Electronic Journal of Statistics
Vol. 8 (2014) 3141–3175
ISSN: 1935-7524
DOI: 10.1214/15-EJS985

Integrated Cumulative Error (ICE) distance for non-nested mixture model selection: Application to extreme values in metal fatigue problems

P. Vandekerkhove[1,2], J. M. Padbidri[3] and D. L. McDowell[4]

[1] *LAMA (UMR 8050), UPEMLV*
Université Paris-Est Marne-la-Vallée
F-77454, Marne-la-Vallée, France
e-mail: Pierre.Vandekerkhove@u-pem.fr

[2] *UMI Georgia Tech – CNRS 2958*
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0405, USA

[3] *QuesTek Innovations, LLC*
1820 Ridge Avenue
Evanston, IL 60201, USA
e-mail: jpadbidri@questek.com

[4] *George W. Woodruff School of Mechanical Engineering*
Georgia Institute of Technology
Atlanta, GA 30332-0405, USA
e-mail: david.mcdowell@me.gatech.edu

Abstract: In this paper, we consider the problem of selecting the most appropriate model, amongst a given collection of mixture models, to describe datasets likely drawn from mixture of distributions. The proposed method consists of finding the quasi-maximum likelihood estimators (QMLEs) of the various models in competition, using the Expectation-Maximization (EM) type algorithms, and subsequently estimating, for every model, a statistical distance to the true model based on the empirical cumulative distribution function (cdf) of the original dataset and the QMLE-fitted cdf. To evaluate the goodness of fit, a new metric, the Integrated Cumulative Error (ICE) is proposed and compared with other existing metrics for accuracy of detecting the appropriate model. We state, under mild conditions, that our estimator of the ICE distance converges at the rate $\sqrt{n}$ in probability along with the consistency of our model selection procedure (ability to detect asymptotically the right model). The ICE criterion shows, over a set of benchmark examples, numerically improved performance from the existing distance-based criteria in identifying the correct model. The method is applied in a material fatigue life context to model the distribution of indicators of the fatigue crack formation potency, obtained from numerical experiments.

AMS 2000 subject classifications: Primary 62G05, 62G20; secondary 62E10.
Keywords and phrases: EM algorithm, extreme value, Gumbel, mixture, model selection, metal fatigue, quasi-maximum likelihood, probability distance.

Received April 2013.

Contents

1. Introduction

Statistical practitioners are frequently interested in fitting mixture models to univariate datasets for which nonparametric density estimates show several clear departures from a description assumed to be accurate using one single standard probability density function (pdf). Assuming the existence of a mixture model accurately describing the dataset, a difficult proposition nevertheless arises of defining a criterion for the best model among all the plausible scenarios, i.e. possible numbers of components and collection of mixed parametric density families. If the mixed densities are supposed to belong to the same parametric density family, the above problem turns into estimating the number of components that best describes the mixture model. This order determination problem has been studied in several ways, see for instance Henna [1], Izenman and Sommer [2], Roedner [3], for various nonparametric techniques or Lindsay [4], Dacunha-Castelle and Gassiat [5], for moment-based methods, Keribin [6] for a penalized maximum likelihood selection method, or Berkhof et al. [7] for a Bayesian approach. To our knowledge, when the mixed densities possibly arise from different parametric families, inducing the exploration of a possibly high number of combinatorial models, there is no existing specific methodology. We nevertheless mention the work of Vuong [8], who proposed asymptotic likelihood tests to select the closest model to fit the given dataset from among a set of competing models based on the Kullback-Leibler (KL) information. However, it is important to note that the methodology developed in that paper fails in providing a total order on the set of competing models, the KL information being a statistical divergence, but not a true probability-distance (lack of symmetry).

The aim of this paper is to develop a finite-sample oriented methodology that can order the models in competition, based on their ability to resample the dataset of interest. For this purpose, we suppose that for each model in competition, we can identify its quasi-maximum likelihood estimator (QMLE), the true model being possibly part of the competing models. The basic idea is then to estimate a true distance between the models in competition and the distribution of the observations, based, for each model, on the comparison between the QMLE-fitted cdf and the empirical cdf of the dataset of interest. In this paper, the main motivation for developing such a mixture modeling method arises from observations of fatigue life distributions of metals.

Typically fatigue damage in material science is defined as the degradation of material properties due to the repeated application of stresses and strains leading to material failure [9]. In metallic materials and alloys, fatigue damage in the high cycle and very high cycle fatigue regimes (fatigue life of few to hundreds of millions of loading cycles) is initiated at the level of structural units called grains, whose size ranges from a few to a few hundred microns (1 micron = 10^{-6} metres). Each of these grains has a specific crystalline structure i.e. atomic arrangement. In the high cycle fatigue regime, damage in materials accumulates due to irreversible motion of defects along specific crystallographic directions. The rate of damage accumulation and consequently crack formation and fatigue life, is due to a combination of variables such as microstructural features (grain size, crystalline structure etc.) and applied loading. The probability of crack formation in a given volume of material is governed by the extreme value probability of a favorable combination of microstructural features [12]. Thus, the high cycle fatigue life of a material manifests as a distribution rather than a unique value for multiple experimental realizations and is described well by extreme value distributions.

The variation in distribution of fatigue life in metals is observed to a greater extent in the high cycle and very high cycle fatigue regimes than for low cycle fatigue life (thousands to tens of thousands of fatigue cycles) [10, 11], observed experimentally for a wide variety of metal alloys [13, 14, 15, 16, 17]. The distribution of the fatigue life also varies with the mechanism of crack initiation in the material [14, 17]. For a more detailed discussion on crack initiation mechanisms in materials, the reader is referred to Suresh [9]. In the high and very high cycle fatigue regime, the fatigue life scatter is described by extreme value distributions. Przybyla and McDowell [12] used a Gumbel distribution to quantify the variation of fatigue indicator parameters obtained from numerical experiments. Other extreme value distributions such as 2 or 3 parameter Weibull distributions have also been used to model the scatter in the observed fatigue life [16, 17, 18, 19]. Mixtures of extreme value distributions have been used to model fatigue life distributions when multiple mechanisms for crack formation have been observed [17, 19] with each extreme value distribution describing the scatter associated with a different crack formation mechanism. It is to be noted that an approach to describing the fatigue life scatter using a mixture of distributions for a single crack formation mechanism has not been considered in any of the above studies.

In the present work, we confine ourselves to quantifying the distribution of the extreme values of stresses that act as driving forces for the motion of defects along specific crystallographic directions within a grain and cause damage accumulation. This stress is referred to as resolved shear stress and can be a potent indicator of the crack formation life and thus, the fatigue life of the material. The extreme value of the resolved shear stress can be influenced by factors such as neighboring grains which could lead to the possibility of a corrupted/mixture of extreme value distributions to describe the distribution of the resolved shear stress. Thus, the approach taken here is to develop a generalized framework of identifying the "best" mixture model, not all of which might be extreme value distributions.

The paper is organized as follows. Section 2 is devoted to a detailed description of the model choice problem which is to be addressed and the methodology proposed in answer, while Section 3 is dedicated to the statement of the asymptotic properties of our method (convergence rate and consistency). The finite-sample performance of the proposed model selection method is studied in comparison with two other model selection procedures, based on the Kolmogorov-Smirnov and Shannon-Jensen distances, for various scenarios through Monte Carlo experiments in Section 4. In Section 5 the proposed method is applied to real datasets obtained from numerical experiments where mixtures of Gumbel and Gaussian distributions are suspected. Appendix A is dedicated to the proofs of the theorems established in Section 3. In Appendix B we provide a brief description of the QMLE and its asymptotic convergence properties, and show that the technical assumptions insuring the validity of our method are fully satisfied when considering mixtures of Gumbel and Normal distributions as applied to datasets obtained from numerical simulations in Section 5.

2. Problem and methodology

Let us suppose that we observe a univariate iid sample $X = (X_1, \ldots, X_n)$ distributed according to an unknown pdf f_0 which is possibly a mixture of pdfs belonging to a given collection

$$\mathcal{M} := \left\{ f_j(x, \vartheta_j) = \sum_{k=1}^{K_j} \pi_{j,k} h_{j,k}(x, \theta_{j,k}), \quad x \in \mathbb{R}, \quad j = 1, \ldots, J \right\}, \tag{1}$$

where, for all $j \in \mathcal{J} := \{1, \ldots, J\}$, respectively, the Euclidean parameter

$$\vartheta_j := (\pi_j; \theta_{j,1}, \ldots, \theta_{j,K_j}), \quad \text{with} \quad \pi_j := (\pi_{j,1}, \ldots, \pi_{j,K_j}),$$

is supposed to belong to a parametric space $\Theta_j := S(K_j) \prod_{k=1}^{K_j} \Phi_{j,k}$, where $S(K_j) := \{\pi_{j,k} > 0,\ 1 \le k \le K_j : \sum_{k=1}^{K_j} \pi_{j,k} = 1\}$, $\Phi_{j,k}$ is a parametric space corresponding specifically to each $\theta_{j,k}$, and $\{h_{j,k}(\cdot, \theta_{j,k}),\ k = 1, \ldots, K_j\}$ is a set of given pdfs.

For simplicity and to avoid technical issues related to the identifiability and degeneracy of the MLEs asymptotic normality in the neighborhood of negligible mixture weights, see example given in Keribin [6], p. 51, we suppose that there are no *nested* models in the collection $\mathcal{M}$, which is stated in the following assumption.

Assumption (NN). We suppose that in the collection $\mathcal{M}$ there do not exist two indices j_1 and j_2 such that $K_{j_1} < K_{j_2}$ and

$$f_{j_1}(x, \vartheta_{j_1}) = f_{j_2}(x, \vartheta_{j_2}) \quad x \in \mathbb{R},$$

when considering

$$\pi_{j_2} = (\pi_{j_1,1}, \ldots, \pi_{j_1,K_j-1}, \underbrace{0, \ldots, 0}_{K_{j_2}-K_{j_1}}).$$

Such a setup holds easily by assuming that the mixture weight vectors are componentwise uniformly lower bounded over the collection $\mathcal{M}$. This last approach matches in fact the natural thinking of a practitioner during the construction process of an interpretable collection of models to investigate (every component having to play a "real" role). Let us state now the basic question of this paper:

Model Selection Problem. *How to select, among the J mixture models of interest considered in $\mathcal{M}$ and fitted by a QML approach, the one that fits the dataset X best, the true density f_0 of which is unknown?*

To answer this question we propose to introduce a new *adapted* statistical distance able to establish a total ordering among the models in competition (discrete topology on models set), with some expected good finite sample properties compared to those usually encountered when using the Kolmogorov-Smirnov (strong sensitivity of the supremum norm to local defects or outliers) and Shannon-Jensen distances (asymptotic tailed-oriented approach suffering of a bandwidth selection problem in finite range applications), see (10–11) for definitions and following discussion.

Proposed Methodology. For each label $j \in \mathcal{J}$, we introduce

$$\hat{\vartheta}_j(X_1^n) := \hat{\vartheta}_j := (\hat{\pi}_{j,1}, \ldots, \hat{\pi}_{j,K-1}; \hat{\theta}_{j,1}, \ldots, \hat{\theta}_{j,K})$$

the QMLE of $\vartheta_{j,*}$, where $\vartheta_{j,*}$ is defined as the minimizer in ϑ_j of the Kullback-Leibler divergence $\mathcal{K}(f_0, f_j(\cdot, \vartheta_j))$ over Θ_j, when considering the j-labelled model in family $\mathcal{M}$, *i.e.*

$$\vartheta_{j,*} := \arg \min_{\vartheta_j \in \Theta_j} \mathcal{K}(f_0, f_j(\cdot, \vartheta_j)). \tag{2}$$

Note that if there exists a label $j_0 \in \mathcal{J}$ such that $\mathcal{K}(f_0, f_{j_0}(\cdot, \vartheta_{j_0,*})) = 0$ then $\vartheta_{j_0,*}$ coincides with ϑ_0. Generally the above estimators are computed by using the Expectation Maximization (EM) algorithm, see *e.g.* Dempster *et al.* [21] and Wu [20], which is by far the most efficient and essential fitting method in missing data problems.

We define next, for every label $j \in \mathcal{J}$, the plug-in QMLE mixture estimate of $f_j(\cdot, \vartheta_{j,*})$, also denoted for convenience $f_{j,*}$, by

$$\hat{f}_j(x) := f_j(x, \hat{\vartheta}(X_1^n)) = \sum_{k=1}^{K_j} \hat{\pi}_{j,k} h_{j,k}(x, \hat{\theta}_{j,k}), \quad x \in \mathbb{R}. \tag{3}$$

Finally, we introduce the collection of pdfs $\mathcal{M}_*$ defined by

$$\mathcal{M}_* := \{f_0(x), f_{j,*}(x), \quad x \in \mathbb{R}, \quad j = 1, \ldots, J\}, \tag{4}$$

composed by f_0 and the *Kullback-closest* pdfs to f_0 picked from $\mathcal{M}$, along with the key Integrated Cumulative Error (ICE) quantity on $\mathcal{M}_*$ defined for all $(f_1, f_2) \in \mathcal{M}_*^2$ by

$$ICE(f_1, f_2) := \frac{1}{2} \int_{\mathbb{R}} |F_1(x) - F_2(x)| dF_{\mathcal{M}_*}(x), \quad x \in \mathbb{R}, \tag{5}$$

where $F_i(x) = \int_{-\infty}^{x} f_i(t)dt$, $i = 1, 2$, and $F_{\mathcal{M}_*} := 1/(J+1) \sum_{j=0}^{J} F_{j,*}$ with the convention $F_{0,*} = F_0$. To differentiate the behavior of the method when f_0 truly belongs to the family $\mathcal{M}$ or the contrary, we propose to introduce two additional assumptions.

