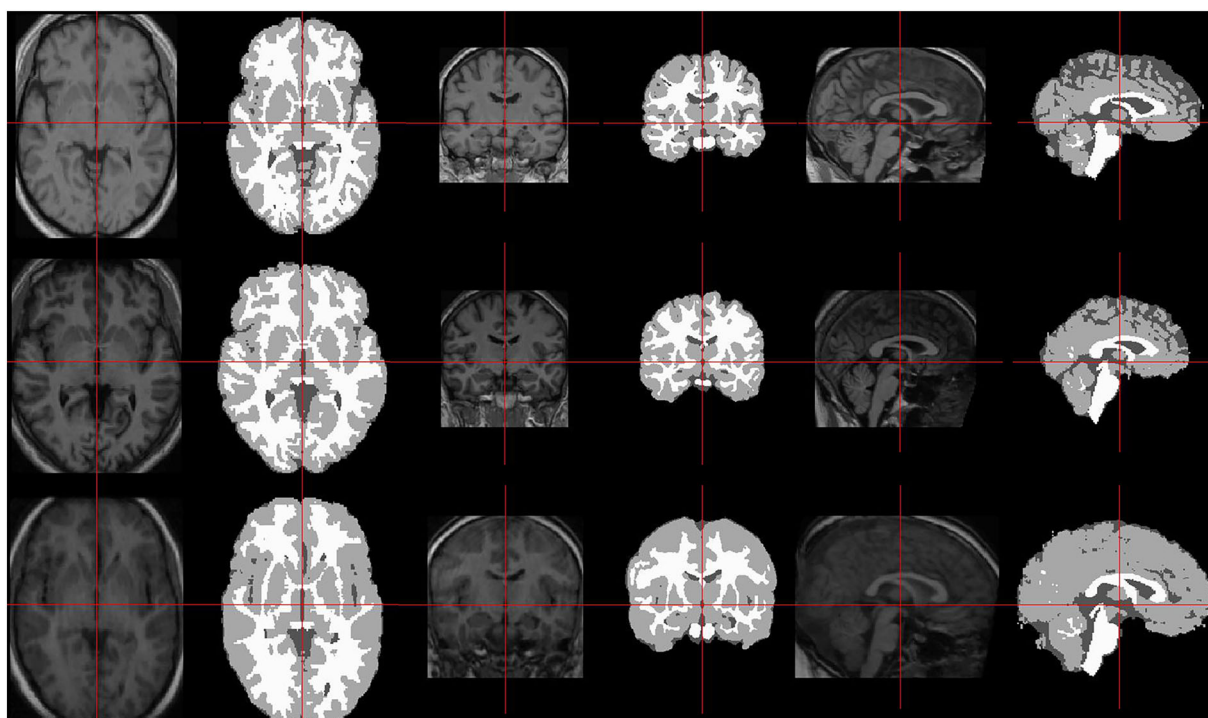


FIGURE 6 Three T1-weighted brain MR volumes from the IBSR database corresponding to three subjects 6, 9, and 18 (top to bottom). Left to right: mid-axial, mid-coronal, and mid-sagittal slices in columns 1, 3, and 5; corresponding ground-truth segmentations in columns 2, 4, and 6

segmentation for each of the 20 volumes is generated by an expert supervisor. Similar to the Brainweb database, we extract the tissue classes from brain MR volumes based on the manual segmentation. From the total of 20 volumes in the database, we select nine for our analysis, with three volumes for each of the three brain MR classes. Three example T1-weighted brain MR volumes from the IBSR database are presented in three rows of Figure 6 (top to bottom). We present the T1-weighted mid-axial, mid-coronal, and mid-sagittal views of three real brain MR volumes in columns 1, 3, and 5, while the ground-truth segmentations are presented in columns 2, 4, and 6 in the same figure. The three rows of Figure 6 represent the volumes from three subjects 6, 9, and 18.

In the following sections, we describe the efficacy of stomped normal distribution in modelling both simulated and real brain MR image volumes. We compare the performance of the proposed StN model with the Gaussian distribution using the KS distance and χ^2 -distance (Ahonen et al., 2006), for measuring the distance between the original data distribution and the fitted probability distributions.

