Neuroscience Informatics 1 (2021) 100002

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Original article

Tsallis generalized entropy for Gaussian mixture model parameter estimation on brain segmentation application

Mehran Azimbagirad [a,*], Luiz Otavio Murta Junior [b]

[a] Laboratory of Medical Information Processing (LaTIM), Faculty of Medicine and Health Sciences, University of Western Brittany, Brest, France
[b] Department of Computing and Mathematics, Faculty of Philosophy, Science and Languages, University of São Paulo, Ribeirão Preto, SP, Brazil

ARTICLE INFO

Article history:
Received 3 May 2021
Received in revised form 1 July 2021
Accepted 2 July 2021

Keywords:
Tsallis entropy
Shannon entropy
Expectation-Maximization
K-means
Gaussian Mixture Model

ABSTRACT

Among statistical models, Gaussian Mixture Models (GMMs) have been used in numerous applications to model the data in which a mixture of Gaussian curves fits them. Several methods have been introduced to estimate the optimum parameters to a GMM fitted to the data. The accuracy of such estimation methods is crucial to interpret the data. In this paper, we proposed a new approach to estimate the parameters of a GMM using critical points of Tsallis-entropy to adjust each parameter's accuracy. To evaluate the proposed method, seven GMMs of simulated random (noisy) samples generated by MATLAB were used. Each simulated model was repeated 1000 times to generates 1000 random values obeying the GMM. In addition, five GMM shaped samples extracted from magnetic resonance brain images were used, aiming for image segmentation application. For comparison assessment, Expectation-Maximization, K-means, and Shannon's estimator were employed on the same dataset. These four estimation methods using accuracy, Akaike information criterion (AIC), Bayesian information criterion (BIC), and Mean Squared Error (MSE) were evaluated. The mean accuracies of the Tsallis-estimator for simulated data, i.e., the mean values, variances, and proportions, were 99.9(±0.1), 99.8(±0.2), and 99.7(±0.3)%, respectively. For both datasets, the Tsallis-estimator accuracies were significantly higher than EM, K-means, and Shannon. Tsallis-estimator, increasing the estimated parameters' accuracy, can be used in statistical approaches and machine learning.

1. Introduction

A statistical model describes a set of variables and relations concerning the generation of some sample data and similar data from a larger population. There are three purposes for a statistical model, according to [1]:

- Prediction
- Extraction of information
- Description of stochastic structures

Generally, these aims are interested in all studies that need to be modeled.

Usually, one acquires a sample from the population, and then a model is guessed for the sample. Then, using several evaluation techniques, e.g., Akaike information criterion (AIC), Bayesian information criterion (BIC), Mean Squared Error (MSE), and another statistical index, the model is evaluated to quantify its performance.

Once a model (with its parameters) is guessed for sample data, several crucial indicators can be obtained. For example, once a relation between spreading a contagious disease and accumulating peoples is modeled, an obtained indicator shows the minimum distance to avoid the infection [2]. Another example is, clustering the data due to keep similar samples in different groups. Such clustering tasks are crucial for medical image processing, especially in neuroscience. For example, in the diagnostic scheme, follow up and predict the brain disease using the acquired image, clustering is a central part of such approaches.

A straightforward example is to cluster a brain image to White Matter (WM), Gray Matter (GM), or Cerebrospinal Fluid (CSF) since each compartment has its functionality for investigation. One step more, only GM can be divided into several sections (according to brain anatomy) to analyze each section in depth. Therefore, fit a model to raw data enables us to deeply explore and interpret that data which can be marketing data or in a higher level in neuroscience.

* Corresponding author. Address for correspondence: Av. 12 Foch, IBRBS, LATIM lab, 29200, Brest, France.
E-mail address: mehran.azimbagirad@univ-brest.fr (M. Azimbagirad).

https://doi.org/10.1016/j.neuri.2021.100002

Once provided a data model, it is possible to separate them using that model's indicators (parameters). For example, in [3,4] they investigated Model-Based Clustering and Classification in both theoretical and applied research. Moraes et al. [5] introduced a new method based on the principle curve model for data clustering. In order to estimate the parameters of the model, usually, two methods are used:

- Maximum Likelihood Estimation (MLE)
- Method of Moments Estimation (MME)

MME is discussed in several books (for more details, see [6]) beyond this study. MLE will shortly be explained in the next section, along with two examples describing this method's implementation.

1.1. Parameter estimation by MLE

As an estimator method for the parameters of a statistical model, MLE has widely been used [7]. Assume that $(X = x_i,\ i = 1..n)$ is a sample data by a Probability Distribution Function (*PDF* or shortly f), e.g., Gaussian (Normal) distribution $\mathcal{N}(x_i; \mu, \sigma^2)$ with unknown parameters mean μ and variance σ^2. The parameters μ and σ^2 can be estimated using the MLE by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable given the model.

Generally, suppose that there is a sample $x_1, x_2, \ldots, x_n$ by a known (or guessed) *PDF* depends on some unknown parameter θ, written $f(X = x_i; \theta)$. The primary goal here is to estimate θ, such that $f(X = x_i; \theta)$ is a reasonable estimate of the model. The model quality can be defined as a specific error calculated by any of the methods mentioned above for the guessed *PDF* on the estimated parameters. For instance, a reasonable estimate of $\theta = (\mu, \sigma^2)$ would be the value such that maximizes *the joint density function* for all observations

$$PDF\,(X = x_1, X = x_2, \ldots, X = x_n; \theta) = f\,(X = x_1, x_2, \ldots, x_n; \theta)\,. \tag{1}$$

Now looking at this function from a different perspective by considering the observed values $x_1, x_2, \ldots, x_n$ to be the fixed parameters, whereas θ be the function's variable and allowed to vary freely

$$\begin{aligned} L\,(\theta) &= f\,(X = x_1, x_2, \ldots, x_n; \theta) \\ &= f\,(x_1; \theta) \times f\,(x_2; \theta) \times \cdots \times f(x_n; \theta) \\ \Rightarrow L\,(\theta) &= \prod_{i=1}^{n} f\,(x_i; \theta)\,, \end{aligned} \tag{2}$$

this function is called the *likelihood*. Now, in the spirit of MLE, one reasonable way to proceed is to treat the likelihood function $L\,(\theta)$ as a function of θ, and find the value of $\tilde{\theta}$ which maximizes it. In practice, it is often more convenient when working with the natural logarithm of the likelihood function, called the *log-likelihood*, defined as follows:

$$\ln L\,(\theta) = \ln \prod_{i=1}^{n} f\,(x_i; \theta) = \sum_{i=1}^{n} \ln f\,(x_i; \theta)\,. \tag{3}$$

Several properties of this function were illustrated in [8]. Different studies are working on different MLE types and its application (refer to the last few [9–11]).