Assumption (S1). The density f_0 does not belong to the collection $\mathcal{M}$, or equivalently $\mathcal{K}(f_0, f_{j,*}) > 0$ for all $j \in \mathcal{J}$.

Assumption (S2). The collection $\mathcal{M}$ contains the true density f_0 , *i.e.* there exists a unique j_0 and a unique parameter $\vartheta_0 \in \Theta_{j_0}$ such that

$$f_0(x) = f_{j_0}(x, \vartheta_0) \quad x \in \mathbb{R},$$

or equivalently $\mathcal{K}(f_0, f_{j,*}) = 0$, if and only if $j = j_0$ anf $\vartheta_{j,*} = \vartheta_0$.

To use a QML based approach for choosing the most appropriate model, we suggest to select among the collection of models $\mathcal{M}$ defined in (1), the one, with label $j_* \in \mathcal{J}$, that minimizes the ICE distance to f_0, *i.e.*

$$j_* := \arg\min_{j \in \mathcal{J}} ICE(f_{j,*}, f_0). \tag{6}$$

Note that under **(S2)**, we have $j_* = j_0$.

In order to implement this procedure we introduce, for all $x \in \mathbb{R}$ and $j \in \mathcal{J}$, the empirical and smoothed versions of the cdfs involved in (5), i.e.

$$\bar{F}_0(x) := \frac{1}{n} \sum_{q=1}^{n} \mathbb{I}_{X_q \leq x}, \quad \hat{F}_j(x) := \int_{-\infty}^{x} \hat{f}_j(t)dt, \quad \text{and}$$

$$\bar{F}_j(x) := \frac{1}{n} \sum_{q=1}^{n} \mathbb{I}_{Y_{j,q} \leq x}, \tag{7}$$

where $Y_j := (Y_{j,1}, \ldots, Y_{j,n})$ is an iid sample drawn from $\hat{f}_j$, and denote for convenience $Y_0 := (Y_{0,1}, \ldots, Y_{0,n}) = (X_1, \ldots, X_n)$. An empirical estimator of $ICE(f_{j,*}, f_0)$ is then naturally defined by the ICE statistic

$$\begin{aligned}\widehat{ICE}(f_{j,*}, f_0) &:= \frac{1}{n(J+1)} \sum_{l=0}^{J} \sum_{i=1}^{n} |\hat{F}_j(Y_{i,l}) - \bar{F}_0(Y_{i,l})| \\ &= \frac{1}{n(J+1)} \sum_{l=0}^{J} \sum_{i=1}^{n} \left| \hat{F}_j(Y_{l,(i)}) - \frac{n_{X,Y_l}(i)}{n} \right|, \qquad x \in \mathbb{R}, \quad (8)\end{aligned}$$

where for all $l = 0, \ldots, J$, $Y_{l,(1)} < \cdots < Y_{l,(n)}$, and $n_{X,Y_l}(i) = \#\{X_q \leq Y_{l,(i)}; q = 1, \ldots, n\}$.

Finally we estimate j_* by $\hat{\jmath}$ defined by:

$$\hat{\jmath} := \arg\min_{j \in \mathcal{J}} \widehat{ICE}(f_j, f_0). \quad (9)$$

Alternative distances and their estimators. The natural counterparts of the ICE distance in the statistical literature are the Kolmogorov-Smirnov (KS) and Shannon-Jensen (SJ) distances. For all $(f_1, f_2) \in \mathcal{M}_*^2$ the KS and SJ distances are defined respectively by: i) $KS(f_1, f_2) := \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)|$ and ii) $SJ(f_1, f_2) := \frac{1}{2}[\mathcal{E}(f_1|m) + \mathcal{E}(f_2|m)]$ where $m = 1/2 f_1 + 1/2 f_2$ and $\mathcal{E}(f_i|m) = \int \log(f_i/m) df_i$, $i = 1, 2$. For implementation and comparison perspective we define the natural estimators of the above distances evaluated at each (f_j, f_0), $j \in \mathcal{J}$:

$$KS \text{ statistic: :} \quad \widehat{KS}(f_j, f_0) := \sup_{x \in \mathbb{R}} |\hat{F}_j(x) - \bar{F}_0(x)|, \quad (10)$$

where $\hat{F}_j$ and $\bar{F}_0$ are defined in (7), and

$$SJ \text{ statistic:} \quad \widehat{SJ}(f_j, f_0) := \frac{1}{2} \left[\widehat{\mathcal{E}}(f_j|m) + \widehat{\mathcal{E}}(f_0|m) \right], \quad (11)$$

where

$$\widehat{\mathcal{E}}(f_j|m) := \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{\hat{f}_j(Y_{i,j})}{\hat{m}(Y_{i,j})} \right), \quad \text{and} \quad \widehat{\mathcal{E}}(f_0|m) := \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{\hat{f}_0(X_i)}{\hat{m}(X_i)} \right),$$

with $\hat{f}_j$ defined in (3), $\hat{f}_0(x) := 1/nb_n \sum_{i=1}^{n} K(X_i - x/b_n)$ is a standard Parzen-Rosenblatt kernel density estimate of f_0 with bandwidth parameter $b_n \to 0$ as $n \to +\infty$, and $\hat{m} := 1/2\hat{f}_j + 1/2\hat{f}_0$.

Note that Babu [24] considers similarly, for model selection purposes, a bootstrap estimate of the KS distance obtained basically by replacing in (10) the smooth cdf $\hat{F}_j(x)$ by the scaled $\bar{F}_j(x)$, both defined in (7).

Discussion. It is well known that in case the original dataset X contains outliers, then the empirical cdf $\bar{F}_0$ turns to be a bad estimator of F_0 in the distribution-tails. Practically this leads to observe sometimes excessively high values of the KS statistic when the approximation $\hat{F}_j$ under the modes of f_0 is

rather acceptable. In addition, especially when the sample size is small and the dispersion of f_0 is large, it is not rare to observe "gaps" in the ordered dataset along intervals normally reasonably weighted by f_0. This situation usually leads to local defects of $\bar{F}_0$ which are generally very strongly punished by the KS statistic and could deteriorate even more if $\bar{F}_j$ is used instead of $\hat{F}_j$, as considered in Babu [24]. On the other hand, the advantage of the KS statistic lies in the fact that his computation is completely tuning-parameter free which makes its implementation and interpretation rather straightforward.

Regarding the SJ statistic, it is interesting to notice that it is a much more robust (with respect to outliers and local sampling imperfections) quantity in the sense that it integrates progressively the dissimilarities of f_j and f_0 versus $m = 1/2f_j + 1/2f_0$ proportionnaly to f_j and f_0, smoothing by the way the impact of local defects and vanishing the influence of outliers. The downside of the SJ statistic lies in the fact that it requires the tuning of a bandwidth parameter, which is known to have a big influence on the quality of the estimation (and final interpretation) especially in small range sample sizes.

We attempt, by proposing the ICE statistic, to keep the best properties of the KS and SJ statistics and avoid as much as possible their shortcomings. Indeed, we clearly see in (8) that the cdf difference originally involved in the KS statistic is averaged according to a probability density distribution reflecting equally the fitted distributions $\hat{f}_j$, $j \in \mathcal{J}$. This step should hopefully prevent the ICE statistic to be oversensitive to outliers and local random defects by providing a more *global-fitting* oriented information similarly to the SJ statistic. On the other hand, since the ICE statistic only relies on closed form pdfs and empirical measures based integration, it does not require any tuning parameter which makes its use particularly simple.

Small sample size: A Monte Carlo alternative. When there is no hope, for some technical reasons, to get large observed datasets $x_1^n := (x_1, \ldots, x_n)$, the interpretation of $\widehat{ICE}$ criterion has to be handled carefully. Indeed, the empirical means involved in definition (8), based on the random samples $(Y_{j,1}, \ldots, Y_{j,n})$'s, may show themselves to be sources of stochastic instability and, as a consequence, badly impact the interpretation of the results provided by the $\widehat{ICE}$ criterion. Nevertheless, this stochastic instability can be easily removed. Indeed, denoting by $\hat{\vartheta}_j(x_1^n)$ the QMLE of $\vartheta_{j,*}$, $j \in \mathcal{J}$, based on the observed sample x_1^n, we can figure out to evaluate the accuracy of the fitted pdf $\hat{f}_j(x, \hat{\vartheta}_j(x_1^n))$ with f_0, by generating independently N iid samples of size n, $Y_{1,\ell}^n := (Y_{1,\ell}, \ldots, Y_{n,\ell})$, for $\ell = 1, \ldots, N$, and estimate the mean value of criterion ICE conditionally on $\{X_1^n = x_1^n\}$, *i.e.* $ICE_j(x_1^n) := E.(\widehat{ICE}(\hat{f}_j, f))|\{X_1^n = x_1^n\})$, by the empirical mean

$$\overline{ICE}_j(x_1^n) := \frac{1}{N} \sum_{\ell=1}^{N} \widehat{ICE}_\ell(\hat{f}_j, f), \tag{12}$$

where $\widehat{ICE}_\ell$ corresponds to expression (8) when taking $Y_{1,\ell}^n$ instead of Y_1^n. Moreover since the samples $(Y_{1,\ell}^n)_{1 \le \ell \le N}$ are mutually independent and the random

variables $0 \leq \overline{ICE}_\ell(\hat{f}_j, f) \leq 1$ we have the Central Limit Theorem (CLT):

$$\sqrt{N}(\overline{ICE}_j(x_1^n) - ICE_j(x_1^n)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_j(x_1^n)), \quad \text{as} \quad N \to \infty,$$

which allows to derive classical parametric bootstrap confidence intervals. Note that averaged criteria $\overline{KS}$ and $\overline{SJ}$ can also be obtained by replacing $\widehat{ICE}_\ell$ in (12) by $\widehat{KS}_\ell$ and $\widehat{SJ}_\ell$ respectively. This last approach will be employed to reinforce the stability of the various criteria during the study of models **M3**–**4** in Section 4.1 and also in the real datasets Section 4.3.

3. Assumptions and asymptotic results

For simplicity, we endow the space $\mathbb{R}^s$, $s \geq 1$, with the $\|\cdot\|_s$ norm defined for all $v = (v_1, \ldots, v_s)$ by $\|v\|_s = \sum_{j=1}^s |v_j|$ where $|\cdot|$ denotes the absolute value. To reduce wastefully heavy expressions due to the dependence on s, we omit to mention it by considering, equally on s, $\|\cdot\|_s = \|\cdot\|$.

We introduce now a basic assumption dealing with the resampling step of our method (see Section B.3 for illustrative examples).

Assumption (G). For all $(j,k) \in \mathcal{J} \times \{1, \ldots, K_j\}$ and all $\theta_{j,k} \in \Phi_{j,k}$, there exists a pdf $f_{j,k}$ and an analytic function $\rho_{j,k}(\cdot, \theta_{j,k})$ such that for any random variable $Y_{j,k} \sim f_{j,k}$ we have $\rho_{j,k}(Y_{j,k}, \theta_{j,k}) \sim f_{j,k}(\cdot|\theta_{j,k})$. In addition, there exists a constant C independent from $(j,k) \in \mathcal{J} \times \{1, \ldots, K_j\}$ such that for all $(\theta, \theta') \in \Phi_k^2$ we have

$$|\rho_{j,k}(x, \theta) - \rho_{j,k}(x, \theta')| \leq C[|x| + 1] \times \|\theta - \theta'\|, \quad x \in \mathbb{R}. \tag{13}$$

In addition, we define for all $j \in \mathcal{J}$ and all $k = 1, \ldots, K_j$:

$$\dot{F}_{j,k}(x, \theta) := \left(\frac{\partial F_{j,k}(x, \theta)}{\partial \theta_1}, \ldots, \frac{\partial F_{j,k}(x, \theta)}{\partial \theta_{d_{j,k}}} \right)^T, \quad \theta \in \Phi_{j,k},$$

where $d_{j,k} := \dim(\Phi_{j,k})$.

Assumption (R). For all $j \in \mathcal{J}$ and all $k = 1, \ldots, K_j$, the cdf $F_{j,k}(x, \theta)$ is a continuously differentiable function of $\theta \in \Phi_{j,k}$ for each $x \in \mathbb{R}$. Moreover, we suppose that there exists a constant $M > 1$ such that

$$\sup_{x \in \mathbb{R}, \theta \in \Phi_{j,k}} \|\dot{F}_{j,k}(x, \theta)\| < M, \quad j \in \mathcal{J} \quad \text{and} \quad k = 1, \ldots, K_j.$$

Note that fully detailed verification of assumptions **(R)** and **(G)** is done in Appendix A for the Normal and Gumbel distributions.

The following theorems establish respectively the metric property of the ICE quantity and the asymptotic behavior of $\widehat{ICE}$ and $\hat{\jmath}$ when the sample size n goes to infinity.

Theorem 1. *Under assumption (R) the ICE quantity is a distance on the finite collection $\mathcal{M}_*$ (inducing a total ordering), i.e. for all pdf f_j, $j = 1, 2, 3$ belonging to $\mathcal{M}_*$ we have*

i) $ICE(f_1, f_2) \geq 0$, *and* $ICE(f_1, f_2) = 0 \Leftrightarrow f_1 = f_2$ λ*-a.e. (definite positiveness),*
ii) $ICE(f_1, f_2) = ICE(f_2, f_1)$ *(symmetry),*
iii) $ICE(f_1, f_3) \leq ICE(f_1, f_2) + ICE(f_2, f_3)$ *(subadditivity).*

Theorem 2. *i) If all the parametric mixture models belonging to the collection $\mathcal{M}$ satisfy conditions (A1–6), given in Appendix A, and assumptions (NN, S1 or S2, G, R) hold, then for all $j \in \mathcal{J}$ we have*

$$\sqrt{n}\left|\widehat{ICE}(f_j, f_0) - ICE(f_j, f_0)\right| = O_P(1).$$

ii) Under (S1), if conditions (A1–6) and assumptions (NE, G, R) hold, then the ICE criterion defined in (9) is quasi-consistent in Probability, i.e.

$$P(\hat{\jmath} = j_*) \to 1, \quad as \quad n \to \infty. \tag{14}$$

iii) Under (S2), if conditions (A1–6) and assumptions (NE, G, R) hold, then the ICE criterion defined in (9) is consistent in Probability, i.e.

$$P(\hat{\jmath} = j_0) \to 1, \quad as \quad n \to \infty. \tag{15}$$

The proofs of Theorem 1 and 2 are relegated to the Appendix A.