5.2 | Analysis of simulated brain MR image data

We extract the brain MR classes based on the anatomical model, that is, the ground-truth segmentation. The figures in Figure 7 show the intensity distributions corresponding to six different simulated brain MR volumes of the Brainweb database, along with the fitted Gaussian and stomped normal densities. For each Brainweb volumes with different noise and INU levels, we estimate the parameters of the fitted normal and StN distributions over the voxel intensity data from extracted tissue classes. Finally, we use the KS distance and χ^2 -distance to measure the performance of the Gaussian and StN distributions in modelling the data distributions. In Table 3, we show the MLE and LIE for the parameters of the two distributions along with their respective KS and χ^2 -distances for goodness of fit.

Each of the three classes, CSF, GM, and WM, emits separate frequencies and hence, should have a different location. Table 3 shows that the intensity distribution of the three classes are equally well identified by both models having same estimate of location parameter μ .

For the scale parameter, however, the estimates vary over the two models. Increase in INU level with fixed noise level causes inflated estimation of the scale parameter (σ) (see CSF parameter estimates) for both Gaussian and stomped normal models. Further, the estimate of the truncation parameter (k) decreases with the increase in noise in StN model. It may be remarked here that both normal and StN models explain the data well, with KS and χ^2 distances being comparatively lower in case of StN, indicating a better fit.

For the grey matter data, INU level is fixed at 40% but the noise varies. In this case also, we observe increase in estimate of σ for both normal and stomped normal models with higher magnitude compared to the CSF cases. The increase in noise level from 3% to 5% results in reduced

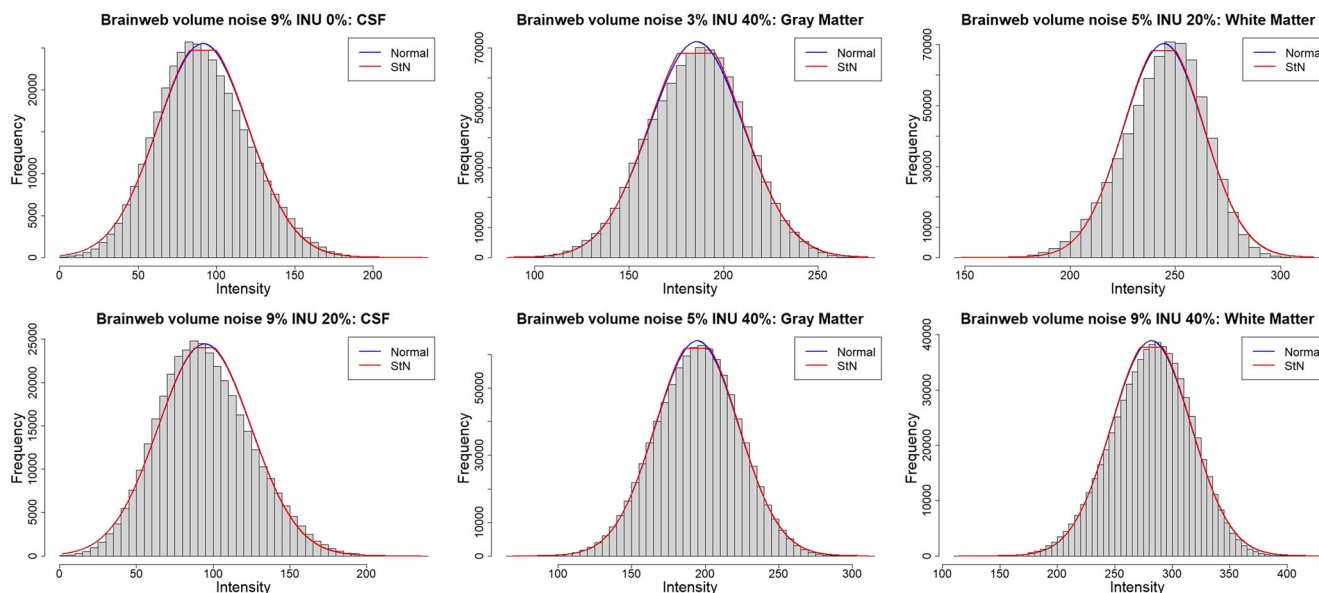


FIGURE 7 The data generating model fitting of image classes in six different Brainweb simulated MR volumes. Left to right: CSF, grey matter, and white matter. In each case, the optimally fitted normal and StN (based on LIE) distributions are presented in blue and red colors, respectively, over the intensity distribution