As a bivariate function in calculus, to find the maximizer point of $L\left(\theta = (\mu, \sigma^2)\right)$, usually, the partial derivative is set to zero to find the optimum points $\tilde{\mu}, \tilde{\sigma}^2$. Since this function has a nonlinear form and many terms, the logarithm form of the function will be much easier to take the partial derivative. Taking the partial derivative of the log-likelihood concerning μ and σ^2 and then setting to 0, gives:

$$\tilde{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } \tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \tilde{\mu})^2}{n}, \tag{4}$$

what is expected as inductive reasoning for the population.

Nevertheless, it may not always be easy to solve the maximization equations. For example, if the *PDF* is a GMM with three components, there will be nine parameters to be estimated. Although numerical methods can be used to approximate a solution for the partial differential equations, the Expectation-Maximization (EM) algorithm [12] is an iterative method to find the maximizer of the likelihood, i.e., the parameters of GMM. The problem of MLE for GMM parameter estimation and how the EM can solve it are discussed in [13], though shortly is demonstrated as follows.

1.2. Parameter estimation by the EM

EM method is an iterative algorithm for parameter estimation for statistical models when some of the involved random variables are not observed or solving the equation $\frac{d \ln L(\theta)}{d\theta} = 0$ has no analytical solution. For many years, the EM algorithm had been used before being introduced by Orchard et al. [14] is a problem for missing information principle provided the theoretical foundation of the underlying idea. The EM was presented by Dempster [15], who gave the proof of general results about the algorithm's behavior. The EM algorithm uses an initial guess, i.e., $\theta^{(0)}$ for θ, then on the first iteration compute

$$Q\,(\theta; \theta^{(0)}) = E_{\theta^{(0)}}[\ln L(\theta)], \tag{5}$$

where E is the expectation and $Q\,(\theta; \theta^0)$ is now maximized concerning θ, that is, $\theta^{(1)}$ is found such that

$$Q\left(\theta^{(1)}; \theta^{(0)}\right) \geq Q\,(\theta; \theta^{(0)}), \tag{6}$$

for all possible θ. Thus, the EM algorithm consists of an E-step (Expectation) and followed by an M-step (Maximization) defined as follows:

I. **E-Step**: Compute $Q\,(\theta; \theta^{(t)})$ where $Q\,(\theta; \theta^{(t)}) = E_{\theta^{(t)}}[\ln L(\theta)]$,
II. **M-Step**: Find $\theta^{(t+1)}$ such that $Q\left(\theta^{(t+1)}; \theta^{(t)}\right) \geq Q\,(\theta; \theta^{(t)})$,

for all possible θ. The E-step and the M-step have repeated alternately until the difference $L\left(\theta^{(t+1)}\right) - L(\theta^{(t)})$ is less than epsilon, where epsilon is a prescribed small quantity. The computation of these two steps simplifies a great deal when it can be shown that the log-likelihood is linear in the sufficient statistic for θ. Several examples are discussed to illustrate these steps in the exponential family case [16] and to modify approaches to improve EM [17,18].

As an available algorithm for complex maximum likelihood computations, the EM algorithm has several appealing properties relative to other iterative algorithms such as Newton-Raphson based approaches. First, it is easily implemented because it relies on complete-data computations; the E-step of each iteration only involves taking expectations over complete-data conditional distributions. The M-step of each iteration only requires complete-data maximum likelihood estimation. Secondly, it is numerically stable and converges to a local maximum or saddle point of MLE. A disadvantage of the EM is its convergence rate, which can be extremely slow if much data are missing. Dempster, Laird, and Rubin [15] showed that convergence is linear, with the rate proportional to the fraction of information. Another drawback of the

EM is the necessity of a large number of initial parameters. For example, for a GMM with three components, nine initial parameters, i.e., $(\mu_1, \sigma_1, \pi_1), (\mu_2, \sigma_2, \pi_2)$ and (μ_3, σ_3, π_3) are needed to start the approach.

1.3. Parameter estimation by K-means

The K-means algorithm is another simple iterative method to estimate a model's parameters, e.g., the means of a population alongside clustering data [19]. A detailed history of K-means and descriptions of several variations are given in [20], though this algorithm is illustrated very shortly. Assume that a population including three components as the same example used in the MLE estimator for GMM. One can use the K-means algorithm to estimate the means (three means) of the components. The algorithm is initialized by three (K) giving points as the initial means representatives or centroids for the components. Then the algorithm iterates between two steps until it converges to the best estimations:

I. **Data Assignment**: Each member of the population is assigned to its closest centroid using the given centroids. Then this step clusters the population into three groups.

II. **Relocation of the means**: For each cluster (group), its mean is recalculated.

The algorithm converges when the assignments (and hence the centroids) no longer change. The number of iterations required for convergence varies and may depend on the population, but this algorithm's complexity can be considered linear in the dataset size as a first cut. Concerning the other estimators like MLE, K-means is faster. Recently, [21] used K-means for functional data clustering based on a combination of a hypothesis test of parallelism and the test for equality of means. Nevertheless, K-means' precision alone may not be acceptable in those studies looking for a high precision approximation since only consider the means and not the variances. This study considers this algorithm to estimate the means and the variances and proportions as follows.

Assume that a dataset is given such that a GMM with three components needs to be fitted to this data. If K-means is employed to estimate three means of the data, then the middle points between the first and the second and the middle point between the second and the third means can be considered the local minimum points of a GMM curve. Using these two points and then segment the data into three parts, the variance and proportion for each part can be calculated. Thus, in addition to the means, the variances and the proportions are estimated using K-means.

1.4. Additive and nonadditive entropy

A statistical model estimating how much information is required, on average, to identify random samples from a distribution is defined by Shannon entropy. Equally, the Shannon entropy equation provides a way to estimate the average minimum number of bits needed to encode a string of symbols based on the symbols' frequency. Since researchers desire to model such systems (any system with a PDF with GMM to estimate the probability of its states), it could be defined as a statistical model. The Shannon entropy and its generalized version to estimate the parameters of a GMM are exemplified as follows.