4. Simulation study

4.1. Large sample Monte Carlo study

In order to study the qualitative finite sample properties of the ICE criterion compared to the KS and SJ criteria, we propose to test these criteria on two families of benchmark examples: i) mixtures of Normals and at most one centered Laplace and ii) mixtures of Normals and Gumbels.

Notations and examples i) We denote by $N[r](\vartheta_r)$ a generic mixture of r Normal distributions where $\vartheta_r := (\pi, (\theta_i)_{i=1}^r)$ is composed by $\pi = (\pi_1, \ldots, \pi_{r-1})$ the mixture weights vector, and $(\theta_i)_{i=1}^r$ the collections of parameters corresponding to the Normals.

We also denote by $NL[q](\vartheta_{q-1}, \lambda)$, a generic mixture of $q-1$ Normal distributions and one centered Laplace distribution, where $\vartheta_{q-1} := (\pi, (\theta_i)_{i=1}^{q-1}, \lambda)$ is composed by $\pi = (\pi_1, \ldots, \pi_{q-1})$ the mixture weights vector, $(\theta_i)_{i=1}^{q-1}$ the collections of parameters corresponding to the Normals and $\lambda \in \mathbb{R}^+ \setminus \{0\}$ the scaling parameter of the centered Laplace distribution which pdf is given by $f_{\mathcal{L}(\lambda)}(x) = \lambda/2 \exp(-\lambda|x|)$, $x \in \mathbb{R}$. We then consider the benchmark models:

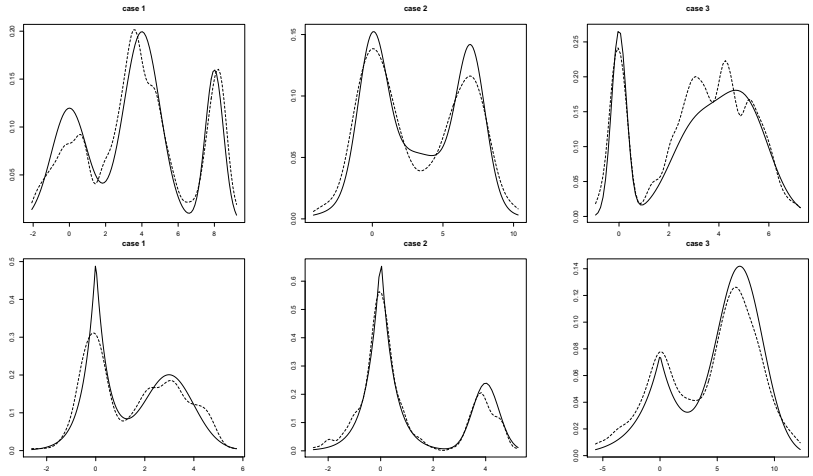

FIG 1. *First row, respectively second row, plot of model* **M1**, *respectively* **M2**, *probability density function (solid) and a representative kernel density estimate (dashed) based on* $n = 300$ *observations under case* $j = 1, 2, 3$ *(columns).*

$$
\begin{aligned}
&\mathbf{M1} : N[3](\vartheta_3) \text{ under}\\
&\qquad \text{case } 1: \quad \pi = (0.2, 0.5), \theta_1 = (8, 0.5), \theta_2 = (4, 1), \theta_3 = (0, 1),\\
&\qquad \text{case } 2: \quad \pi = (0.4, 0.3), \theta_1 = (7, 1), \theta_2 = (3, 3), \theta_3 = (0, 1),\\
&\qquad \text{case } 3: \quad \pi = (0.4, 0.3), \theta_1 = (5, 1), \theta_2 = (3, 1), \theta_3 = (0, 0.3).\\
&\mathbf{M2} : NL[2](\vartheta_1, \lambda) \text{ under}\\
&\qquad \text{case } 1: \quad \pi = 0.5, \theta_1 = (3, 1), \lambda = 2,\\
&\qquad \text{case } 2: \quad \pi = 0.3, \theta_1 = (4, 0.5), \lambda = 2,\\
&\qquad \text{case } 3: \quad \pi = 0.7, \theta_1 = (7, 2), \lambda = 0.5.
\end{aligned}
$$

For model **M1**, we consider three models in competition labeled by j: when $j = 1, 2, 3$ we consider respectively $N[2]$, $N[3]$, $NL[3]$, which gives $j_0 = 2$ for **M1**.

For model **M2**, we consider three models in competition labeled by j: when $j = 1, 2, 3$ we consider respectively, $N[2]$, $NL[2]$, $N[3]$, which gives $j_0 = 2$ for **M2**.

Remarks. The interest of comparing mixtures of Normals and Normals with one centered Laplace distribution lies in the fact that the EM algorithms linked to these models have closed iterative forms allowing a fast investigation on a varied collection of models. In addition, it is worth to notice that they also provide very challenging modeling problems when a Laplace component has, for example, to be validated against a peaky Normal component (see the plots in Fig. 1).

ii) We denote by $NG[r, q](\vartheta_{r,q})$ a generic mixture of r Normal and q Gumbel distributions where $\vartheta_{r,q} := (\pi_{r,q}, (\theta_i)_{i=1}^r, (\phi_j)_{j=1}^q)$ is composed by $\pi_{r,q} := (\pi_1, \ldots, \pi_{r+q})$ the mixture weights vector, and $(\theta_i)_{i=1}^r$, respectively $(\phi_j)_{j=1}^q$, the collections of parameters corresponding to the Normal and Gumbel distributions

respectively. We then consider the benchmark models:

$$\mathbf{M3}: NG[1,2](\vartheta_{1,2}), \text{ with } \pi_{1,2}=(1/2,1/4), \theta_1=(2,0.5), \phi_1=(0,2), \phi_2=(4,3).$$
$$\mathbf{M4}: NG[2,1](\vartheta_{2,1}), \text{ with } \pi_{2,1}=(1/3,1/3), \theta_1=(0,1), \theta_2=(2,0.5), \phi_1=(3,2).$$

For these two models we consider three models in competition labeled by j: when $j=1,2,3$ we consider respectively $NG[1,1]$, $NG[2,1]$, $NG[1,2]$ which gives $j_0=3$ for **M3** and $j_0=2$ for **M4**.

EM and GEM algorithms i) For the fitting of mixtures of Normals and at most one centered Laplace distribution, we use the standard closed form Expectation-Maximization (EM) algorithms (the MLE for the centered Laplace distribution being explicit) as defined in Dempster *et al.* [21] p. 4. The EM algorithms are initialized at arbitrarily values supposed to reflect the guess of a practitioner and run until the difference between two successive iteration outputs is inferior to 10^{-9} or when one of the estimated proportion mixture is less than 0.1 to insure the no-nested model condition **(NN)** especially in the over-parametrization case $N[3]$ under model **M2**.

ii) For the fitting of mixtures of Normal and Gumbel distributions, unfortunately there do not exist closed form solutions for updating the parameters of the Gumbel components during the maximization step of the EM algorithm. This concern forces us to implement a generalized EM algorithm (GEM), see Dempster et al. [21], p. 7, in which a discretized search in the neighborhood of the current parameters is performed in order to ensure the increase of the log-likelihood function. The search is performed over a grid of uniformly spaced points and the weighted log-likelihood of each component is calculated at each of these points. The limits of the grid are defined as ± 0.1 with increments of 0.01 for the location parameter and ± 0.01 with increments of 0.001 for the scale parameter. The parameters of a component of the mixture are updated, if necessary, by identifying the argument of the maximum of the weighted log-likelihood evaluated over the grid of points specified. The support of the area/grid over which the weighted log-likelihood of a component is maximized, is also updated along with the updated parameters. The weights of the components of the model are updated in the successive iteration based on the updated values of the parameters. The method is assumed to have converged when the global log-likelihood of the model being estimated varies within a tolerance of 10^{-10} for fifty successive iterations.

In our Monte Carlo simulation experiment, we generate 100 samples of size n from models **M1–4** and their different cases. For each sample, we compute the QMLE for the various models in competition, estimate their distances $\widehat{ICE}$, $\widehat{KS}$ and $\widehat{JS}$ when considering **M1–2** and $\overline{ICE}$, $\overline{KS}$ and $\overline{JS}$ (based on $N=100$ resamplings) when considering **M3–4**, to the true model, and pick the models minimizing respectively these quantities. In Tables 1–3 we report, for different values of n, the selection rate obtained by each model when using these criteria. For clarity purposes, the entries of Tables 1–3 corresponding to the true model are colored in green.

TABLE 1

Within brackets selection rates of competing models $j = 1, 2, 3$ associated respectively to $\widehat{ICE}$, $\widehat{KS}$, $\widehat{SJ}$ under **M1** *and cases 1, 2, 3*

Model j	n	case 1	case 2	case 3
$j = 1$	50	$\{0, 0, 0\}$	$\{0, 11, 0\}$	$\{4, 10, 0\}$
$j = 2$	50	$\{82, 69, 57\}$	$\{77, 63, 57\}$	$\{74, 54, 60\}$
$j = 3$	50	$\{18, 31, 43\}$	$\{23, 26, 43\}$	$\{22, 36, 40\}$
$j = 1$	100	$\{0, 0, 0\}$	$\{1, 8, 0\}$	$\{3, 10, 0\}$
$j = 2$	100	$\{90, 64, 69\}$	$\{84, 73, 61\}$	$\{78, 52, 61\}$
$j = 3$	100	$\{10, 36, 31\}$	$\{15, 19, 39\}$	$\{19, 38, 39\}$
$j = 1$	200	$\{0, 0, 0\}$	$\{0, 8, 0\}$	$\{0, 4, 0\}$
$j = 2$	200	$\{95, 79, 83\}$	$\{87, 67, 61\}$	$\{87, 68, 68\}$
$j = 3$	200	$\{5, 21, 17\}$	$\{13, 25, 39\}$	$\{13, 28, 32\}$

TABLE 2

Within brackets selection rates of competing models $j = 1, 2, 3$ associated respectively to $\widehat{ICE}$, $\widehat{KS}$, $\widehat{SJ}$ under **M2** *and cases 1, 2, 3*

Model j	n	case 1	case 2	case 3
$j = 1$	100	$\{39, 34, 27\}$	$\{34, 29, 33\}$	$\{32, 19, 25\}$
$j = 2$	100	$\{22, 24, 30\}$	$\{62, 61, 65\}$	$\{46, 52, 51\}$
$j = 3$	100	$\{39, 36, 49\}$	$\{4, 10, 2\}$	$\{22, 30, 23\}$
$j = 1$	200	$\{14, 23, 17\}$	$\{18, 14, 25\}$	$\{18, 17, 20\}$
$j = 2$	200	$\{38, 36, 29\}$	$\{82, 82, 75\}$	$\{65, 63, 57\}$
$j = 3$	200	$\{10, 36, 31\}$	$\{0, 4, 0\}$	$\{17, 20, 23\}$
$j = 1$	500	$\{2, 3, 6\}$	$\{5, 6, 9\}$	$\{16, 20, 18\}$
$j = 2$	500	$\{53, 49, 40\}$	$\{95, 92, 91\}$	$\{75, 71, 71\}$
$j = 3$	500	$\{45, 48, 54\}$	$\{0, 2, 0\}$	$\{9, 9, 11\}$
$j = 1$	1000	$\{2, 1, 1\}$	$\{0, 2, 0\}$	$\{3, 10, 8\}$
$j = 2$	1000	$\{63, 59, 41\}$	$\{100, 95, 93\}$	$\{95, 89, 90\}$
$j = 3$	1000	$\{35, 40, 58\}$	$\{0, 3, 7\}$	$\{2, 1, 2\}$

Comments on Tables 1–3

Table 1. It is interesting to notice first that case 1, which is a clearly designed 3 components Normal mixture, is well identified by all the criteria despite the small sample size under consideration. For case 2 and 3, which have been deliberately setup to mimic a two components mixture model, we observe that $\widehat{KS}$ falls in that trap about 10% of the trials when the other criteria do not. Overall the $\widehat{ICE}$ criterion outperforms very clearly the other criteria (15% to 20% more efficient) and $\widehat{JS}$ seems to have more difficulties in distinguishing a (peaky) Laplace component vs a Normal component due to the kernel smoothing step involved in its definition (11).