TABLE 3 Parameter estimates and the Kolmogorov–Smirnov (KS) and χ^2 -distances for the fitted Gaussian and stomped normal distributions in three tissue classes from six different simulated brain MR volumes of Brainweb database

Volume Noise% INU%	Tissue class	Number of voxels	Gaussian MLE			Stomped normal LIE		
			Parameters ^a	KS-dist	χ^2 -dist	Parameters ^b	KS-dist	χ^2 -dist
N-9% INU 0%	CSF	371945	(91.440, 29.074)	0.0233	2328.62	(91.440, 28.983, 0.291)	0.0229	2315.16
N-9% INU 20%		371945	(94.545, 30.358)	0.0238	2542.27	(94.545, 30.318, 0.216)	0.0238	2537.50
N-3% INU 40%	GM	902912	(185.736, 25.033)	0.0119	2189.79	(185.736, 24.848, 0.393)	0.0108	2046.35
N-5% INU 40%		902912	(194.567, 28.129)	0.0084	1141.95	(194.567, 28.024, 0.309)	0.0075	1077.79
N-5% INU 20%	WM	674777	(244.497, 19.139)	0.0317	8828.09	(244.497, 19.071, 0.305)	0.0314	8772.71
N-9% INU 40%		674777	(281.862, 34.624)	0.0167	2909.13	(281.862, 34.524, 0.284)	0.0165	2881.58

^aValues represent estimates of (μ, σ) .

^bvalues represent estimates of (μ, σ, k) .

estimate of the truncation parameter in StN model. Similar to the CSF cases, both normal and StN distributions fit the data well; but the discrepancy between the data and the StN distribution is less in terms of both KS and χ^2 measures, indicating a better fit.

Increasing both noise (5% \rightarrow 9%) and INU level (20% \rightarrow 40%) in white matter data, we observe more than 75% increase in estimated scale parameter σ for both Gaussian and StN models. Under the StN model, these increments cause reduction in truncation parameter estimates. As in the previous two cases, here also we observe that both normal and StN distributions fit the data well, with stomped normal distribution having smaller KS and χ^2 distances.

From the above discussion, we can also observe that the StN is at least as good as the normal model, having more robustness in fitting data with different types of peaks ranging from platykurtic to mesokurtic distributions. Indeed, for the grey matter data, where the peaks are quite flat, StN distribution provides better fitting than the popular choice of normal distribution.

5.3 | Analyzing real brain MR images from IBSR database

The nine subfigures in Figure 8 present the brain MR tissue intensity distributions corresponding to nine different healthy subjects from the IBSR database, along with the fitted Gaussian and stomped normal models. The first column of the figure corresponds to the CSF from the three volumes from subjects 4, 5, and 6. Similarly, the second column represents the volumetric GM tissue data from subjects 9, 10, and 11, while the third column presents the volumetric WM tissue data from subjects 16, 17, and 18. The comparative performance of StN model with Gaussian model in explaining the peakedness of the real data is measured by KS and χ^2 -distances as explained in the previous section. The parameter estimates are presented in Table 4, along with the KS and χ^2 -distances at the estimated parameter values.