1.4.1. Additive entropy

The entropy of a system is usually calculated from a probability distribution, where p_i is the probability of finding the system in each possible state i. Therefore, $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$ where n is the total number of states. For example, Shannon entropy described as

$$S = -\sum_{i=1}^{n} p_i \ln(p_i). \tag{7}$$

This formalism is restricted to the domain of validity of the Boltzmann-Gibbs-Shannon (BGS) statistics [22]. Generally, systems that obey BGS statistics are called extensive systems. Assume that a system is decomposed into two independent statistical subsystems A and B. Then, the probability of the composite system is $P^{\{A+B\}} = P^A.P^B$. It has been verified that the Shannon entropy has the additive property (or extensivity in some sense)

$$S(A+B) = S(A) + S(B), \tag{8}$$

hence for three subsystems C_1, C_2 and C_3 it is defined

$$S(C_1 + C_2 + C_3) = S(C_1) + S(C_2) + S(C_3). \tag{9}$$

Therefore, those probabilities are divided into three parts to calculate the Shannon entropy of each subsystem. For example, $C_1 = \{p_1, p_2, ...p_t\}$, $C_2 = \{p_{t+1}, p_{t+2}, ...p_k\}$ and $C_3 = \{p_{k+1}, p_{k+2}, ..., p_n\}$ then

$$\begin{aligned} S(C_1) &= -\sum_{i=1}^{t} \frac{p_i}{pC_1} \ln\left(\frac{p_i}{pC_1}\right), \\ S(C_2) &= -\sum_{i=t+1}^{k} \frac{p_i}{pC_2} \ln\left(\frac{p_i}{pC_2}\right), \\ S(C_3) &= -\sum_{i=k+1}^{n} \frac{p_i}{pC_3} \ln\left(\frac{p_i}{pC_3}\right), \end{aligned} \tag{10}$$

where pC_i is the sum of all probabilities in C_i. Finding the optimum t and k which maximize the Shannon entropy, is desired to estimate the best statistical model describing the system. Therefore, solving the subproblem

$$\arg\max_{t,k} S(C_1 + C_2 + C_3), \tag{11}$$

gives us $\tilde{t}$ and $\tilde{k}$ which can be used to estimate the parameters of GMM, i.e., Shannon estimator illustrated in methodology. One can solve the problem denoted in (11) through different approaches. The first way to find $\tilde{t}$ and $\tilde{k}$ is to check all possible combinations of $\tilde{t}$ and $\tilde{k}$ from 1 to n where $1 < \tilde{t} < \tilde{k} < n$. Another way is to find firstly $\tilde{t}$ and then $\tilde{k}$ i.e., the system has only two subsystems. Thus, it is possible to only look for $\tilde{t}$ to maximize the subproblem and then search for $\tilde{k}$ to maximize (11).

1.4.2. Nonadditive entropy

Some extension appears to become necessary for a specific class of physical systems, which entail long-range interactions, long-time memory, and fractal-type structures. Inspired by multifractal concepts, Tsallis has proposed a generalization of the BGS statistics [23]. Tsallis statistics are currently considered useful in describing nonadditive systems' thermostatic properties, measuring the uncertainty of random variables [24], and monitoring defects of a moving metallic surface [25]. The original idea of Tsallis entropy is based on a generalized entropic form,

$$S_q = \frac{1 - \sum_{i=1}^{n} p_i^q}{q-1}, \tag{12}$$

where n is the total number of possibilities of the states of the system and the real number q is an entropic index that characterizes the degree of nonadditivity (or nonextensivity). This expression

meets the BGS entropy in the limit $q \to 1$ proved in [23]. Tsallis entropy is nonextensive in such a way that for a statistically independent system (including two subsystems A and B), the entropy of the system is defined by the following pseudo additivity entropic rule

$$S_q(A+B) = S_q(A) + S_q(B) + (1-q) \times S_q(A).S_q(B), \quad (13)$$

where S_q is defined by Eq. (12). To find the maximum S_q for the same as classical entropy, an optimization problem should be solved as

$$\arg\max_{q,t} S_q(A+B). \quad (14)$$

In this paper, we employ the Tsallis entropy concept and the maximization criteria of its definition to estimate the parameters of a system in which its states' probability obeys a GMM distribution. Then the accuracy of this approach will be compared by EM, K-means, and classical entropy on several examples.

2. Methodology

2.1. Tsallis entropy estimator

Assume that for a general system by n states, frequency of state i is given by f_i $(i = 1, \ldots, n)$. The statistical model to fit the normalized histogram (divided f_i by total frequency N) is guessed by a GMM with three components C_1, C_2 and C_3 (defined in section 1.4.1). Therefore, GMM has a *PDF* by

$$PDF(x_i) = \sum_{j=1}^{3} \pi_j \times \mathcal{N}(x_i; \mu_j, \sigma_j^2),\ i = (1, 2, \ldots, n), \quad (15)$$

where x_i is the probability of the state i, π_j is the proportion of component C_j, $\mathcal{N}(\mu_j, \sigma_j^2)$ is a Gaussian distribution with mean μ_j and variance σ_i^2. Using Eq. (13):

$$\begin{aligned} S_q(C_1, C_2, C_3) &= S_q(C_1) + S_q(C_2) + S_q(C_3) \\ &+ (1-q) \times \left[S_q(C_1).S_q(C_2) + S_q(C_2).S_q(C_3) + S_q(C_1).S_q(C_3)\right] \\ &+ (1-q)^2 \times \left[S_q(C_1).S_q(C_2).S_q(C_3)\right], \end{aligned} \quad (16)$$

then, the maximum of S_q is provided by

$$\arg\max_{q,t,k} S_q(C_1, C_2, C_3). \quad (17)$$

One can solve the problem denoted in (17) by applying the same approaches described in section 1.4.2. Assume that, $\tilde{t}$ and $\tilde{k}$ are obtained such that maximize (17). Then, C_1, C_2 and C_3 are defined such that

$$\begin{aligned} C_1 &= \{p_1, p_2, \ldots p_{\tilde{t}}\}, \\ C_2 &= \left\{p_{\tilde{t}+1}, p_{\tilde{t}+2}, \ldots p_{\tilde{k}}\right\}, \\ C_3 &= \{p_{\tilde{k}+1}, p_{\tilde{k}+2}, \ldots, p_{\tilde{k}}\}, \end{aligned} \quad (18)$$

and then the parameters of the GMM are estimated by

$$\begin{aligned} \mu_j &= N \times mean\, C_j, \\ \sigma_j^2 &= N \times variance\, C_j,\ j = 1, 2, 3 \\ \pi_j &= sum\, C_j. \end{aligned} \quad (19)$$

These expressions reveal our proposal to estimate the GMM parameters. It should be mentioned that in (18), for estimating variance, one may use $(N-1)$ instead of N, but in large numbers, that does not change the results. Shannon's estimator is defined as Tsallis estimator but using Eq. (10) and Eq. (11). The algorithm of the Tsallis estimator can be seen in Appendix A.