Table 2. The main difficulty in that collection of cases is that our criteria have to "decide" at some point if the zero-centered peaky component corresponds to a Laplace distribution or a peaky gaussian distribution (see Fig. 1 second row). This explains the fact that much more data is needed to get a valid conclusion. In case 1, where the Laplace component is strongly weighted and widely overlapped by the Normal component, we observe that there is no really valid conclusion under $n = 500$. For case 1, the $\widehat{ICE}$ criterion outperforms sligthly the other criteria when $\widehat{JS}$ struggles more in rejecting the over-fitted model $j = 3$. For

TABLE 3
Within brackets selection rates of competing models $j = 1, 2, 3$ associated respectively to $\overline{ICE}$, $\overline{KS}$, $\overline{SJ}$ under **M3 and M4**

Model j	n	**M3**	**M4**
$j = 1$	100	$\{13, 25, 1\}$	$\{2, 10, 5\}$
$j = 2$	100	$\{36, 43, 56\}$	$\{54, 42, 58\}$
$j = 3$	100	$\{51, 32, 43\}$	$\{44, 48, 37\}$
$j = 1$	300	$\{13, 21, 2\}$	$\{1, 6, 2\}$
$j = 2$	300	$\{29, 37, 52\}$	$\{70, 66, 55\}$
$j = 3$	300	$\{58, 43, 46\}$	$\{29, 28, 43\}$
$j = 1$	1000	$\{8, 14, 2\}$	$\{0, 1, 0\}$
$j = 2$	1000	$\{17, 45, 46\}$	$\{87, 75, 61\}$
$j = 3$	1000	$\{75, 41, 52\}$	$\{13, 24, 39\}$

cases 2 and 3, we can observe in Fig. 1 that the kernel density estimators fit pretty well the probability density curve of the mixtures, which is reflected in the very good selection rate obtained by all the criteria (we even obtained a 100% of success for $\widehat{ICE}$ during our $N = 100$ experiments when considering case 2 with $n = 1000$).

Table 3. The interesting aspect in model **M3** is that its third Gumbel component has a very heavy right-side tail compared to the cases considered in **M1** and **M2**, when model **M4** allows to illustrate the behavior of our method in a kind of well designed situation (**M1**–**2** like) including a more reasonably tailed Gumbel distribution.

Since the datasets generated from **M3** are basically very spread-out, we encounter very often situations, especially for small sample sizes, where nearest neighbors are relatively far from each over. This concern impacts very badly the quality estimation of the cumulative distribution and the probability density function (see left side of Fig. 2) as it is discussed at the end of Section 2. In this very challenging example, we observe that criterion $\widehat{JS}$ excludes very early the underparametrized model $NG[1, 1]$ while criteria $\widehat{ICE}$ and $\widehat{KS}$, both based on the empirical cdf, struggle in rejecting it (still about 10% of $NG[1, 1]$ models are selected when $n = 1000$). On the other hand, we have to recognize the good behavior of the $\widehat{ICE}$ criterion which actually is the only one able to detect significantly the right model by having a success rate of 75% for $n = 1000$ when other criteria stabilize their success rates around 30–50% for $n = 100, 300, 1000$ which are values close to the prior probability of picking the right model.

The study of model **M4** confirms the major influence of the distribution-tails on the finite sample size performances of the various criteria. Indeed, the fact that the mixture model **M4** has only one Gumbel component with a scale parameter equal to 2, when there was two Gumbel components with scale parameters equal to 2 and 3 in **M3**, allows a better estimation of the cdf and pdf (compare both sides of Fig. 2). As a consequence, the results displayed in Table 3 corresponding to model **M4** are clearly better than those corresponding to model **M3**. Again we have to recognize the superiority of $\widehat{ICE}$ on its competitors since it gets respectively 70% and 87% of success rate when $\widehat{KS}$, respectively $\widehat{JS}$, only reach 66% and 75%, respectively 55% and 61%, for $n = 300$ and $n = 1000$.

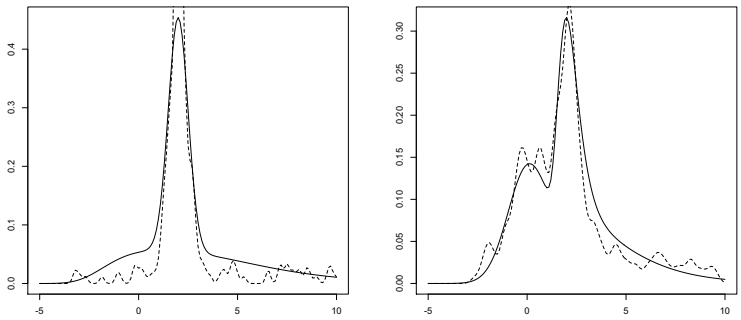

FIG 2. *Left side, respectively right side, plot of model* **M3**, *respectively* **M4**, *probability density function (solid) and a representative kernel density estimate (dashed) based on* $n = 300$ *observations.*

4.2. Simulation methodology

We attempt to relate the distribution of fatigue life typically observed in materials through numerical simulations that take into account the effect of the microstructure of the material i.e., by modeling individual grains. A schematic of the microstructure generated is shown in Fig. 3. The different color codes of each grain in Fig. 3 indicate that different orientations are assigned to each grain. For simplicity, we have considered an idealized grain structure (cubes) subject to monotonic deformation with linear elastic material properties. We have considered the elastic properties of a material with a face centered cubic crystalline structure (austenitic stainless steel at room temperature) for the purpose of these simulations. The single crystal elastic constants have been adopted from [26].

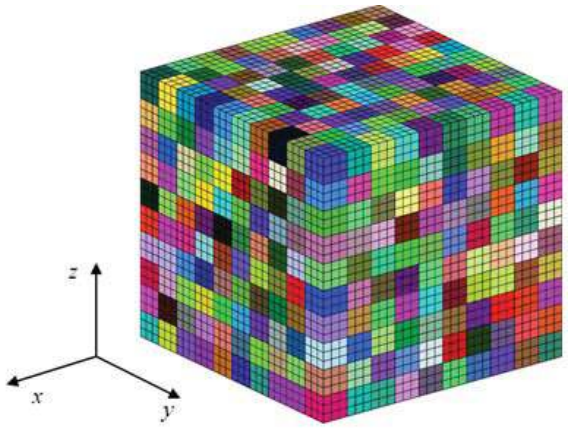

FIG 3. *Realization of the idealized microstructure used for the finite element simulations.*

The distribution of fatigue crack initiation potencies for different microstructures is usually quantified by the distribution of fatigue indicator parameters [12], which consider grain averaged parameters that include microscale cyclic plasticity obtained from computational experiments. The distribution of the microscale cyclic plasticity in a computational volume has been shown to vary with the imposed deformation amplitude [25]. In the present context of using linear elastic material models subject to tension, i.e. no cyclic plasticity, the simulations would qualitatively approach the high cycle and very high cycle fatigue regime where the plasticity averaged over the volume of the computational cell would be very low. Further, here we characterize the distribution of the extreme values of the shear stress resolved along specific crystallographic (slip) directions, averaged over a grain. Thus, the extremal value of this grain averaged resolved shear stress would be an indicator of the potency of fatigue crack formation in a microstructural realization. Also, while inter-granular interactions are being considered here, the effect of grain size distribution will not be accounted for due to the simplified microsctructure assumed. The computational cell shown in Fig. 3, serves as a statistical volume element (SVE) i.e. a single computational cell is not large enough to capture all the statistical variations of the fatigue crack formation potency generated due to microstructural variations. Thus, multiple realizations are required to obtain the distribution of the extremal values of the fatigue crack formation potency.

The boundary value problem is solved by the finite element method [27] using the software ABAQUS [28]. We explore the distribution of the extreme values of the grain averaged resolved shear stresses for different boundary conditions and computational cell size. In this study, the computational cell is subject to uniaxial tension using two boundary conditions viz. free surface boundary conditions and generalized periodic boundary conditions [12, 29]. The computational cell size is varied from 5 grains in each dimension to 10 grains in each dimension with 27 8-node brick elements with linear interpolation and reduced integration (C3D8R) finite elements for each grain. For each microstructural realization, the grain averaged values of shear stress along all possible slip directions at the peak tensile strain (0.2%) form the random variable, from which the extremal value is selected. The collection of such extreme values for multiple realizations is referred to as the real dataset in the subsequent subsection.

4.3. Application to a real data set: Results and discussion

Let consider a set of r^3 *Input* random variables

$$U(r) := \left\{U_{k,l,m},\ (k,l,m) \in \mathcal{S}_r{}^3\right\},$$

with $\mathcal{S}_r := \{1,\ldots,r\}$, valued in a measurable space $(U,\mathcal{B}_U)$, where for each triplet $(k,l,m) \in \mathcal{S}_r^3$, $U_{k,l,m}$ represents the resolved shear stresses averaged over a grain in a particular realization of the numerical experiments. Thus, the

set of grain averaged resolved shear stresses obtained for all the grains from a realization are given by:

$$R_{k,l,m} := \xi(k, l, m, U(r), \partial B),$$

where $\xi(\cdot)$ is treated as a black-box function whose entries are the location of a grain in the computational cell (k, l, m), the random input in the block $U(r)$, and the boundary condition ∂B. The input $U(r)$ is considered random, since each computational realization has randomly assigned grain orientations. We denote by

$$X(r) := \max_{1 \le k,l,m \le r} R_{k,l,m}.$$

Let us suppose that we repeat the experiment n times and collect n extreme values $X_1^n(r) := (X_1(r), \ldots, X_n(r))$, which forms the dataset. Our aim is to statistically model the distribution of $X(r)$'s for different levels of discretization r, and boundary conditions ∂B. For our simulations, (k, l, m) vary from 5 to 10 yielding a computational cell size of 125 to 1000 grains and $n = 300$. Here, since $R_{k,l,m}$ is the set of grain averaged resolved stresses for all the grains in a given realization of the microstructure, the dataset $X_1^n(r)$ represents the collection of the extreme values of the grain averaged resolved shear stress (indicative of fatigue crack formation potency) for n realizations of the microstructure.

The method of identifying the best model to estimate the underlying mixture of $X_1^n(r)$ is similar to the one outlined in the previous section i.e., we start by assuming various possible models that would describe the underlying mixture and estimate the parameters of the model using the EM algorithm. However, the tolerance used as criterion for the convergence of the method, is changed to 10^{-6}. A comparison of the pdfs of the dataset and a random variable of size $n = 300$ generated from the estimated parameters, using nonparametric density estimates, is shown in Figs. 4 and 5, for different assumed mixture models. It can be seen that the EM algorithm estimates the parameters of the mixture model accurately, not only for the location and scale parameters for the dominant distribution, but also the parameters of the smaller components of the mixture that present as perturbations (bumps). From the nonparametric density estimates (Figs. 4 and 5), it can be seen that the dataset is a mixture of many distributions (possibly 3–4).

We recall that the dataset is a collection of the extreme values of the grain averaged resolved shear stresses in a computational cell for multiple microstructural realizations. In the present case, the interaction with neighboring grains, which have different crystallographic orientations, coupled with the boundary conditions introduce perturbations from homogeneous deformation of the assembly. This could be the primary reason of the convergent extreme value distribution being corrupted which manifests as a mixture of distributions. Nevertheless, assuming the existence of a model that has mixture of distributions, an approach to order the different possible mixture models is presented next.

Since the exact model describing the underlying mixture is not known a priori, a collection of competing models is considered to obtain the best fit to

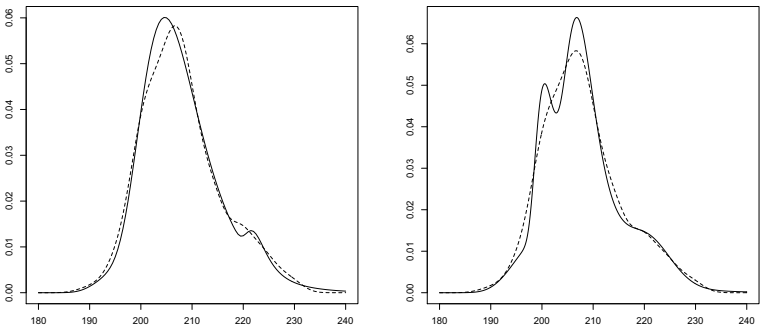

FIG 4. *Left side, respectively right side, plot of the fitted mixture model (solid) when model (a) $NG[0,3]$, respectively (b) $NG[1,3]$, is assumed for the dataset $X_1^{300}(5)$ along with a nonparametric density estimate (dashed).*

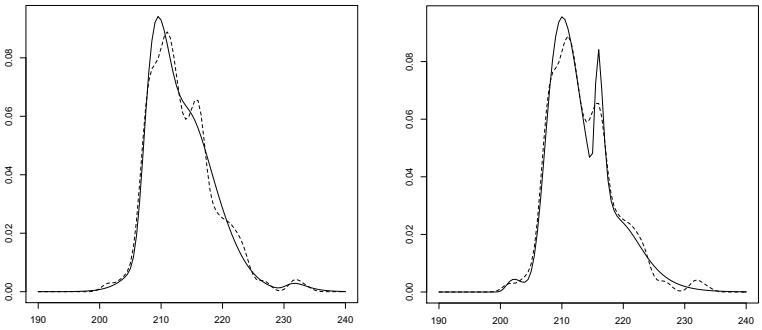

FIG 5. *Left side, respectively right side, plot of the fitted mixture model (solid) when model (a) $NG[1,3]$, respectively (b) $NG[0,4]$, is assumed for the dataset $X_1^{300}(10)$ along with a nonparametric density estimate (dashed).*

the dataset $X_1^n(r)$. The models in competition are model 1: $NG[0,1]$, model 2: $NG[1,1]$, model 3: $NG[0,2]$, model 4: $NG[1,2]$, model 5: $NG[0,3]$, model 6: $NG[2,2]$, model 7: $NG[1,3]$ and model 8: $NG[0,4]$. The parameters of each model are estimated using the EM algorithm described in Section 4.1. Subsequent to determining the parameters of each model $j \in \{1,\dots,8\}$, a random sample of size n is generated to compute the criterion $\widehat{\overline{ICE}_j}$ from the dataset $X_1^n(r)$. This process is repeated for $N = 100$ samples in order to compute the Monte Carlo estimate $\overline{ICE}_j(r) := \overline{ICE}_j(X_1^n(r))$, see definition (12), based on the dataset $X_1^n(r)$ for each model j. This approach assumes significance in the current context of a small sample size ($n = 300$) relatively to the observation domain which is approximately equal to $[200, 240]$. The model that most accurately fits the dataset is assessed based on the minimum $\overline{ICE}$ for all the models considered. This approach of considering models in competition allows the ranking of the different models considered. The $\overline{ICE}_j$ for all the models $j \in \{1,\dots,8\}$ and different sizes of the computational cell are plotted in Fig. 6.