For the CSF data, volume from subject no. 6 depicts substantial improvement achieved by StN model with a 51% gain in KS distance and 71% in χ^2 -distance over the Gaussian counterpart. In this case, the estimates of μ and σ from the Gaussian models are 46.36 and 17.01, respectively, whereas the same from StN model are 46.36 and 9.17. Since stomped normal model shrinks the variance from the Gaussian

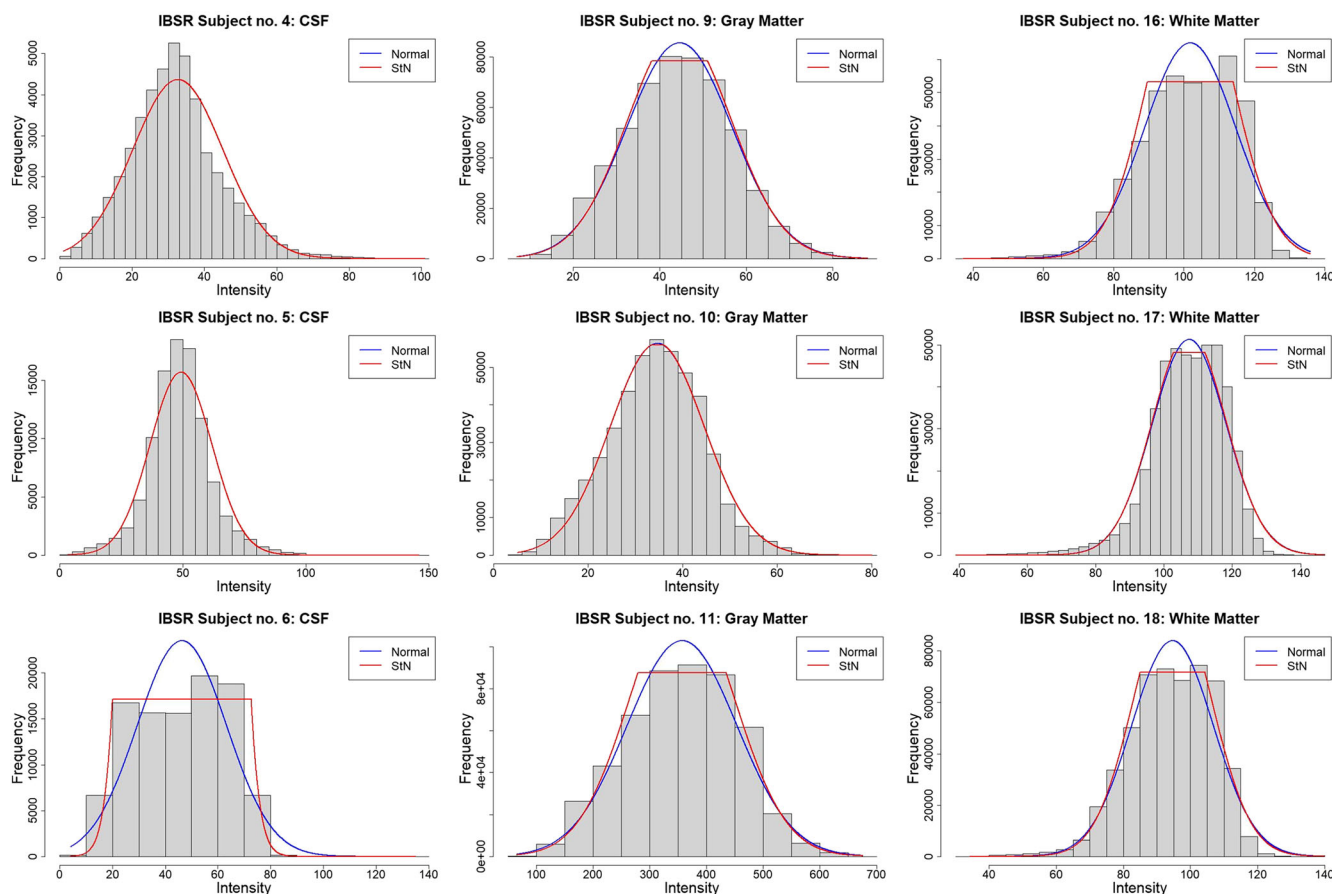


FIGURE 8 The data generating model fitting of image classes in IBSR brain MR volumes from nine different healthy subjects. Left to right: CSF, grey matter, and white matter. In each case, the optimally fitted normal and StN (based on LIE) distributions are presented in blue and red colors, respectively, over the intensity distribution