2.2. Evaluation method

We used two strategies to evaluate Tsallis entropy estimation for GMM. First, using a collection of simulated data generated by MATLAB to evaluate the robustness of the method. This survey generated GMM shape data with several selections of the parameters and random (noisy) data. Therefore, the simulations include noise factors, making challenging models. We also repeated the simulations to distribute the noise around all parameters. The second strategy was the application of the Tsallis estimator tool in segmentation. It has been shown that the maximization of either BGS or Tsallis entropy of a system (e.g., image) leads to disjoint the system states. We used five Magnetic Resonance images [26] as five systems. Then, we extract the histogram of the images which have GMM shape. Then, we estimated the parameters of each GMM using the proposed tool. In addition, EM and K-means were employed and Shannon's estimators for comparison purposes.

We defined three errors for the mean, variance, and proportion estimated by EM, K-means, Shannon, and Tsallis estimators by

$$\begin{aligned} error_\mu &= \frac{|\,\mu - \tilde{\mu}\,|}{n}, \\ error_\sigma &= \frac{|\,\sigma - \tilde{\sigma}\,|}{s}, \\ error_\pi &= \left|\pi - \tilde{\pi}\right|, \end{aligned} \quad (20)$$

where $|\,.\,|$ is an absolute value, n is the maximum possible values for μ, and s is the maximum possible values for σ^2. The reason for defining these errors was the simplicity of showing accuracy. Therefore, the accuracy of the parameter θ is defined as

$$Accuracy_\theta = (1 - error_\theta) \times 100, \quad (21)$$

by percentage unit.

In addition to the accuracy defined above, we also estimated AIC [27], BIC [28], and MSE as another statistical index to assess the models provided using the estimators, i.e., EM, K-means, Tsallis and Shannon as:

$$AIC = n \times \log\left(\frac{RSS}{n}\right) + 2k, \quad (22)$$

$$BIC = n \times \log(RSS/n) + k \times \log(n), \quad (23)$$

where k is the number of parameters in the model and *RSS* is the residual sum of squares.

Choosing the best q for the Tsallis estimator needs a training step the same as a learning step in neural networks. Depending on the GMM artifacts, may the entropic index vary in a limited interval [29] to give the best results. In the datasets used in this study, we used 5% of the dataset as the training sample to find the optimum q in the interval $(-1, 1)$.

In practice, when calculating Eq. (12), for the small values of p_i, S_q may increase exponentially. To avoid this calculation error, we proposed an additional parameter α as a conditional limitation for p_i in Eq. (12) searching in interval $(10^{-6}, 10^{-2})$. Therefore, the recent equation is maximized by two parameters q and α. Thus, in the training step, these two parameters are chosen as the optimum parameters of the following equation

$$\arg\max_{q,t,k,\alpha} S_q(C_1, C_2, C_3). \quad (24)$$

Therefore, q and α are two parameters that can tune the accuracy of the Tsallis estimator.

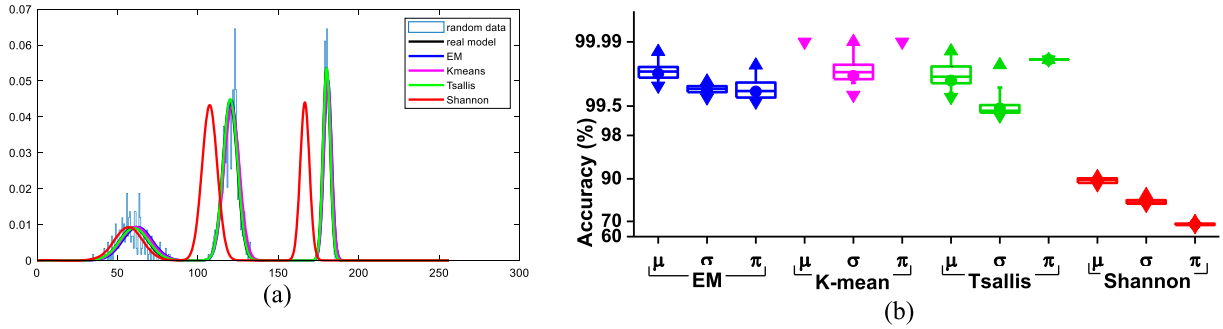

Fig. 1. (*a*) Histogram of a sample including 1000 random values of a GMM with the probability of the GMM for 1 to 256. $\mu_1 = 60$, $\mu_2 = 120$ and $\mu_3 = 180$, $\pi_1 = 0.2$, $\pi_2 = 0.5$ and $\pi_3 = 0.3$, $\sigma_1^2 = 75$, $\sigma_2^2 = 20$ and $\sigma_3^2 = 5$. (*b*) Box chart of the methods' accuracy, including max, min, and mean of calculated accuracy.

3. Results

Using MATLAB software and **mvnrnd** function, we generated 1000 simulated random data by GMM in which $\mu_1 = 60$, $\mu_2 = 120$ and $\mu_3 = 180$ and $\pi_1 = 0.2$, $\pi_2 = 0.5$ and $\pi_3 = 0.3$. However, to generate more challenging simulations, the variances were chosen such that finding the local minimums turns difficult. For the first simulation, we run 1000 times the function generator to produce 1000 random data by $\sigma_1^2 = 75$, $\sigma_2^2 = 20$ and $\sigma_3^2 = 5$. One of the random samples and the GMM are presented in Fig. 1 (a). The mean accuracies of each parameter and method are presented in Fig. 1 (b).

As shown in Fig. 1, K-means had the best performance through the EM, and Tsallis are closed to K-means. Nevertheless, Shannon could not accomplish a good performance as the others.

For the simulation dataset, the found optimum q was -0.3 and $\alpha = 10^{-2}$ using the first sample as the training dataset. To make a more demanding challenge for estimating the parameters, we changed the variances to join data and generated 1000 random values. One sample of each selection of variances and the GMM is presented in Fig. 2, along with each method's mean accuracy for each parameter.

In Fig. 2, the samples from 2 to 7 were gradually more challenging due to estimate the parameters of the Gaussian mixture models. Nevertheless, Tsallis and Shannon estimators' accuracy was getting better for the complex samples, unlike the EM and K-means.

Averagely, for all seven simulated data, the accuracies of the methods are presented in Fig. 3 using violin plot according to the definition of the accuracy Eq. (21).

Another scenario to evaluate the parameter estimator methods was to calculate the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Mean Squared Error (MSE) on the six simulated data. The results are presented in Table 1.

In Table 1, the Tsallis estimator obtained a better score in AIC for all simulation data than the EM and K-means and Shannon and BIC and MSE.