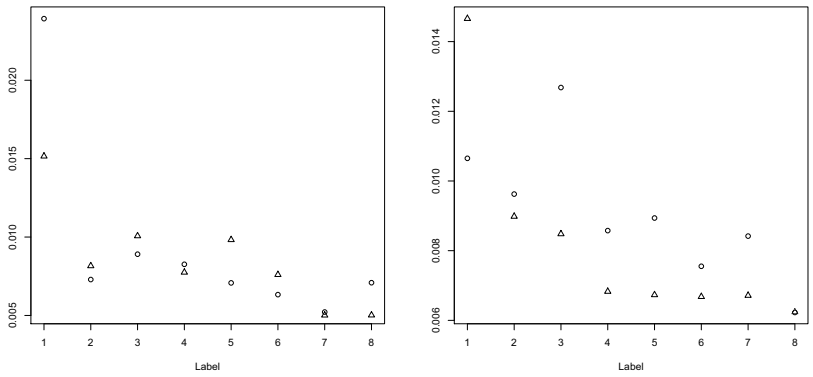

FIG 6. *Variation of* $\overline{ICE}$ *criterion for different assumed mixture models for computational cells consisting of 125 grains (left side) and 1000 grains (right side). Points depicted by triangles and circles correspond to free surface BC and periodic BC respectively.*

The values of the $\overline{ICE}$ for all the models and combinations of computational cell size and Boundary Conditions (BC) are listed in Table 4 (at the end of Appendix B).

From Fig. 6, it is apparent that a better fit to the dataset is obtained by using a higher number of distributions. In almost all cases, a single Gumbel distribution provides the least accurate fit to the dataset. This is particularly important since the fatigue life distributions are often fitted to a single extreme value distribution [16, 17, 18, 19]. We have previously surmised that the departure of the dataset from a single extreme value distribution is due to perturbations induced by inter-granular interactions and boundary conditions. It is, however, noteworthy that a single Gumbel distribution shows a marked increase of accuracy in modeling the dataset for the larger assembly. Thus, it can be argued that the the periodic BC might partially mitigate the perturbations introduced due to the boundary conditions, thus increasing the accuracy of one extreme value distribution in describing the dataset.

It is also noteworthy that the same mixture models predict the distribution most accurately for both boundary conditions for a given size of the computational cell – model 7 ($NG[1,3]$) for the 125 grain assembly and model 8 ($NG[0,4]$) for the 1000 grain assembly. We conjecture that a normal distribution acts as an attractor for the 125 grain assembly since the extreme values are being sampled from a smaller population of outliers.

Conjecture on multi-regime model. One possible physical explanation regarding the mixture of Gumbel and Normal distributions selected by the ICE criterion, should be the existence of a multi-regime model. Let us define, for simplicity, a generic model with L so-called *regimes*. We conjecture that there exist two types of attraction domains, *i.e.*, the so-called *Gumbel attraction domain* $\mathcal{A}_{Gumbel} = \{\mathcal{C}_i\}_{i=1,\ldots,L_1}$ and *Normal attraction domain* $\mathcal{A}_{Normal} = \{\mathcal{C}_i\}_{i=L_1+1,\ldots,L}$, justified as follows:

- given $\{U(r) \in \mathcal{C}_i\}$, $i = 1, \ldots, L_1$, the set of resolved shear stresses from a numerical realization, $R_{k,l,m}$, $(k,l,m) \in \mathcal{S}^3$, is quasi-homogeneous (mixing enough and marginally approximately equally distributed) in a such a way that the Dabrowski's [30] convergence theorem for extreme value of mixing random sequences applies, *i.e.* $\mathcal{L}(X(r)|U(r) \in \mathcal{C}_i) \simeq \mathcal{G}(\mu_i, \beta_i)$, $i = 1, \ldots, L_1$.
- given $\{U(r) \in \mathcal{C}_j\}$, $j = L_1+1, \ldots, L$, the set of resolved shear stresses from a numerical realization, $R_{k,l,m}$, $(k,l,m) \in \mathcal{S}^3$, is not quasi-homogeneous, as it is supposed to hold under $\mathcal{A}_{Gumbel}$, and contains a small collection of Gaussian outliers. In such a case, the extreme values being taken from among a small population of Gaussian random variables, and the convergence for extreme values of Gaussian samples being known to converge at the (possibly) slowest rate, *i.e.* $O(\log(n)^{-1})$, see Han and Ferreira [31], we suppose that these maxima are themselves approximately Gaussian, *i.e.* $\mathcal{L}(X(r)|U(r) \in \mathcal{C}_i) \simeq \mathcal{N}(m_i, \sigma_i^2)$, $j = L_1 + 1, \ldots, L$.

In conclusion, we have the following *mixture* model structure:

$$\begin{aligned} X(r) \quad &= \quad \sum_{i=1}^{L_1} \underbrace{\max_{1\le k,l,m\le r} R_{k,l,m}}\, \underbrace{\mathbb{I}_{U(r)\in\mathcal{C}_i}} + \sum_{j=L_1+1}^{L} \underbrace{\max_{1\le k,l,m\le r} R_{k,l,m}}\, \underbrace{\mathbb{I}_{U(r)\in\mathcal{C}_j}} \\ &\overset{distribution}{\simeq} \quad \sum_{i=1}^{L_1} \quad f_{\mathcal{G}(\mu_i,\beta_i)} \quad \times \quad \pi_i \quad + \sum_{j=L_1+1}^{L} \quad f_{\mathcal{N}(m_j,\sigma_j^2)} \quad \times \quad \pi_j, \end{aligned}$$

where $\pi_j = P(U(r) \in \mathcal{C}_j)$, $j = 1, \ldots, L$.

Comment. Recall that the EM algorithm computes for each $X_i(r)$, $i = 1, \ldots, n$, the probabilities to belong to the groups characterized by conditions $(\mathcal{C}_j)_{1\le j\le L}$, providing a very useful exploratory tool to posteriorly investigate the particular structure of each block based on its extreme value observation.

5. Summary and conclusions

In this work, we have addressed the problem of identifying the model that best fits a dataset among a given finite collection of non-nested mixture models. The parameters of the models are identified using EM or GEM algorithms. A GEM algorithm, involving a maximization step based on a neighborhood screening method, has been especially implemented to deal with mixtures of Normal and Gumbel distributions.

To test the accuracy of the models in describing the mixture, a novel metric, the Integrated Cumulative Error (ICE) distance has been defined. The estimator of the ICE distance has been shown to be asymptotically consistent and practically more efficient in identifying the correct model than commonly used approaches, such as the ones based on the Kolmogorov-Smirnov or Shannon-Jensen statistics.

The approach developed in this paper is used to identify a mixture that fits best the distribution of indicators of fatigue crack formation potency (grain av-

eraged resolved shear stresses). The observations show that a mixture model characterizes the distribution of interest more accurately than a single extreme value distribution, which is the hypothesis commonly assumed. It is to be noted that the methods developed in this work have not been applied, yet, to experimentally observed fatigue life distributions. The solution in this case is direct, since observations of significant deviations from an assumed unique extreme value distribution that characterizes the fatigue life distributions are widely found in literature. The use of computational models is motivated from the numerous constraints of performing a large number of experiments in the regime of very high cycle fatigue life of a material. However, correlating distributions from computational models with experimental observations for the same material would improve predictions of fatigue life distributions. Nevertheless, the use of the methods developed in this work would better characterize the overall fatigue life distribution of a material, which would be informative for minimum life based design approaches. Finally, since the approach developed here is general to the number and types of distributions that form a mixture, it can be used for characterizing fatigue life distributions through multiple failure mechanisms as well. Thus, the statistical framework developed in this work could contribute significantly to current design approaches.

Acknowledgments

The authors are grateful for support of the Carter N. Paden, Jr. Distinguished Chair in Metals Processing.

Appendix A

Let us recall some basic results on the empirical cdf $\bar{F}(x) = 1/n \sum_{k=1}^{n} \mathbb{I}_{X_k \leq x}$. From well known results on empirical processes (see, *e.g.*, Shorack and Wellner [34]), for a general distribution function f_0, we have

$$\sqrt{n}\|\bar{F} - F\|_\infty = O_P(1). \tag{16}$$

Lemma 1. *Under assumption (R) we have, for all* $j \in \mathcal{J}$,

$$\|\hat{F}_j - F_{j,*}\|_\infty = O(\|\hat{\vartheta} - \vartheta_*\|).$$

Proof. Let us consider

$$\begin{aligned}
|\hat{F}_j(x) - F_{j,*}(x)| &\leq \sum_{k=1}^{K} \left(\hat{\pi}_k |F_{j,k}(x, \hat{\theta}_k) - F_{j,k}(x, \theta_{k,*})| + F_{j,k}(x, \theta_*) |\hat{\pi}_{j,k} - \pi_{j,k}| \right) \\
&\leq \sum_{k=1}^{K} \left(\sup_{x \in \mathbb{R}, \theta_k \in \Theta_k} \|\dot{F}_k(x, \theta_k)\|_k \times \|\hat{\theta}_k - \theta_{k,*}\| + |\hat{\pi}_{j,k} - \pi_{j,k}| \right) \\
&\leq \max(M, 1) \|\hat{\vartheta} - \vartheta_*\|.
\end{aligned} \tag{17}$$

where the last inequality holds because of assumption **(R)**. □

Proof of Theorem 1. i) Let us suppose that $f_1 \neq f_2$ on a set $\mathcal{E}$ with $\lambda(\mathcal{E}) > 0$, then there exists at least one point $x_0 \in \mathbb{R}$ such that $F_1(x_0) \neq F_2(x_0)$, and (F_1, F_2) being continuous functions over $\mathbb{R}$, there also exists $\varepsilon > 0$ such that $|F_1(x) - F_2(x)| > 0$ on $]x_0 - \varepsilon, x_0 + \varepsilon[$. To conclude, it can be deduced that

$$ICE(f_1, f_2) \geq \int_{x_0-\varepsilon}^{x_0+\varepsilon} |F_1(x) - F_2(x)| \frac{dF_1 + dF_2}{J+1} > 0.$$

ii) The symmetry property is straightforward by noticing $|F_1 - F_2| = |F_2 - F_1|$ and that $F_{\mathcal{M}_*} \propto \sum_{j=0}^{J} F_{j,*}$ is invariant by permutation of indices.

iii) The subadditivity is a direct consequence of the triangular inequality for the absolute value and the fact that $F_{\mathcal{M}_*}$ equally considers all the F_j belonging to $\mathcal{M}_*$. □

Proof of Theorem 2. i) For simplicity, let us drop the dependence on j in our expression, *i.e.*, $f_{j,*} := f_*$, $F_{j,*} := F_*$, $\hat{f}_j := \hat{f}$, $\hat{F}_j := \hat{F}$. Now, denote $\Psi(\cdot) := |F_*(\cdot) - F_0(\cdot)|$, and consider the following decomposition

$$\Delta(\hat{f}, f) := |\widehat{ICE}(\hat{f}, f) - ICE(f_*, f)| \leq \frac{1}{J+1}(T_1 + \sum_{l=1}^{J} T_2(l)),$$

where for all $l \in \mathcal{J}$,

$$\begin{aligned} T_1 &:= \left| \frac{1}{n} \sum_{i=1}^{n} \left(|\hat{F}(X_i) - \bar{F}_0(X_i)| \right) - E_{f_0}(\Psi) \right|, \\ T_2(l) &:= \left| \frac{1}{n} \sum_{l=1}^{n} \left(|\hat{F}(Y_{i,l}) - \bar{F}_0(Y_{i,l})| \right) - E_{f_{l,*}}(\Psi) \right|, \end{aligned}$$

with $E_{f_0}(\Psi) := \int_{\mathbb{R}} \Psi(x) dF_0(x)$, $E_{f_*}(\Psi) := \int_{\mathbb{R}} \Psi(x) dF_*(x)$. We denote by $\mathcal{F}_n$ the σ-algebra generated by the random variables $(X_1, \ldots, X_n)$.

We note that

$$T_1 \leq D_1 + R_{1,1} + R_{1,2}, \tag{18}$$

where

$$D_1 := \left| \frac{1}{n} \sum_{i=1}^{n} [\Psi(X_i) - E_f(\Psi)] \right|, \quad R_{1,1} := \frac{1}{n} \sum_{i=1}^{n} \left| \hat{F}(X_i) - F_*(X_i) \right|,$$
$$R_{1,2} := \frac{1}{n} \sum_{i=1}^{n} \left| \bar{F}_0(X_i) - F_0(X_i) \right|.$$

According to the central limit theorem, we have $D_1 = O_P(1/\sqrt{n})$, and from (16) and subsequently Lemma 1, we obtain $R_{1,1} \leq \max(M, 1)\|\vartheta_n - \vartheta_*\| = O_P(1/\sqrt{n})$ and $R_{1,2} \leq \|\bar{F}_0 - F_0\|_\infty = O_P(1/\sqrt{n})$. Let us note that $M_1 := D_1 + R_{1,1} + R_{1,2} = O_P(1/\sqrt{n})$.

For the treatment of $T_2(l)$, $l \in \mathcal{J}$, we drop, for simplicity and without loss of generality, the dependence on l in our expressions, *i.e.*, $(Y_1, \ldots, Y_n) := (Y_{1,l}, \ldots, Y_{n,l})$, $f_{l,*} := f_*$, $K := K_l$, and for $k = 1, \ldots, K$, $\rho_k(\cdot,\cdot) := \rho_{k,l}(\cdot,\cdot)$, $(\hat{\pi}_k, \hat{\theta}_k) := (\hat{\pi}_{k,l}, \hat{\theta}_{k,l})$, and $(\pi_{k,*}, \hat{\theta}_{k,*}) := (\pi_{k,l,*}, \hat{\theta}_{k,l,*})$.