TABLE 4 Parameter estimates and the Kolmogorov–Smirnov (KS) and χ^2 -distances for the fitted Gaussian and stomped normal distributions in three tissue classes from brain MR volumes of nine different subjects in the IBSR database

Subject no.	Tissue class	Number of voxels	Gaussian MLE			Stomped normal LIE		
			Parameters ^a	KS-dist	χ^2 -dist	Parameters ^b	KS-dist	χ^2 -dist
4	CSF	45786	(32.645, 12.551)	0.0649	1088.39	(32.645, 12.551, 0.028)	0.0649	1088.39
5		98948	(49.305, 12.585)	0.0598	3479.27	(49.305, 12.585, 0.024)	0.0598	3479.06
6		100181	(46.358, 17.005)	0.0813	12533.25	(46.358, 9.175, 2.873)	0.0399	3736.18
9	GM	523709	(44.596, 12.206)	0.0272	4515.35	(44.596, 11.982, 0.534)	0.0237	4288.88
10		468438	(34.622, 9.945)	0.0342	5609.69	(34.622, 9.940, 0.139)	0.0344	5620.12
11		500011	(357.255, 96.947)	0.0288	8123.81	(357.255, 91.302, 0.852)	0.0189	6287.81
16	WM	422847	(101.796, 12.997)	0.0651	18162.53	(101.796, 11.849, 1.030)	0.0386	12845.26
17		464782	(107.397, 10.847)	0.0534	22181.21	(107.397, 10.749, 0.428)	0.0481	21349.95
18		514781	(94.693, 12.230)	0.0582	18802.09	(94.693, 11.561, 0.838)	0.0367	14060.77

^aValues represent estimates of (μ, σ) .^bvalues represent estimates of (μ, σ, k) .

counterpart, chances of generating outlier observations are also reduced in this case. Additionally, the truncation parameter is estimated to be 2.87, resulting in the positive zone as (20,72.72) with a coverage of 0.9009.

For fitting the GM intensity data, both StN and Gaussian models provide comparable fit in all three volumes, with an improved fitting achieved by the StN for volume of subject 11. The KS and χ^2 -distances under the StN model improve by 34% and 22%, respectively, compared to the Gaussian. The positive zone of the StN distribution, in this case, is (279.46,435.04) with coverage 0.5454.

With the white matter intensity data, stomped normal model clearly outperforms the popular Gaussian model in volumes for all three subjects 16, 17, and 18. In volumes of subjects 16 and 18, the flat-top distributions are better described by the StN models with nearly 37% and 41% improvement in KS distances respectively and the same figures are observed to be 29% and 25% for χ^2 -distances. The positive zone for volume of subject 16 is estimated to be (89.59,114.0), whereas the same for subject 18 is (84.98,104.4). The corresponding coverage probabilities are 0.6147 and 0.5393.

6 | DISCUSSION

This paper contributes to the existing literature from two significant aspects. First, we have characterized a new probability model for MRI data from rough-set theory concepts well known in the pattern recognition literature. The stomped family of platykurtic distributions described in this paper incorporates the two-way decision structure using a modal set as the positive region. Any observation in the positive zone of a class could be definitely identified as its member with equal likelihood, as opposed to the unimodal case where the decision is subjected to the interpretation of closeness to the unique mode. We have discussed in detail the major statistical properties of the stomped normal distribution including its moments and quantiles.

Second, we have developed an iterative method for parameter estimation, due to the lack of ML estimation for the proposed new family of probability distributions. For this purpose, we have performed an extensive simulation study over nine different parameter combinations and six different sample sizes. We have elaborated on the parameter estimation methods for StN distribution, which can be used, in general, for any stomped distribution.

Finally, we have demonstrated the appropriateness of StN distribution over the popular choice of Gaussian distribution through the analysis of simulated as well as real 3D MR images of human brains. Indeed, the stomped normal family provides robust alternative to the Gaussian model (see Section 5.3) as it fits well to data distributions of any shape between flat-top (platykurtic, e.g., Figure 8, bottom-right) to single peak (mesokurtic, e.g., Figure 8, top-left). The coverage probabilities of the positive region estimated under StN distribution show that a large chunk of the data can be identified as positive region for a class, which remains ambiguous for unimodal models.