In the second dataset for evaluating the Tsallis estimator, five T1-weighted volumetric MRI images [26] were used to estimate the parameters of the GMM fitted to the histogram of the images. Each image has a dimension of $240 \times 240 \times 48$, i.e., $2,764,800$ voxels (values). Also, each image includes three brain compartments; WM, GM, or CSF. That means each voxel has an intensity level (a number in (0,255)), which belongs to one of these compartments. Therefore, it is expected to have a GMM shape with three components (three bell shapes) for the image voxels histogram. Fig. 4 shows the mid slice of each volume image and its histogram.

Since the actual parameters of the GMM fitted to the histograms are not available, ground truth is needed to compare the estimated parameters with it. Ground-truth is provided by an expert who knows each voxel in the image belongs to which tissue, i.e., WM, GM, or CSF. The estimations of the parameters were calculated using this ground truth as reference which are presented in Table 2.

For the actual images (dataset), the optimum q was found by -0.5 and $\alpha = 10^{-2}$ using the first sample (image). The accuracies of the estimated parameters by all methods are presented in Fig. 5.

The accuracy of the methods which estimate the parameters of the GMM fitted to the histogram of five actual images is presented in Table 3. This table shows the accuracy of our suggestion is higher than EM, K-means and Shannon estimators. That is, the Tsallis estimator gives the best thresholds to segment the images.

4. Discussion

Using Tsallis entropy concept for several applications such as segmentation [30], characterizing threshold channel bank profiles [31], Monitoring defects of a moving metallic surface [25] etc. had been introduced. However, to the best of our knowledge, for the first time, we adopted Tsallis entropy to estimate the parameters of a GMM, considering the optimum points on the maximization of Tsallis entropy in the fitting model scheme. Moreover, we tested it on several datasets with GMM shapes. Therefore, one can use this estimator in statistical tools, e.g., Origin, SAS, R, and MATLAB, as a

Table 1
Akaike information criterion (AIC), Bayesian information criterion (BIC), and Mean Squared Error (MSE) estimated for EM, K-means, Tsallis, and Shannon on six simulation data, S1 to S6. Lower score showing a better model. The best scores are highlighted.

	MSE				AIC				BIC			
	EM	K-means	Tsallis	Shannon	EM	K-means	Tsallis	Shannon	EM	K-means	Tsallis	Shannon
S1	4.91E-06	7.79E-06	2.33E-07	0.00018	−12217.5	−11756.3	−15265.4	−8563.0	−12202.8	−11741.6	−15250.6	−8548.2
S2	8.59E-07	1.53E-06	5.31E-08	7.75E-05	−13960.1	−13384.0	−16744.0	−9458.6	−13945.4	−13369.3	−16729.3	−9443.9
S3	3.34E-07	7.85E-07	2.12E-08	3.00E-05	−14904.7	−14051.0	−17664.6	−10408.0	−14890.0	−14036.3	−17649.9	−10393.3
S4	1.63E-07	4.17E-07	1.12E-08	1.12E-05	−15623.0	−14683.3	−18303.7	−11396.4	−15608.3	−14668.6	−18288.9	−11381.7
S5	8.37E-08	2.06E-07	8.36E-09	4.49E-06	−16288.8	−15390.2	−18592.5	−12306.9	−16274.1	−15375.4	−18577.8	−12292.1
S6	5.12E-08	1.15E-07	7.39E-09	2.02E-06	−16781.1	−15973.8	−18716.0	−13103.8	−16766.4	−15959.0	−18701.2	−13089.1
S7	4.05E-08	9.35E-08	7.63E-09	1.41E-06	−17013.9	−16178.2	−18684.5	−13465.5	−16999.2	−16163.5	−18669.7	−13450.8