We propose to couple now the sample $(Y_1, \ldots, Y_n)$ with a sample $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ which is iid according to f_*, in the following way:

$$\left\{ \begin{array}{l} Y_i = \displaystyle\sum_{k=1}^{K} \mathbb{I}_{\hat{p}_{k-1} < U \leq \hat{p}_k} \rho_k(Z_{k,i}, \hat{\theta}_k), \\ \check{Y}_i = \displaystyle\sum_{k=1}^{K} \mathbb{I}_{p_{k-1} < U \leq p_k} \rho_k(Z_{k,i}, \hat{\theta}_k), \\ \tilde{Y}_i = \displaystyle\sum_{k=1}^{K} \mathbb{I}_{p_{k-1} < U \leq p_k} \rho_k(Z_{k,i}, \theta_{k,*}), \end{array} \right. \tag{19}$$

where $p_k = \sum_{l=0}^{k} \pi_l$ and $\hat{p}_k = \sum_{l=0}^{k} \hat{\pi}_l$, with the convention $\pi_0 = 0$ and $\hat{\pi}_0 = 0$. Then the term T_2 can be treated as follows:

$$T_2 \leq D_2 + R_{2,1} + R_{2,2} + R_{2,3} + R_{2,4}, \tag{20}$$

where

$$D_2 := \left| \frac{1}{n} \sum_{i=1}^{n} \left[\Psi(\tilde{Y}_i) - E_{f_*}(\Psi) \right] \right|,$$

$$R_{2,1} := \frac{1}{n} \sum_{i=1}^{n} \left| \hat{F}(Y_i) - F_*(Y_i) \right|, \quad R_{2,2} := \frac{1}{n} \sum_{i=1}^{n} \left| \bar{F}(Y_i) - F(Y_i) \right|,$$

$$R_{2,3} := \left| \frac{1}{n} \sum_{i=1}^{n} \left[\Psi(Y_i) - \Psi(\check{Y}_i) \right] \right|, \quad R_{2,4} := \left| \frac{1}{n} \sum_{i=1}^{n} \left[\Psi(\check{Y}_i) - \Psi(\tilde{Y}_i) \right] \right|,$$

The three first terms in the right hand side of (20) being similar to the three first terms in (18) we can state that $M_2 := D_2 + R_{2,1} + R_{2,2} = O_P(1/\sqrt{n})$.

Term $R_{2,3}$. We note that

$$R_{2,3} \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{Y_i \neq \check{Y}_i},$$

where, denoting $\Delta\pi := \sum_{k=1}^{K} |\hat{\pi}_k - \pi_k|$,

$$\mathbb{I}_{Y_i \neq \check{Y}_i} = \sum_{k=1}^{K} (\mathbb{I}_{\hat{p}_{k-1} \wedge p_{k-1} < U_i < \hat{p}_{k-1} \vee p_{k-1}} + \mathbb{I}_{\hat{p}_k \wedge p_k < U_i < \hat{p}_k \vee p_k}), \quad \text{and}$$

$$\mathcal{L}\left(\mathbb{I}_{Y_i \neq \check{Y}_i} \,\middle|\, \mathcal{F}_n \right) \sim \mathcal{B}(\Delta\pi).$$

Let us remark that:

$$\Delta\pi \leq \|\hat{\vartheta} - \vartheta_*\|. \tag{21}$$

Term $R_{2,4}$. Using the mean value Theorem and the fact that Ψ' is uniformly bounded on $\mathbb{R}$, we obtain

$$R_{2,4} := \frac{1}{n}\sum_{i=1}^{n} \|\Psi'\|_\infty |\check{Y}_i - \tilde{Y}_i|,$$

where

$$|\check{Y}_i - \tilde{Y}_i| \leq \sum_{k=1}^{K} |\rho_k(Z_{k,i}, \hat{\theta}_k) - \rho_k(Z_{k,i}, \theta_{k,*})| \leq C \sum_{k=1}^{K} [|Z_{k,i}| + 1] \times \|\hat{\theta}_k - \theta_{k,*}\|.$$

Let us denote $W_i := \sum_{k=1}^{K} (E(|Z_{k,i}| + 1)$, $m := E(W_1)$ and $V := \mathrm{Var}(W_1)$. To conclude, we prove that there exists a constant $\gamma > 0$ such that for all $\varepsilon > 0$ there exists an integer N_ε such that $P(\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma) \leq \varepsilon$, for all $n \geq N_\varepsilon$. Since $\|\hat{\vartheta} - \vartheta_*\| = O_P(1/\sqrt{n})$ there exists $\kappa > 0$ such that for all $\delta > 0$ there exists N_δ ensuring $P(A_n^c) \leq \delta$ for all $n \geq N_\delta$ where $A_{n,\kappa} := \{\sqrt{n}\|\hat{\vartheta} - \vartheta_*\| < \kappa\}$. Let us consider $\gamma > 0$ large enough such that:

$$\frac{\kappa}{(\gamma - \kappa)^2} \leq \frac{\varepsilon}{4J}, \quad \frac{V}{\left(\frac{\gamma}{C\kappa} - m\right)^2} \leq \frac{\varepsilon}{4J}, \quad \text{and} \quad \gamma > \max(1, M)\kappa. \tag{22}$$

Then,

$$\begin{aligned} P(\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma) &= P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \cap A_{n,\kappa}\right) \\ &\quad + P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \cap A_{n,\kappa}^c\right) \\ &\leq P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \,\middle|\, A_{n,\kappa}\right) + P\left(A_{n,\kappa}^c\right). \end{aligned} \tag{23}$$

Since $D_1 = O_P(1/\sqrt{n})$ and $R_{1,2} = O_P(1/\sqrt{n})$, we can choose ξ such that $2\xi/(1 - \delta) = \varepsilon/4(J + 1)$ and there exists a non-negative integer N_ξ, such that for all $n \geq N_\xi$, $P(\sqrt{n}D_1 \geq \gamma) \leq \xi$ and $P(\sqrt{n}R_{1,2} \geq \gamma) \leq \xi$ (we suppose here that γ is large enough to satisfy these two conditions).

We now establish an upper bound for the first term in the right hand side of (23) by

$$\begin{aligned} &P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \,\middle|\, A_{n,\kappa}\right) \\ &\leq P(\sqrt{n}M_1 \geq \gamma | A_{n,\kappa}) \\ &+ \sum_{l=1}^{J} \big(P(\sqrt{n}M_2(l) \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{2,3}(l) \geq \gamma | A_{n,\kappa}) \\ &+ P(\sqrt{n}R_{2,4}(l) \geq \gamma | A_{n,\kappa})\big). \end{aligned} \tag{24}$$

Using similar reasoning for establishing bounds on the terms involving M_2 (as used for terms involving M_1 in (24)), we note, according to the definition of $R_{1,1}$ and the last condition in (22), that

$$\begin{aligned}
&P(\sqrt{n}M_1 \geq \gamma | A_{n,\kappa}) \\
&\leq P(\sqrt{n}D_1 \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{1,1} \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{1,2} \geq \gamma | A_{n,\kappa}) \\
&\leq \frac{P(\sqrt{n}D_1 \geq \gamma)}{P(A_{n,\kappa})} + P\left(A^c_{n,\gamma \max(M,1)^{-1}} | A_{n,\kappa}\right) + \frac{P(\sqrt{n}R_{1,2} \geq \gamma)}{P(A_{n,\kappa})} \\
&\leq \frac{2\xi}{1-\delta} \leq \frac{\varepsilon}{4(J+1)}.
\end{aligned} \tag{25}$$

The terms involving $R_{2,3}$ in (24) are handled by applying Tchebychev's inequality:

$$\begin{aligned}
&P(\sqrt{n}R_{2,3} \geq \gamma | A_{n,\kappa}) \\
&\leq P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{Y_i \neq \check{Y}_i} \geq \gamma | A_{n,\kappa}\right) \\
&= P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{I}_{Y_i \neq \check{Y}_i} - E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | \mathcal{F}_n)\right) \geq \gamma - \sqrt{n}E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | \mathcal{F}_n) \,\middle|\, A_{n,\kappa}\right) \\
&\leq \frac{\mathrm{Var}\left(\sum_{i=1}^{n} \mathbb{I}_{Y_i \neq \check{Y}_i} \mid A_{n,\kappa}\right)}{n(\gamma - \sqrt{n}E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | A_{n,\kappa}))^2} \\
&\leq \frac{\kappa}{(\gamma - \kappa)^2} \leq \frac{\varepsilon}{4J},
\end{aligned} \tag{26}$$

since $E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | A_{n,\kappa}) \leq \kappa/\sqrt{n}$ due to remark (21), and first condition in (22) is supposed. The $R_{2,4}$-dependent terms in (24) are handled in a similar manner by using again Tchebychev's inequality:

$$\begin{aligned}
&P(\sqrt{n}R_{2,4} \geq \gamma | A_{n,\kappa}) \\
&\leq P\left(\frac{C' \left\|\hat{\vartheta}_k - \vartheta_{k,*}\right\|}{\sqrt{n}} \sum_{i=1}^{n} W_i \geq \gamma \,\middle|\, A_{n,\kappa}\right) \\
&\leq P\left(\sum_{i=1}^{n} (W_i - m) \geq \gamma \frac{\sqrt{n}}{C'|\hat{\vartheta}_k - \vartheta_{k,*}|} - m \,\middle|\, A_{n,\kappa}\right) \\
&\leq P\left(\sum_{i=1}^{n} (W_i - m) \geq \frac{\gamma n}{C\kappa} - m\right) \\
&\leq \frac{\mathrm{Var}\left(\sum_{i=1}^{n} W_i\right)}{n\left(\frac{\gamma}{C\kappa} - \frac{m}{n}\right)^2} \\
&\leq \frac{V}{\left(\frac{\gamma}{C\kappa} - m\right)^2} \leq \frac{\varepsilon}{4J},
\end{aligned} \tag{27}$$

according to the second condition in (22). The proof of i) is completed by collecting results (22–27) and taking $\delta = \varepsilon/4$.

ii) First we have:

$$\begin{aligned} P(\hat{\jmath} \neq j_*) &= P\left(\cup_{1\leq j\neq j_*\leq J}\left\{\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right\}\right) \\ &\leq \sum_{1\leq j\neq j_*\leq J} P\left(\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right). \end{aligned} \tag{28}$$

Next, for all $1 \leq j \neq j_* \leq J$, since $\Delta_{j_*,j} := ICE(f_j, f) - ICE(f_{j_*}, f) > 0$ (possibly "arbitrarily" small), we suggest to write

$$\begin{aligned} &\left\{\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right\} \\ &= \left\{\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f) + ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f) > \Delta_{j_*,j}\right\} \\ &\subseteq \left\{|\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f)| + |ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f)| > \Delta_{j_*,j}\right\} \\ &\subseteq \left\{|\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f)| > \Delta_{j_*,j}\right\} \\ &\cup \left\{|ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f)| > \Delta_{j_*,j}\right\} \end{aligned}$$

Finally noticing that, according to i) in Theorem 1, for all $j \in \mathcal{J}$, there exists $K > 0$ such that for all $\delta := \varepsilon/2(J-1) > 0$, there exists an integer N_δ such that for all $n \geq N_\delta$: $P(|\widehat{ICE}(\hat{f}_j, f) - ICE(f_j, f)| \geq K/\sqrt{n}) \leq \delta$, we can define

$$n_j := \min\left\{n \in \mathbb{N} :\ \Delta_{j_*,j} \geq \frac{K}{\sqrt{n}}\right\},$$

which provides us, for all $\varepsilon > 0$, the existence of an integer $N_\varepsilon := \max(N_\delta, n_1, \ldots, n_J)$ such that, according to (28), for all $n \geq N_\varepsilon$:

$$\begin{aligned} P(\hat{\jmath} \neq j_*) &\leq \sum_{1\leq j\neq j_*\leq J}\ \sum_{k=j_*,j} P\left(|\widehat{ICE}(\hat{f}_k, f) - ICE(f_k, f)| > \frac{K}{\sqrt{n}}\right) \\ &\leq \sum_{1\leq j\neq j_*\leq J}\ \sum_{k=j_*,j} \varepsilon/2(J-1) = \varepsilon. \end{aligned}$$

which concludes the proof.

iii) The proof is entirely similar to the proof of ii) when replacing j_* by j_0 and noticing that $ICE(f_{j_0}, f) = 0$ □

Appendix B: Supplementary materials

B.1. Behavior of the MLE when the model is possibly misspecified

In this section, we briefly recall some basic material, from White [40], regarding the asymptotic behavior of the MLE when the model is possibly misspecified. The first assumption defines the structure which generates our observations.

Assumption (A1). The iid sample $X_1^n = (X_1, \ldots, X_n)$, $n \geq 1$, is distributed according to a cdf F_0 on $\mathbb{R}$ whose density, with respect to the Lebesgue measure, is denoted f_0. Since F_0 is not known apriori, we choose a family of cdfs which may or may not contain the true structure of F_0. It is usually easy to choose this family to satisfy the next assumption.

Assumption (A2). The family of cdfs $F(\cdot, \vartheta)$ admits a density $f(\cdot, \vartheta)$ (which will sometimes be denoted for convenience $f_\vartheta(\cdot)$) with respect to the Lebesgue measure on $\mathbb{R}$, which is measurable in x for all ϑ in Θ a compact subset of $\mathbb{R}^p$, and continuous in ϑ for all $x \in \mathbb{R}$.