The present work has several possibilities of theoretical extensions and applications. There can be real-life scenarios where heavy-tail distributions would be required to model a data distribution with extreme values and a “definitely belongs to” zone. In such cases, heavy-tailed stomped class of distributions, for example, stomped-t, can be studied (Banerjee & Maji, 2017). Multivariate extensions of the stomped distributions would be useful for different image and biological data analysis, including medical image analysis and interpretation.

Application wise, the present work not only shows the goodness of fit of stomped normal distribution, but also leads to an interesting classification problem. Any unidentified object in an image could be easily classified to an image class under the stomped normal model in case of

minimum overlap between the data distributions. In case of overlapping data distributions between two image classes where the pixels are very differently spatially aligned, the classification problem can be solved with the application of a finite mixture of the stomped class of distributions (Banerjee & Maji, 2015a). Additionally, in order to model the spatial structure in the image, stochastic models such as hidden Markov random field (Besag, 1986) can be used.

Although we have demonstrated the application of StN distribution only for MR image data modelling, the framework can be useful for modelling image classes in many other fields of study as well. For example, in machine learning, robotics, cognitive science, and artificial intelligence, where data generating models play a critical role (Ghahramani, 2015; Wang & Blei, 2019), the stomped family of distributions could be very useful.

ACKNOWLEDGEMENTS

The authors are thankful to the Reviewers and the Editors for their insightful comments, which helped to improve the quality and readability of the manuscript.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

AUTHOR CONTRIBUTIONS

Abhirup Banerjee: Conceptualization, Data curation, Methodology, Software, Visualization, Analysis, Writing-Original draft preparation, Writing-Reviewing and Editing. **Sujay K. Mukhoti:** Conceptualization, Methodology, Visualization, Analysis, Writing-Original draft preparation, Writing-Reviewing and Editing.

SUPPORTING INFORMATION

Supporting Information, containing detailed proofs of the theorems and lemmas and additional quantitative results, is available as part of the online article.

DATA AVAILABILITY STATEMENT

The implementations of the statistical properties and the parameter estimations of the stomped normal distribution are available as a R package, named “StoNproperties”, at <https://github.com/abhirupb88/StoNproperties>.

ORCID

Abhirup Banerjee  <https://orcid.org/0000-0001-8198-5128>

Sujay Mukhoti  <https://orcid.org/0000-0002-8089-4248>

REFERENCES

- Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037–2041.
- Altman, R. M., & Petkau, A. J. (2005). Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine*, 24(15), 2335–2344.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.
- Banerjee, A., & Maji, P. (2013a). Contraharmonic mean based bias field correction in MR images. *Computer Analysis of Images and Patterns: Springer Berlin Heidelberg*, pp. 523–530.
- Banerjee, A., & Maji, P. (2013b). Rough set based homogeneous unsharp masking for bias field correction in MRI. *Image Analysis and Processing – ICIAP 2013* (pp. 542–551). Springer Berlin Heidelberg.
- Banerjee, A., & Maji, P. (2013c). Rough sets for bias field correction in MR images using contraharmonic mean and quantitative index. *IEEE Transactions on Medical Imaging*, 32(11), 2140–2151.
- Banerjee, A., & Maji, P. (2015a). Rough sets and stomped normal distribution for simultaneous segmentation and bias field correction in brain MR images. *IEEE Transactions on Image Processing*, 24(12), 5764–5776.
- Banerjee, A., & Maji, P. (2015b). Rough sets for finite mixture model based HEp-2 cell segmentation. *Rough Sets and Knowledge Technology* (pp. 459–469). Springer International Publishing.
- Banerjee, A., & Maji, P. (2016). Rough-probabilistic clustering and hidden Markov random field model for segmentation of HEp-2 cell and brain MR images. *Applied Soft Computing*, 46, 558–576.
- Banerjee, A., & Maji, P. (2017). Stomped-t: A novel probability distribution for rough-probabilistic clustering. *Information Sciences*, 421, 104–125.
- Banerjee, A., & Maji, P. (2018). Spatially constrained Student's t-distribution based mixture model for robust image segmentation. *Journal of Mathematical Imaging and Vision*, 60(3), 355–381.
- Banerjee, A., & Maji, P. (2019). Segmentation of bias field induced brain MR images using rough sets and stomped-t distribution. *Information Sciences*, 504, 520–545.