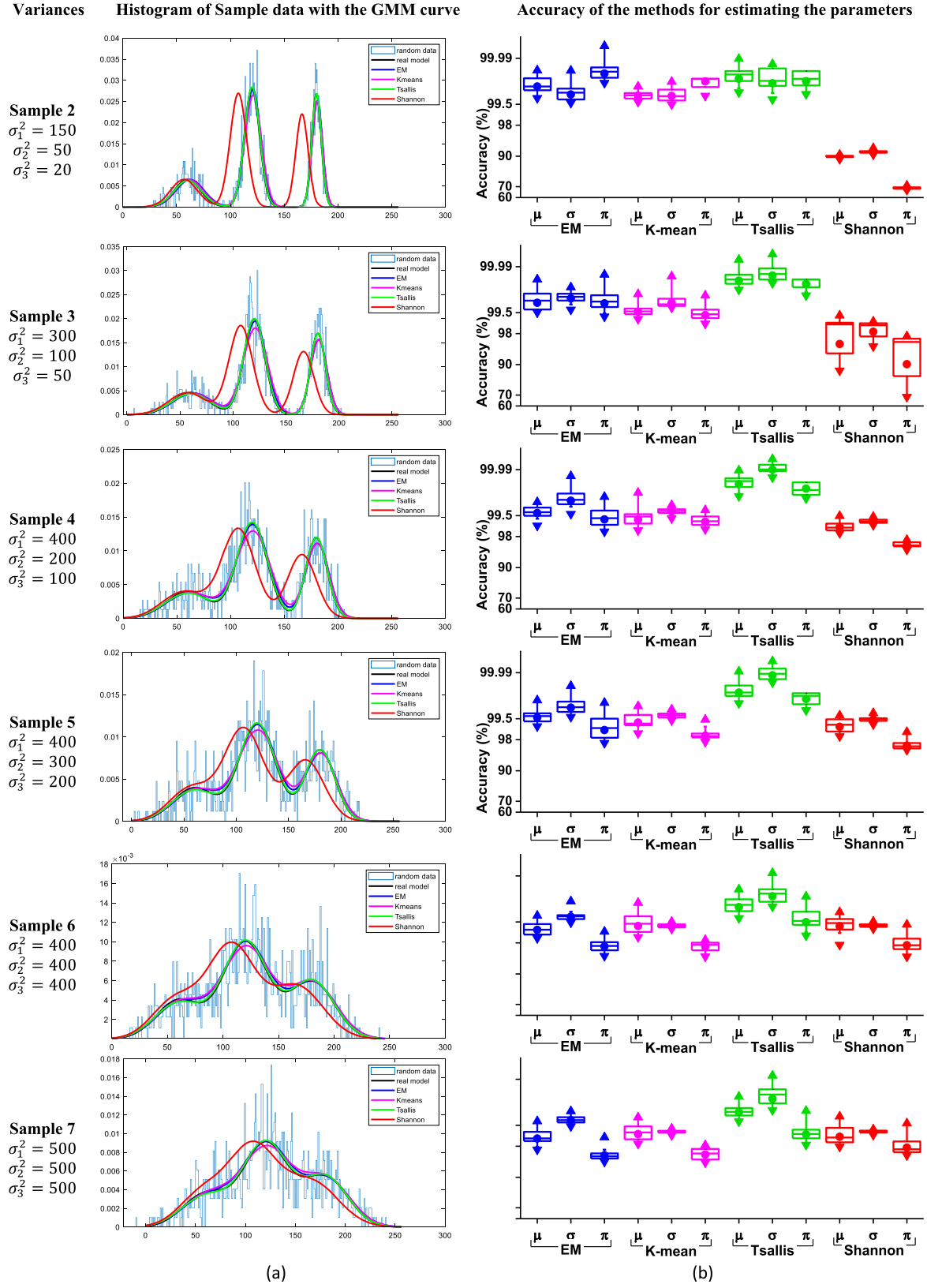

Fig. 2. (*a*) Histograms of 1000 random datasets of five GMMs with $\mu_1 = 60$, $\mu_2 = 120$ and $\mu_3 = 180$, $\pi_1 = 0.2$, $\pi_2 = 0.5$ and $\pi_3 = 0.3$, and different variances. (*b*) Box chart of the accuracy of the methods, including max, min, and mean of calculated accuracies.

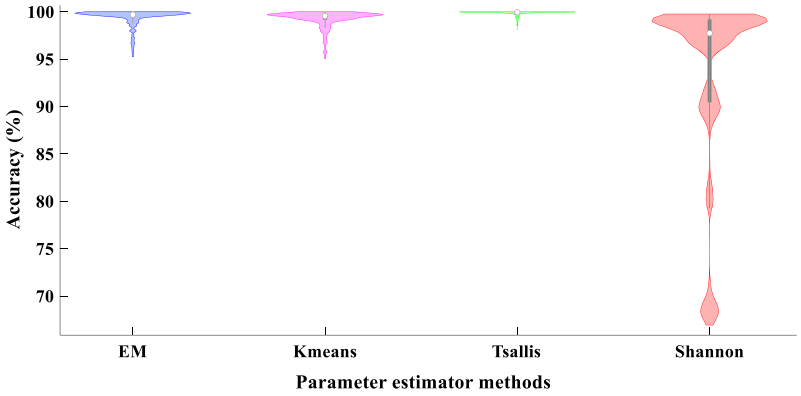

Fig. 3. Violin chart of the accuracies of four-parameter estimation for seven simulations.

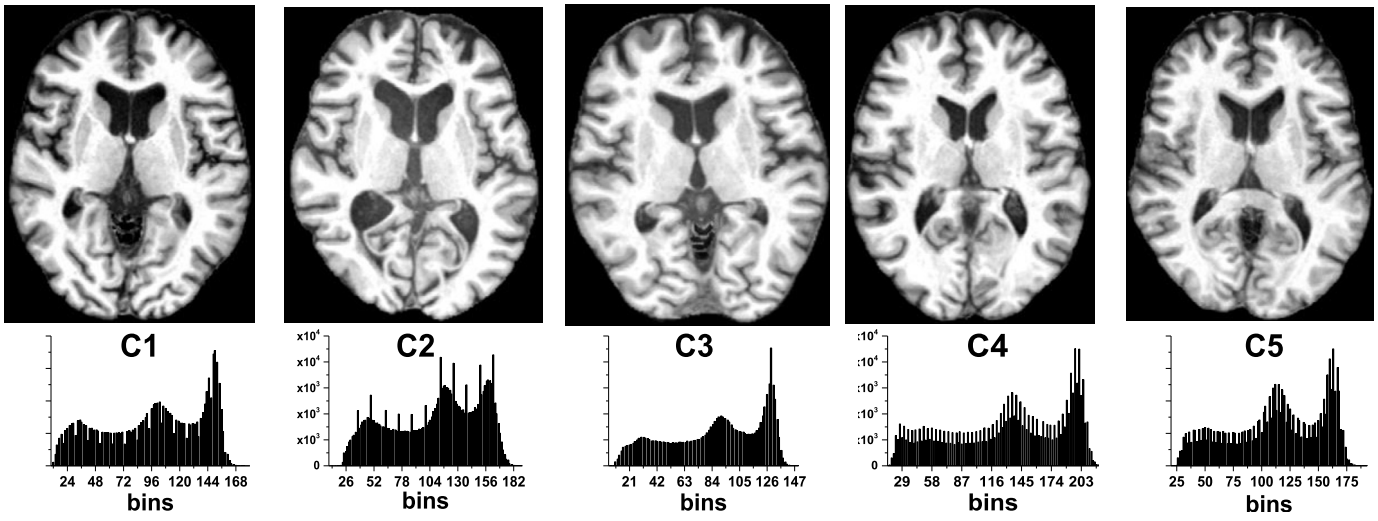

Fig. 4. One slice of each real image [26] used in this study, with their histogram indicated below. It can be seen from the histogram that each image has different properties, such as distribution values of each tissue, noise, and artifacts.

Table 2
The parameters of the GMM fitted to the histograms of the real images, estimated by the manual task.

Parameters	Images				
	C1	C2	C3	C4	C5
μ_1	49.86	59.06	45.31	67.57	60.82
μ_2	106.27	116.80	91.29	144.49	114.52
μ_3	144.32	150.95	123.41	194.44	158.78
σ_1^2	561.99	501.20	454.82	963.91	532.67
σ_2^2	372.08	479.88	281.12	663.71	402.71
σ_3^2	124.42	306.35	121.98	152.21	117.43
π_1	0.36	0.31	0.33	0.35	0.27
π_2	0.37	0.43	0.36	0.40	0.45
π_3	0.27	0.26	0.31	0.25	0.28

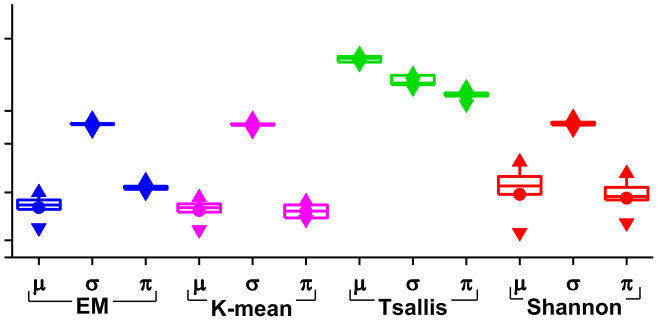

Fig. 5. Box chart of the methods' accuracy, including max, min, and mean of calculated accuracy.

parameter estimator of a GMM model as well as image software for segmentation purposes.