Next, we define the quasi-log-likelihood of the sample as

$$L_n(X_1^n, \vartheta) := \frac{1}{n} \sum_{i=1}^{n} \log f(X_i, \vartheta), \tag{29}$$

and we define a quasi-maximum likelihood estimator (QMLE) as the parameter $\hat{\vartheta}_n$ which solves the maximization problem

$$\hat{\vartheta}_n = \arg\max_{\vartheta \in \Theta} L_n(X_1^n, \vartheta). \tag{30}$$

In Theorem 2.1, White [40] establishes, under Assumptions A1 and A2, the *existence*, for all $n \geq 1$, of a measurable QMLE $\hat{\vartheta}_n$. Given the existence of a QMLE, let us precisely define its properties. It is well known that when $\{F(\cdot, \vartheta),\ \vartheta \in \Theta\}$ contains the true structure ($F(\cdot) := F(\cdot, \vartheta_0)$ for some ϑ_0 in the interior of Θ), the MLE is consistent for ϑ_0 under suitable conditions, see *e.g.* Theorem 2 in Wald [39], Theorem 5.a in LeCam [33]. Without this restriction, Akaike [23] has noted that since $L_n(X_1^n, \vartheta)$ is a natural estimator for $E(\log(f(X_1, \vartheta))$, $\hat{\vartheta}_n$ is a natural estimator of ϑ_* that minimizes the Kullback Leibler [32] divergence ($\mathcal{K}$), *i.e.*

$$\vartheta_* := \arg\min_{\vartheta \in \Theta} \mathcal{K}(f, f_\vartheta), \quad \text{where} \quad \mathcal{K}(f, f_\vartheta) := E\left(\log\left[\frac{f(X_1)}{f(X_1, \vartheta)}\right]\right). \tag{31}$$

To support Akaike's observation that $\hat{\vartheta}_n$ is a natural estimator for ϑ_*, White [40] imposes the following additional condition.

Assumption (A3).
i) $E(\log(f_0(X_1))$ exists.
ii) $|\log(f(x, \vartheta))| \leq m(x)$ for all $\vartheta \in \Theta$, where m is integrable with respect to f_0.
iii) $K(f, f_\vartheta)$ has a unique minimum at point ϑ_* in Θ.

When assumption (A3) ii) holds, ϑ_* is globally identifiable. In Theorem 2.2, White [40] establishes, under assumptions A1–A3, the strong ϑ_*-consistency of the QMLE defined in (30), *i.e.*

$$\hat{\vartheta}_n \xrightarrow{a.s.} \vartheta_*, \quad \text{as} \quad n \to \infty. \tag{32}$$

With additionnal conditions (given below), White [40] also shows that the QMLE is asymptotically normally distributed. When the partial derivatives exist, we define the matrices

$$\begin{aligned} A_n(\vartheta) &:= \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log(f(X_i,\vartheta))}{\partial\vartheta_k \partial\vartheta_l} \right\}_{k,l=1,\ldots,p}, \\ B_n(\vartheta) &:= \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log(f(X_i,\vartheta))}{\partial\vartheta_k} \times \frac{\partial \log(f(X_i,\vartheta))}{\partial\vartheta_l} \right\}_{k,l=1,\ldots,p}. \end{aligned}$$

If expectation also exists, we define the matrices

$$\begin{aligned} A(\vartheta) &:= \left\{ E\left(\frac{\partial^2 \log(f(X_1,\vartheta))}{\partial\vartheta_k \partial\vartheta_l} \right) \right\}_{k,l=1,\ldots,p}, \\ B(\vartheta) &:= \left\{ E\left(\frac{\partial \log(f(X_1,\vartheta))}{\partial\vartheta_k} \times \frac{\partial \log(f(X_i,\vartheta))}{\partial\vartheta_l} \right) \right\}_{k,l=1,\ldots,p}. \end{aligned}$$

Finally, when the appropriate inverse exists, define

$$\begin{aligned} C_n(\vartheta) &:= A_n(\vartheta)^{-1} B_n(\vartheta) A_n(\vartheta)^{-1}, \\ C(\vartheta) &:= A(\vartheta)^{-1} B(\vartheta) A(\vartheta)^{-1}. \end{aligned}$$

Assumption (A4). The collection $\{\partial \log((f(x,\vartheta))/\partial\vartheta_k,\ k = 1,\ldots,p\}$ are measurable functions of x for each $\vartheta \in \Theta$ and continuously differentiable functions of ϑ for each x in $\mathbb{R}$.

Assumption (A5). The two collections $\{|\partial^2 \log((f(x,\vartheta))/\partial\vartheta_k \partial\vartheta_l|,\ k,l = 1, \ldots,p\}$ and $\{|\partial \log((f(x,\vartheta))/\partial\vartheta_k \times \partial \log((f(x,\vartheta))/\partial\vartheta_l|,\ k,l = 1,\ldots,p\}$, are dominated by functions integrable with respect to f_0 for all $x \in \mathbb{R}$ and $\vartheta \in \Theta$.

Assumption (A6).
i) The parameter ϑ_* is an interior point of Θ.
ii) The $p \times p$ matrix $B(\vartheta_*)$ is nonsingular.
iii) The parameter ϑ_* is a regular point of $A(\vartheta)$.

Under assumptions A1–6, White ([40], Theorem 3.2) establishes the asymptotic normality of the QMLE, *i.e.*

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_*) \xrightarrow{d} \mathcal{N}(0, C(\vartheta_*)), \quad \text{as} \quad n \to +\infty. \tag{33}$$

Remark. It is important to recall that if we suppose $g(\cdot) = f(\cdot,\vartheta_0)$ for $\vartheta_0 \in \Theta$, then the QMLE $\hat{\vartheta}_n$ is simply called MLE and if assumptions A1–A6 hold, the MLE is consistent and asymptotically normally distributed according to (30) and (33) when replacing ϑ_* by ϑ_0.

B.2. Mixtures of Gaussian and Gumbel distributions

Identifiability In this paragraph, we propose to establish the identifiability of finite univariate mixtures of Gaussian and Gumbel distributions. Let us first establish a more general result on two-group univariate mixtures. Consider $\mathcal{L}$ and $\mathcal{G}$ two families of distribution functions defined by:

$$\mathcal{L} := \{L(x;\theta) : \ x \in \mathbb{R}, \theta \in \Theta\}, \quad \mathcal{G} := \{G(x;\phi) : \ x \in \mathbb{R}, \phi \in \Psi\} \tag{34}$$

where Θ and Ψ denote parametric spaces. We consider the set $\mathcal{H}$ of all finite mixtures sourcing their distributions in groups $\mathcal{L}$ and $\mathcal{G}$ defined by:

$$\mathcal{H} := \Bigg\{ H(x) = \sum_{i=1}^{n_1} c_i L_i(x) + \sum_{j=n_1+1}^{n_1+n_2} c_j G_j(x), \tag{35}$$

$$\sum_{i=1}^{n_1+n_2} c_i = 1, \ c_i > 0, (n_1+n_2) \in \mathbb{N}^* \times \mathbb{N}^* \Bigg\}. \tag{36}$$

The class of mixture models H is said identifiable if and only if H has the unique representation property:

$$\sum_{i=1}^{n_1} c_i L_i(x) + \sum_{j=n_1+1}^{n_1+n_2} c_j G_j(x) = \sum_{k=1}^{n_1'} c_k' L_k'(x) + \sum_{l=n_1'+1}^{n_1'+n_2'} c_l' G_l'(x)$$

which implies $n_1 = n_1'$, $n_2 = n_2'$, and for each i, $1 \leq i \leq n_1$ and respectively, each j, $1 \leq j \leq n_2$, there is some $1 \leq k \leq n_1$ and respectively, some $1 \leq l \leq n_2$, such that $F_i = F_k'$ and $G_j = G_l'$.

Theorem 3. *Let $\mathcal{F}$ and $\mathcal{G}$ two families of cdfs with respective transforms $\alpha(t)$ and $\gamma(t)$ defined for t respectively in D_α and D_β (the domains of definition of α and γ) such that the mappings $L \to \alpha$ and $G \to \gamma$ are linear and one-to-one. We denote by I_α and I_γ, the largest interval contained respectively in D_α and D_β. Let us denote for all $F \in \mathcal{F} \cup \mathcal{G}$ by ρ_F, its associated transform. Suppose that there exists a total ordering of $\mathcal{F} \cup \mathcal{G}$, denoted by $\preceq$ and satisfying $G \prec F$ if $(F,G) \in \mathcal{F} \times \mathcal{G}$, such that for all $(F_1,F_2) \in (\mathcal{F} \cup \mathcal{G})^2$ the condition $F_1 \prec F_2$ implies (i) $I_{\rho_{F_1}} \subseteq I_{\rho_{F_2}}$ (ii) the existence of a certain t_1 in the closure of $D_{\rho_{F_1}}$ (t_1 being independent of ρ_{F_2}) such that*

$$\lim_{t \to t_1} \frac{\rho_{F_1}(t)}{\rho_{F_2}(t)} = 0,$$

then the class of finite mixture $\mathcal{H}$ is identifiable.

Proof. The proof of this result is entirely similar to the proof of Theorem 2 in Teicher [38]. □

Corollary 4. *Let $\mathcal{F}$ be the family of Gaussian cdfs and $\mathcal{G}$, the family of Gumbel cdfs, then the class of finite mixtures sourcing their distributions in groups $\mathcal{F}$ and $\mathcal{G}$ is identifiable.*

Proof. Let us consider $\alpha(\cdot)$ and $\gamma(\cdot)$ the respective moment generating functions of $\mathcal{F}$ and $\mathcal{G}$, where $D_\alpha = \mathbb{R}$ and $D_\gamma \subset \mathbb{R}$. We order each family lexicographically by: $F_1 \sim \mathcal{N}(m_1, \sigma_1^2) \prec F_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ if $\sigma_1 > \sigma_2$ or if $\sigma_1 = \sigma_2$ but $m_1 < m_2$, and $G_1 \sim \mathcal{G}(\mu_1, \beta_1) \prec F_2 \sim \mathcal{G}(\mu_2, \beta_2)$ if $\beta_1 > \beta_2$ or if $\beta_1 = \beta_2$ but $\mu_1 < \mu_2$. We cross order families $\mathcal{F}$ and $\mathcal{G}$ by: $G \in \mathcal{G} \prec F \in \mathcal{F}$ since $D_{\psi_G} \subset D_{\alpha_F}$.

Case $(F_1, F_2) \in \mathcal{G} \times \mathcal{G}$. The moment generating function of a Gumbel distribution $\mathcal{G}(\mu, \beta)$ is given by

$$\gamma(t) = e^{\mu t}\Gamma(1 - \beta t), \qquad t < 1/\beta,$$

and, for $i = 1, 2$, the largest interval contained in $D_{\rho_{F_i}}$ is $I_{\rho_{F_i}} = (-\infty, 1/\beta_i)$, which implies $I_{\rho_{F_1}} \subseteq I_{\rho_{F_2}}$ if $F_1 \prec F_2$. If $\beta_2 < \beta_1$ then there exist $t_1 = 1/\beta_1$ such that

$$\lim_{t \to t_1} \frac{\rho_{F_2}(t)}{\rho_{F_1}(t)} = \lim_{t \to \frac{1}{\beta_1}} \frac{e^{\mu_2 t}\Gamma(1 - \beta_2 t)}{e^{\mu_1 t}\Gamma(1 - \beta_1 t)} = 0.$$

If $\beta_2 = \beta_1$ and $\mu_1 < \mu_2$ there exists $t_1 = -\infty$ such that

$$\lim_{t \to t_1} \frac{\rho_{F_2}(t)}{\rho_{F_1}(t)} = \lim_{t \to -\infty} \frac{e^{\mu_2 t}}{e^{\mu_1 t}} = 0.$$

Case $(F_1, F_2) \in \mathcal{F} \times \mathcal{F}$. The moment generating function of a Normal distribution $\mathcal{N}(m, \sigma^2)$ is given by

$$\alpha(t) = e^{mt + \frac{1}{2}\sigma^2 t^2}, \qquad t \in \mathbb{R},$$

and for $i = 1, 2$, $D_{\rho_{F_i}} = I_{\rho_{F_i}} = (-\infty, +\infty)$. In that case, application of Theorem 3 is straightforward by taking $t_1 = +\infty$ (the calculations are similar to Teicher [38] who considers the Laplace transform instead of the moment generating function).

Case $(F_1, F_2) \in \mathcal{G} \times \mathcal{F}$. Since our total ordering is completed by $D_{\rho_{F_1}} \subset D_{\rho_{F_2}} \Rightarrow F_1 \prec F_2$, we have for all $F_1 \sim \mathcal{G}(\mu_1, \beta_1) \in \mathcal{G}$ and all $F_2 \sim \mathcal{N}(m_2, \sigma_2^2)$, $I_{\rho_{F_1}} = (-\infty, 1/\beta_1) \subseteq I_{\rho_{F_2}} = (-\infty, +\infty)$, and there exists $t_1 = 1/\beta_1$ such that

$$\lim_{t \to t_1} \frac{\rho_{F_2}(t)}{\rho_{F_1}(t)} = \lim_{t \to \frac{1}{\beta_1}} \frac{e^{mt + \frac{1}{2}\sigma^2 t^2}}{e^{\mu_1 t}\Gamma(1 - \beta_1 t)} = 0,$$

which concludes the proof. □

B.3. Assumptions checking

Assumption (G). For the Normal distribution and the Gumbel distribution, it is enough to simulate random variables according to $\mathcal{N}(0, 1)$ and $\mathcal{G}(0, 1)$ distributions respectively, and consider the transformation $\rho(y, m, \sigma) := (y - m)/\sigma$

and $\rho(y,\mu,\beta) := (y-\mu)/\beta$. Moreover the condition (13) is clearly satisfied in the Gaussian and Gumbel case since generically for all $(m,m') \in [\underline{m},\overline{m}]^2$ and $(\sigma,\sigma') \in [\underline{\sigma},\overline{\sigma}]^2$ we have

$$\begin{aligned}
|\rho(x,m,\sigma) - \rho(y,m',\sigma')| &= \left|\frac{x-m}{\sigma} - \frac{x-m'}{\sigma'}\right| \\
&\leq (|x|+m)\left|\frac{\sigma'-\sigma}{\sigma\sigma'}\right| + \left|\frac{m-m'}{\sigma'}\right| \\
&\leq (|x|+\overline{m})\left|\frac{\sigma'-\sigma}{\underline{\sigma}^2}\right| + \left|\frac{m-m'}{\underline{\sigma}}\right| \\
&\leq \frac{\max(1,\overline{m})}{\min(\underline{\sigma},\underline{\sigma}^2)}(|x|+1)(|\sigma-\sigma'|+|m-m'|) \\
&= C(|x|+1)\|\theta-\theta'\|.
\end{aligned}$$

for $C := \max(1,\overline{m})/\min(\underline{\sigma},\underline{\sigma}^2)$.