- Banerjee, A., & Maji, P. (2020). A spatially constrained probabilistic model for robust image segmentation. *IEEE Transactions on Image Processing*, 29, 4898–4910.
- Beheshti, I., Sone, D., Farokhian, F., Maikusa, N., & Matsuda, H. (2018). Gray matter and white matter abnormalities in temporal lobe epilepsy patients with and without hippocampal sclerosis. *Frontiers in Neurology*, 9, 107.
- Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48(3), 259–302.
- Bickel, P. J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics*, 847–858.
- Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., & Evans, A. C. (1998). Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3), 463–468.
- Fraley, C., Raftery, A., & Wehrens, R. (2005). Incremental model-based clustering for large datasets with small clusters. *Journal of Computational and Graphical Statistics*, 14(3), 529–546.
- Frost, C., Kenward, M. G., & Fox, N. C. (2004). The analysis of repeated 'direct' measures of change illustrated with an application in longitudinal imaging. *Statistics in Medicine*, 23(21), 3275–3286.
- García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., & Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis*, 17(1), 1–18.
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553), 452–459.
- Greenspan, H., Ruf, A., & Goldberger, J. (2006). Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Transactions on Medical Imaging*, 25(9), 1233–1245.
- Kawaguchi, A., & Yamashita, F. (2017). Supervised multiblock sparse multivariable analysis with application to multimodal brain imaging genetics. *Biostatistics*, 18(4), 651–665.
- Kim, G.-W., Kim, Y.-H., & Jeong, G.-W. (2017). Whole brain volume changes and its correlation with clinical symptom severity in patients with schizophrenia: A DARTEL-based VBM study. *PLOS ONE*, 12(5), 1–15.
- Kwan, R. K.-S., Evans, A. C., & Pike, G. B. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11), 1085–1097.
- Li, K., & Luo, S. (2017). Functional joint model for longitudinal and time-to-event data: An application to Alzheimer's disease. *Statistics in Medicine*, 36(22), 3560–3572.
- Liang, Z., MacFall, J. R., & Harrington, D. P. (1994). Parameter estimation and tissue segmentation from multispectral MR images. *IEEE Transactions on Medical Imaging*, 13(3), 441–449.
- McCormick, T. H., Rudin, C., & Madigan, D. (2012). Bayesian hierarchical rule modeling for predicting medical conditions. *The Annals of Applied Statistics*, 6(2), 652–668.
- Nguyen, T. M., Wu, Q. M. J., & Ahuja, S. (2010). An extension of the standard mixture model for image segmentation. *IEEE Transactions on Neural Networks*, 21(8), 1326–1338.
- Ozenne, B., Subtil, F., Østergaard, L., & Maucourt-Boulch, D. (2015). Spatially regularized mixture model for lesion segmentation with application to stroke patients. *Biostatistics*, 16(3), 580–595.
- Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31(2), 153–163.
- Wang, X., Nan, B., Zhu, J., & Koeppe, R. (2014). Regularized 3D functional regression for brain image data via Haar wavelets. *The Annals of Applied Statistics*, 8(2), 1045.
- Wang, Y., & Blei, D. M. (2019). The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528), 1574–1596.
- Zou, K. H., Wells III, W. M., Kikinis, R., & Warfield, S. K. (2004). Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in Medicine*, 23(8), 1259–1282.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Banerjee, A., & Mukhoti, S. (2023). Characterization of a robust probabilistic framework for brain magnetic resonance image data distributions. *Stat*, 12(1), e541. <https://doi.org/10.1002/sta4.541>