Generally, Tsallis estimator for the GMM parameters may have different accuracy for each parameter. For example, in Fig. 1 and Fig. 2, the means estimations' accuracy is not the same as the variances. It is possible to adopt the Tsallis estimator to achieve high precision for a parameter of interest. In the training step, the optimum (q, α) should be searched such that the optimum parameter is estimated with the highest accuracy. For example, if the proportions of a GMM are pursued, in the training step, (q, α) is obtained to maximize the accuracy of only the proportions. Generally, we recommend a strategy to choose the optimum (q, α) in machine learning approaches which are using the training data (commonly 30%) to find the best (q, α), then using them to estimate the parameters of the model. This flexibility feature of the Tsallis estimator could get better image segmentation results compared to the EM, K-means, and Shannon.

In the simulation data (see Fig. 1 and 2), we adopt only one q and α parameters for the Tsallis estimator. Since each simulation had different variances, it was possible to increase the accuracy of the Tsallis estimator by choosing (q, α), individually for each simulation. Besides, by increasing the variances to make it difficult for methods to estimate parameters, Tsallis and Shannon's estimators' accuracy increased instead, showing the robustness of these meth-

Table 3
The mean accuracy of the parameter estimations by four methods on the five real subjects, μ = mean, σ = variance, and π = proportion, the parameters of GMM.

	Methods											
	EM			K-means			Tsallis			Shannon		
Parameters	μ	σ	π	μ	σ	π	μ	σ	π	μ	σ	π
Image 1	84.49	99.10	90.87	83.42	99.09	81.02	99.97	99.86	99.75	90.26	99.15	94.33
Image 2	87.77	98.85	92.90	86.23	98.82	87.23	99.96	99.87	99.79	93.78	98.89	87.77
Image 3	90.02	99.28	90.94	88.54	99.27	80.38	99.97	99.92	99.83	96.13	99.33	91.36
Image 4	76.46	99.08	91.65	75.54	99.05	83.90	99.97	99.83	99.63	74.10	99.06	78.78
Image 5	86.99	99.12	90.05	86.48	99.11	86.11	99.96	99.92	99.76	93.12	99.17	89.38
Mean	85.15	99.08	91.28	84.04	99.07	83.73	99.97	99.88	99.75	89.48	99.12	88.32

ods. Also, for the real data, it has been shown that [29,32], Tsallis entropy for estimating the critical points of a GMM has higher precision than Shannon entropy.

Artifacts such as noise, inhomogeneity of the magnetic field, and partial volume effect directly impact the intensity histogram of the image voxel. These artifacts can be seen in Fig. 4, all five histograms. If one uses the histogram for segmentation purposes, such artifacts affect the accuracy of the results accordingly. For Tsallis entropy, precisely, the critical points of its maximizer (see Eq. (17)), usually, outliers (noises) can not move these points since they decrease the value of S_q in Eq. (16). Still, we should mention that using only Tsallis entropy and, consequently, only image histograms, especially 3D images, may not be a robust decision [32,33]. For this reason, it is recommended to combine Tsallis entropy with other approaches to improve the results of segmentation. However, our study is beyond segmentation and its subbranches. Therefore, we only used image histograms as some examples for the Gaussian mixture model to estimate the curve parameters that can fit them. Indeed, we proposed an estimator approach for curve fitting, which segmentation can be an application for that.

We introduced and evaluated an estimator for GMM with 3 components since one of its applications is to segment brain image into three materials, i.e., the WM, the GM and the CSF. Finding two parameters (q, α) for this estimator increases the computation and consequently CPU time. Although results showed an acceptable accuracy and speed in the simulated samples and natural subjects, its robustness still needs to be evaluated in GMM with more than three components and more GMM dimensions as a future study.

5. Conclusions

In this paper, we introduced a tool using the optimum parameters that maximize Tsallis generalized entropy to estimate a GMM with three components. This tool, the same as machine learning approaches, at first, needs a training step to obtain two required values (q, α). These values can tune the Tsallis estimator's accuracy for either all or a specific parameter of the GMM. The Tsallis estimator's goodness was compared with other well-known estimators, i.e., Expectation-Maximization, K-means, and Shannon entropy. Using a series of simulation and real datasets, we estimated the accuracy, Akaike information criterion (AIC), Bayesian information criterion (BIC), and Mean Squared Error (MSE) of all investigated methods. The accuracy comparison results showed that the precision of the Tsallis estimator averagely was higher than Shannon, EM, and K-means. Besides, for the AIC, BIC, and MSE, the proposed estimator tool had a better performance than the other methods. The proposed tool's innovation is this estimator's ability to increase the accuracy for a specific parameter among all parameters of a model. Although results showed an acceptable accuracy in the simulated samples and actual subjects, its robustness still needs to be evaluated in GMM with more than three components or more dimensions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Tsallis estimator algorithm on Gaussian Mixture model to estimate the parameters of the probability distribution function:

Tsallis Estimator

Input: *Data (X), q, α*

Solve $\mathbf{argmax}_{t_1,t_2}(S_{q,\alpha})$ *to obtain* $\tilde{t}_1, \tilde{t}_2$ *by:*
  for *all* t_1 *and* t_2 $(1 < t_1 < t_2 < k)$ ***do***
    Calculate the parameters $(\pi_i, \mu_i, \sigma_i^2)$, $i = 1, 2, 3$ *using* t_1, t_2 *for a GMM*
    Calculate S_q *by using GMM and Eq. (15)*
  end
using $\tilde{t}_1, \tilde{t}_2$ *optimum, estimate* $(\tilde{\pi}_i, \tilde{\mu}_i, \tilde{\sigma}_i^2)$, $i = 1, 2, 3$
Output: $(\tilde{\pi}_i, \tilde{\mu}_i, \tilde{\sigma}_i^2)$, $i = 1, 2, 3$

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neuri.2021.100002.