Assumption (R). For the Gaussian pdf $F_{\mathcal{N}(m,\sigma^2)}(\cdot) := F_{\mathcal{N}}(\cdot,\theta)$ where $\theta = (m,\sigma) \in [\underline{m},\overline{m}] \times [\underline{\sigma},\overline{\sigma}]$ we have, for all $x \in \mathbb{R}$,

$$\begin{aligned}
\left|\frac{\partial}{\partial m}F_{\mathcal{N}}(x,\theta)\right| &= \left|f_{\mathcal{N}(0,1)}\left(\frac{x-m}{\sigma}\right)\frac{1}{\sigma}\right| \leq \frac{1}{\sqrt{\pi}\sigma^2} \leq \frac{1}{\sqrt{\pi}\min(\underline{\sigma}^2,\underline{\sigma}^3)}, \\
\left|\frac{\partial}{\partial \sigma}F_{\mathcal{N}}(x,\theta)\right| &= \left|f_{\mathcal{N}(0,1)}\left(\frac{x-m}{\sigma}\right)\frac{1}{\sigma^2}\right| \leq \frac{1}{\sqrt{\pi}\sigma^3} \leq \frac{1}{\sqrt{\pi}\min(\underline{\sigma}^2,\underline{\sigma}^3)}.
\end{aligned}$$

For the Gumbel pdf $F_{\mathcal{G}(\mu,\beta)}(\cdot) := F_{\mathcal{G}}(\cdot,\theta)$ where $\theta = (\mu,\beta)$ we have, for all $x \in \mathbb{R}$,

$$\begin{aligned}
\left|\frac{\partial}{\partial \mu}F_{\mathcal{G}}(x,\theta)\right| &= \left|f_{\mathcal{G}(0,1)}\left(\frac{x-\mu}{\beta}\right)\frac{1}{\beta}\right| \leq \frac{1}{e\beta^2} \leq \frac{1}{e\min(\underline{\beta}^2,\underline{\beta}^3)}, \\
\left|\frac{\partial}{\partial \beta}F_{\mathcal{G}}(x,\theta)\right| &= \left|f_{\mathcal{G}(0,1)}\left(\frac{x-\mu}{\beta}\right)\frac{1}{\beta^2}\right| \leq \frac{1}{e\beta^3} \leq \frac{1}{e\min(\underline{\beta}^2,\underline{\beta}^3)}.
\end{aligned}$$

Assumption (A3) ii). For any Gaussian pdf $f_{\mathcal{N}(m,\sigma^2)}$ with parameters $(m,\sigma) \in [\underline{m},\overline{m}] \times [\underline{\sigma},\overline{\sigma}]$, we have the following upper-bound:

$$\begin{aligned}
f_{\mathcal{N}(m,\sigma^2)}(x) \leq \frac{1}{\sqrt{2\pi\underline{\sigma}^2}}\Bigg(&\mathbb{I}_{\underline{m}\leq x\leq\overline{m}} + \exp\left(-\frac{1}{2}\left(\frac{x-\underline{m}}{\overline{\sigma}}\right)^2\right)\mathbb{I}_{x\leq\underline{m}} \\
&+ \exp\left(-\frac{1}{2}\left(\frac{x-\overline{m}}{\overline{\sigma}}\right)^2\right)\mathbb{I}_{x\geq\overline{m}}\Bigg) := b_{\mathcal{N}}(x).
\end{aligned}$$

For any Gumbel pdf $f_{\mathcal{G}(\mu,\beta)}$ with parameters $(\mu,\beta) \in [\underline{\mu},\overline{\mu}] \times [\underline{\beta},\overline{\beta}]$, we propose a similar upper-bound whose construction is detailed next. For this purpose, we note that for $u \in (0,+\infty)$, the function $r(u) := \exp(-u)u$ is strictly increasing on $(0,1]$ and strictly decreasing on $(1,+\infty)$. Thus for all $x > 0$, since $\exp(x/\beta) >$

TABLE 4
$\overline{ICE}$ values for different mixtures, various computational cell sizes (125 and 1000 grains) and boundary conditions (FS=Free Surface, PBC=Periodic Boundary Condition)

Model label	Model	125-FS	125-PBC	1000-FS	1000-PBC
1	$NG[0,1]$	0.015154	0.023928	0.014659	0.010650
2	$NG[1,1]$	0.008158	0.007290	0.008982	0.009624
3	$NG[0,2]$	0.010069	0.008906	0.008480	0.012682
4	$NG[1,2]$	0.007748	0.008262	0.006828	0.008576
5	$NG[0,3]$	0.009824	0.007076	0.006731	0.008937
6	$NG[2,2]$	0.007598	0.006328	0.006679	0.007552
7	$NG[1,3]$	0.005013	0.005218	0.006713	0.008418
8	$NG[0,4]$	0.005018	0.007088	0.006239	0.006226

$\exp(x/\overline{\beta}) > 1$ we obtain $r(\exp(x/\overline{\beta})) > r(\exp(x/\beta))$. Next for all $x \leq 0$, since $\exp(x/\beta) < \exp(x/\overline{\beta}) \leq 1$ we also obtain $r(\exp(x/\overline{\beta})) > r(\exp(x/\beta))$. Using this observation, we establish easily that:

$$f_{\mathcal{G}(\mu,\beta)}(x) \leq \frac{1}{\underline{\beta}} \left(\mathbb{I}_{\underline{\mu} \leq x \leq \overline{\mu}} + r\left(\exp\left(\frac{x-\underline{\mu}}{\overline{\beta}} \right) \right) \mathbb{I}_{x \leq \underline{\mu}} + r\left(\exp\left(\frac{x-\overline{\mu}}{\overline{\beta}} \right) \right) \mathbb{I}_{x \geq \overline{\mu}} \right)$$
$$:= b_{\mathcal{G}}(x).$$

In conclusion, we have

$$\log\left(\sum_{i=1}^{n_1} f_{\mathcal{N}(m_i,\sigma_i^2)}(x) + \sum_{i=n_1+1}^{n_2} f_{\mathcal{G}(\mu_i,\beta_i)}(x) \right)$$
$$\leq \log\left(n_1 b_{\mathcal{N}}(x) + n_2 b_{\mathcal{G}}(x) \right)$$
$$\leq \log(n_1+n_2) + \log\left(b_{\mathcal{N}}(x) \right) + \log\left(b_{\mathcal{G}}(x) \right) := m(x),$$

which implies that f_0 must have to integrate $\exp(x/\bar{\beta})$ over $\mathbb{R}$. Note that this condition always holds if f_0 is a mixture of Normal and Gumbel distributions.

Assumption (A3) iii). The identifiability property established in Section B.2 is a necessary condition, but cannot insure that **(A3)** iii) is automatically satisfied.

Assumption (A4–5). Checking assumption A4 is straightforward. We can prove, similarly to the result established for **(A3)**, that **(A5)** is satisfied if f_0 admits exponential moments.

The remaining standard assumptions involving f_0, *i.e.* A1, A3 i) and iii), A6, are generally imposed.

References

[1] HENNA, J. (1985). On estimating the number of constituents of a finite mixture of continuous distributions. *Ann. Inst. Statist. Math.*, **37**, 235–240. MR0799237

[2] IZENMAN, A. J. and SOMMER, C. (1988). Philatelic mixtures and multivariate densities. *Journal of the American Math. Soc.*, **83**, 941–953.

[3] Roeder, K. (1994). A graphical technique to determining the number of components in a mixture of normals. *J. American Statist. Assoc.*, **89**, 487–495. MR1294074
[4] Lindsay, B. G. (1983). Moment matrices: Application in mixtures. *Ann. Statist.*, **17**, 722–740. MR0994263
[5] Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *Ann. Statist.*, **27**, 1178–1209. MR1740115
[6] Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhya Series A*, **62**, 49–66. MR1769735
[7] Berkhof, J., van Mechelen, I. and Gelman, A. (2003). A Bayesian approach to the selection and testing of mixture models. *Statistica Sinica*, **13**, 423–442. MR1977735
[8] Vuong, Q. H. (1989). Likelihood ratio test for model selection and non-nested hypothesis. *Econometrica*, **57**, 307–333. MR0996939
[9] Suresh, S. (1998). *Fatigue of Materials, 2nd ed.*, Cambridge University Press, Cambridge, UK.
[10] McDowell, D. L. (1996). Basic issues in the mechanics of high cycle metal fatigue. *Int. J. Frac.*, **80**, 103–145.
[11] Schijve, J. (2005). Statistical distribution functions and fatigue of structures. *Int. J. Fat.*, **27**, 1031–1039.
[12] Przybyla, C. P. and McDowell, D. L. (2010). Microstructure-sensitive extreme value probabilities for high cycle fatigue of Ni-base superalloy IN100. *Int. J. Plast.*, **26**, 372–394.
[13] Berger, C. and Kaiser, B. (2006). Results of very high cycle fatigue tests on helical compression springs. *Int. J. Fat.*, **28**, 1658–1663.
[14] Marines, I., Bin, X. and Bathias, C. (2003). An understanding of very high cycle fatigue of metals. *Int. J. Fat.*, **25**, 1101–1107.
[15] Miao, J., Pollock, T. M. and Jones, J. W. (2009). Crystallographic fatigue crack initiation in nickel-based superalloy René 88DT at elevated temperature. *Acta Mat.*, **57**, 5964–5974.
[16] Jha, S. K., Caton, M. J. and Larsen, J. M. (2008). Mean vs. life-limiting fatigue behavior of a nickel-based superalloy. *Superalloys 2008 – Proceedings of the 11th International Symposium on Superalloys*, 565–572.
[17] Sakai, T., Lian, B., Takeda, M., Shiozawa, K., Oguma, N., Ochi, Y., Nakajima, M. and Nakamura, T. (2010). Statistical duplex SN characteristics of high carbon chromium bearing steel in rotating bending in very high cycle regime. *Int. J. Fat.*, **32**, 497–504.
[18] Schijve, J. (1994). Fatigue predictions and scatter. *Fatigue Fract. Enng. Mater. Struct.*, **17**, 381–396.
[19] Ravi Chandran, K. S., Chang, P. and Cashman, G. T. (2010). Competing failure modes and complex SN curves in fatigue of structural materials. *Int. J. Fat.*, **32**, 482–491.
[20] Wu, C. F. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103. MR0684867

[21] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Stat. Soc. B*, **39**, 1–38. MR0501537
[22] Ahmad, K. E., Jahseen, Z. F. and Modhesh, A. A. (2010). Estimation of a discriminant function based on small sample size from a mixture of two Gumbel distributions. *Comm. Statist.–Simulation and Computation*, **39**, 713–725. MR2785661
[23] Akaike, H. (1973). Information theory and an extension of the likelihood principle. *Proceedings of the second International symposium of Information Theory*. Ed. Petrov, B. N. and Csáki, F., Akadémiai Kiado, Budapest. MR0483125
[24] Babu, G. J. (2011). Resampling method for model fitting and model selection. *J. Biopharma. Statist.*, **21**, 1177–1186. MR2861900
[25] McDowell, D. L. (2007). Simulation-based strategies for microstructure-sensitive fatigue modeling. *Mat. Sci. Engg. A*, **468–470**, 4–14.
[26] Vandermeulen, W., Scibetta, M., Leenaers, A., Schuurmans, J. and Gérard, R. (2008). Measurement of the Young modulus anisotropy of a reactor pressure vessel cladding. *J. Nuc. Mat.*, **372, 2–3**, 249–255.
[27] Hughes, T. J. R. (2000). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover publications. MR1008473
[28] ABAQUS FEA, V6.7.1. *D S Simulia, Dassault Systèmes, Providence, RI.*
[29] Mesarovic, S. Dj. and Padbidri, J. (2005). Minimal kinematic boundary conditions for simulations of disordered microstructures. *Phil. Mag.*, **85**, 65–78.
[30] Dabrowski, A. R. (1990). Extremal point processes and intermediate quantile functions. *Probab. Theory Related Fields*, **85**, 365–386. MR1055761
[31] Han, L. and Ferreira, A. (2006). *Extreme Value Theory*. Springer, New-York. MR2234156
[32] Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Ann. Math. Statist.*, **22**, 79–86. MR0039968
[33] LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *University of California Publications in Statistics*, **1**, 277–330. MR0054913
[34] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR0838963
[35] Budka, M., Gabrys, B. and Musial, K. (2011). *On Accuracy of PDF Divergence Estimators and Their Applicability to Representative Data Sampling. Entropy*, **13**, 1229–1266. MR2820970
[36] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester. MR0838090
[37] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New-York. MR1385671
[38] Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Stat.*, 34, 1265–1269. MR0155376

[39] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **60**, 595–603. MR0032169

[40] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25. MR0640163