References

[1] S. Konishi, G. Kitagawa, Information Criteria and Statistical Modeling, Springer Science & Business Media, 2008.
[2] T. Zhao, C. Cheng, H. Liu, C. Sun, Is one- or two-meters social distancing enough for COVID-19? Evidence for reassessing, Publ. Health 185 (2020) 87, https://doi.org/10.1016/j.puhe.2020.06.005.
[3] S. Frühwirth-Schnatter, S. Ingrassia, A. Mayo-Iscar, Special issue on "Advances on model-based clustering and classification", Adv. Data Anal. Classif. 13 (1) (2019) 1–5, https://doi.org/10.1007/s11634-019-00355-w.
[4] S. Frühwirth-Schnatter, Panel data analysis: a survey on model-based clustering of time series, Adv. Data Anal. Classif. 5 (4) (2011) 251–280, https://doi.org/10.1007/s11634-011-0100-0.
[5] E.C.C. Moraes, D.D. Ferreira, G.B. Vitor, B.H.G. Barbosa, Data clustering based on principal curves, Adv. Data Anal. Classif. 14 (1) (2020) 77–96, https://doi.org/10.1007/s11634-019-00363-w.
[6] M.L. Hazelton, Methods of moments estimation, in: M. Lovric (Ed.), International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 816–817.
[7] X. He, W. Chen, W. Qian, Maximum likelihood estimators of the parameters of the log-logistic distribution, Stat. Pap. 61 (5) (2020) 1875–1892, https://doi.org/10.1007/s00362-018-1011-3.
[8] F.W. Scholz, Maximum likelihood estimation, https://doi-org.ez67.periodicos.capes.gov.br/10.1002/0471667196.ess1571.pub2, 2006.
[9] W. Shiyu, Two-Stage maximum likelihood estimation in the misspecified restricted latent class model, Br. J. Math. Stat. Psychol. 71 (2) (2018) 300–333, https://doi-org.ez67.periodicos.capes.gov.br/10.1111/bmsp.12119.
[10] J. Snoke, T.R. Brick, A. Slavković, M.D. Hunter, Providing accurate models across private partitioned data: secure maximum likelihood estimation 12 (2) (2018) 877–914, https://doi.org/10.1214/18-AOAS1171.

[11] R. Sheikhrabori, M. Aminnayeri, M.J.Q. Ayoubi, R.E. International, Maximum likelihood estimation of change point from stationary to nonstationary in autoregressive models using dynamic linear model 34 (1) (2018) 27–36, https://doi-org.ez67.periodicos.capes.gov.br/10.1002/qre.2233.
[12] M.P. Becker, I. Yang, K. Lange, EM algorithms without missing data, Stat. Methods Med. Res. 6 (1) (1997) 38–54, https://doi.org/10.1177/096228029700600104.
[13] M. Azimbagirad, Segmentação por entropia de Tsallis através de MRF para o parcelamento de ressonância magnética cerebral, Universidade de São Paulo, 2019.
[14] T. Orchard, M.A. Woodbury, A missing information principle: theory and applications, in: Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, 1972, pp. 697–715, https://apps.dtic.mil/dtic/tr/fulltext/u2/1022173.pdf.
[15] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B (1977) 1–38, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
[16] G. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley & Sons, 2008.
[17] M. Ranalli, R. Rocci, Mixture models for mixed-type data through a composite likelihood approach, Comput. Stat. Data Anal. 110 (2017) 87–102, https://doi.org/10.1016/j.csda.2016.12.016.
[18] J. Zhao, H.-J. Kim, H.-M. Kim, New EM-type algorithms for the Heckman selection model, Comput. Stat. Data Anal. 146 (2020) 106930, https://doi.org/10.1016/j.csda.2020.106930.
[19] R. Maronna, Charu C. Aggarwal and Chandan K. Reddy (eds.): Data clustering: algorithms and applications, Stat. Pap. 57 (2) (2016) 565–566, https://doi.org/10.1007/s00362-015-0661-7.
[20] R.C. Dubes, A.K. Jain, Algorithms for Clustering Data, Prentice Hall, Englewood Cliffs, 1988.
[21] A.Z. Zambom, J.A.A. Collazos, R. Dias, Functional data clustering via hypothesis testing k-means, Comput. Stat. 34 (2) (2019) 527–549, https://doi.org/10.1007/s00180-018-0808-9.
[22] J.N. Kapur, P.K. Sahoo, A.K. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Comput. Vis. Graph. Image Process. 29 (3) (1985) 273–285, https://doi.org/10.1016/0734-189X(85)90125-2.
[23] C. Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer Science & Business Media, 2009.
[24] G. Rajesh, S.M. Sunoj, Some properties of cumulative Tsallis entropy of order α, Stat. Pap. 60 (3) (2019) 933–943, https://doi.org/10.1007/s00362-016-0855-7.
[25] M.R.B. Dias, A.O.C. Junior, C.P. Dias, S.A. de Carvalho, J.A.O. Huguenin, L. da Silva, Monitoring defects of a moving metallic surface through Tsallis entropic segmentation, Physica A 534 (2019) 122175, https://doi.org/10.1016/j.physa.2019.122175.
[26] A.M. Mendrik, et al., MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans [online]. Available at http://mrbrains13.isi.uu.nl/.
[27] F.A. Brigatto, L.E.d.M. Lima, M.D. Germano, M.S. Aoki, T.V. Braz, C.R. Lopes, High resistance-training volume enhances muscle thickness in resistance-trained men, J. Strength Conditioning Res. (2020), https://doi.org/10.1519/jsc.0000000000003413, Publish ahead of print.
[28] G. Zhao, R. Zhu, S. Jiang, N. Xu, H. Bao, Y. Wang, Using the anterior capsule of the hip joint to protect the tensor fascia lata muscle during direct anterior total hip arthroplasty: a randomized prospective trial, BMC Musculoskelet. Disord. 21 (1) (2020) 21, https://doi.org/10.1186/s12891-019-3035-9.
[29] P. Diniz, L. Murta-Junior, D. Brum, D. de Araujo, A. Santos, Brain tissue segmentation using q-entropy in multiple sclerosis magnetic resonance images, Braz. J. Med. Biol. Res. 43 (1) (2010) 77–84, https://doi.org/10.1590/S0100-879X2009007500019.
[30] W. Shi, Y. Miao, Z. Chen, H. Zhang, Research of automatic medical image segmentation algorithm based on Tsallis entropy and improved PCNN, in: 2009 International Conference on Mechatronics and Automation, 9–12 Aug. 2009, 2009, pp. 1004–1008.
[31] A. Gholami, H. Bonakdari, A. Mohammadian, A method based on the Tsallis entropy for characterizing threshold channel bank profiles, Physica A 526 (2019) 121089, https://doi.org/10.1016/j.physa.2019.121089.
[32] M. Azimbagirad, F.H. Simozo, A.C.S. Senra Filho, L.O. Murta Junior, Tsallis-entropy segmentation through MRF and Alzheimer anatomic reference for brain magnetic resonance parcellation, Magn. Reson. Imaging 65 (2020) 136–145, https://doi.org/10.1016/j.mri.2019.11.002.
[33] J. Xu, B.M.W. Tsui, Interior and sparse-view image reconstruction using a mixed region and voxel based ML-EM algorithm, in: 2011 IEEE Nuclear Science Symposium Conference Record, 23–29 Oct. 2011, 2011, pp. 4070–4076.