# Bootstrap non-nested mixture model selection - Application to extreme value modeling in metal fatigue problems

Pierre Vandekerkhove, Jagan Padbiri, David Mcdowell

# Bootstrap non-nested mixture model selection - Application to extreme value modeling in metal fatigue problems

Pierre Vandekerkhove*§, Jagan M. Padbidri†and David L. McDowell †‡

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

*§Université Paris-Est, LAMA (UMR 8050), UPEMLV, F-77454, Marne-la-Vallée, France.*

*†George W. Woodruff School of Mechanical Engineering,*
Georgia Institute of Technology, Atlanta, GA 30332-0405, USA.

*‡Institute for materials, IPST at Georgia Institute of Technology,*
Atlanta, GA 30332-0620, USA.

March 28, 2013

**Abstract**

In this paper, we consider the problem of selecting the most appropriate model from many possible models to describe datasets involving mixtures of distributions. The proposed method consists of finding the maximum likelihood estimators (MLEs) of different assumed mixture models that describe a dataset, using the Expectation-Maximization (EM) algorithm, and subsequently using bootstrap sampling technique to identify the distance between the empirical cumulative distribution function (cdf) of the dataset and the MLE fitted cdf. To test 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. The ICE metric shows a markedly improved performance, from the existing metrics, in identifying the correct

mixture model. The method is applied to model the distribution of indicators of the fatigue crack formation potency, obtained from numerical experiments.

# 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 probability density function. 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 combinatory 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 information criterion. 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 criterion being a Statistical divergence and 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 found at most one time among the competing models. The basic idea is to build a true distance for the models in competition from the dataset, based, for each model, on the comparison between the QMLE-fitted cdf and the empirical cdf from the dataset of interest. In this paper,

the motivation for developing a method to identify mixtures of distributions arises from observations of fatigue life distributions of metals.

Fatigue damage is defined as the degradation of material properties due to the repeated application of stresses and strains leading to material failure [9]. Metallic materials and alloys are typically composed of structural units called grains, whose size ranges from a few to a few hundred microns (1 micron = $10^{-6}$ metres). The crystalline structure i.e. atomic arrangement, within each grain can be assumed to be uniform. However, the crystalline structure leads to non-isotropic material response (stiffness) for imposed deformation along different directions, termed as anisotropy. Further, the directions describing the atomic arrangement (orientation) are different for different grains in the metal. The irreversible motion of defects along specific crystallographic directions, determined by the crystalline structure and orientation and the resultant accumulation of damage in the material is one of the primary mechanisms of fatigue crack formation in metals in the high cycle fatigue regime (fatigue life of tens to hundreds of millions of loading cycles). The accumulation of damage with fatigue loading in a material is due to a combination of microstructural features (grain size, orientation, inter-granular interactions) and applied loading. The probability of damage accumulation and subsequent crack formation in a given volume of material is governed by the extreme value probability of such a favorable combination existing in the given material volume [12]. Thus, the fatigue life of a material manifests as a distribution rather than a unique value for multiple experimental realizations of identical applied loading conditions.

The variation in distribution of fatigue life in metals is observed to a greater extent in the regimes described as high cycle and very high cycle fatigue life than for low cycle fatigue life (thousands to tens of thousands of fatigue cycles) [10, 11]. This has been experimentally observed 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]. Due to the probability of crack formation being governed by the extreme values of the driving forces in a material volume, extreme value distributions can be used to chararacterize the distributions of both the observed driving forces and the resultant fatigue life. Przybyla and McDowell [12] used a Gumbel distribution to quantify the variation of fatigue indicator parameters obtained from numerical experiments. Other extreme value distributions, 2 or 3 parameter Weibull distributions, have been used by various researchers 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 the scatter associated with each crack formation mechanism described by an extreme value distribution. It is to be noted that an approach to describing the fatigue life observations using a mixture of distributions for a given crack formation mechanism has not been considered in any of the above studies i.e. the above works assume that the fatigue life distributions for a given crack initiation mechanism can be accurately described by a single extreme value distribution function.

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 the specific crystallographic directions) in a given material volume, sampled from multiple instantiations of numerical experiments. It is to be noted that the mechanism mentioned above, is one of several for crack formation in a material in the high cycle fatigue regime. For a more detailed discussion on crack initiation mechanisms in materials, the reader is referred to Suresh [9]. For the mechanism of crack formation considered here: accumulation of damage along crystallographic planes in the high cycle fatigue regime, the majority of the fatigue life of the material is spent in forming a crack of the size of a grain. Thus, the distribution of extreme values of stresses that drive the motion of defects in these crystallographic directions resulting from applied loading within a given material volume serve as indicators of the distribution of the material fatigue life. For a material which is subjected to fatigue loading, since the extreme values of the stresses in a given material volume are influenced by the grains neighboring the grain in which the extremal values occur, the possibility of a corrupted/mixture of extreme value distributions to describe the distribution of fatigue life, for the same applied mechanical loading, arises. 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 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 a mixture of Gumbel and Gaussian distributions is suspected. Appendix A1 is dedicated to a brief description of the QMLE and its asymptotic conver-

gence properties while we show in Appendix A2 that technical assumptions insuring the validity of our method are fully satisfied when considering mixtures of Gumbel and Gaussian distributions as applied to datasets obtained from numerical simulations in Section 5.

## 2 Problem and methodology

Let us suppose that we observe an univariate i.i.d. sample $X = (X_1, \ldots, X_n)$ distributed according to an unknown probability distribution function (pdf) $f_0$ which is possibly a mixture of pdfs belonging to the collection

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

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

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

is supposed to belong to a parametric space $\Theta_j := S(K_j) \prod_{k=1}^{K_j} \Phi_{j,k}$, where $S(K_j) := \left\{ \pi_{j,k} > 0,\ 1 \leq k \leq K_j : \ \sum_{k=1}^{K_j} \pi_{j,k} = 1 \right\}$, $\Phi_{j,k}$ is a parametric space associated specific to each $\theta_{j,k}$, and $\{f_{j,k}(\cdot|\theta_{j,k}),\ k = 1, \ldots, K_j\}$ is a set of given pdfs with respect to the Lebesgue measure, denoted by $\lambda$, on $\mathbb{R}$.
For simplicity we suppose that there are no *nested* models in the collection $\mathcal{M}$, which is stated in the following assumption.

**(NE).** 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 framework can be ensured by assuming that the components of vector forming the weights of the mixture are all uniformly minorized over the collection of models $\mathcal{M}$.

We denote, for all $j \in \mathcal{J}$, by $\hat{\vartheta}_j(X_1^n) := \hat{\vartheta}_j := (\hat{\pi}_{j,1}, \ldots, \hat{\pi}_{j,K-1}; \hat{\theta}_{j,1}, \ldots, \hat{\theta}_{j,K})$ the Quasi-Maximum Likelihood Estimator (QMLE) of $\vartheta_{j,*}$, *i.e.*, the minimizer of the Kullback-Leibler divergence $\mathcal{K}(f_0, f_j(\cdot, \vartheta))$ over $\Theta_j$, respectively defined by (27) and (28) when considering the model with label $j$ in the family $\mathcal{M}$. Note that these estimators are generally 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.

Next, we define the QMLE plug-in mixture estimate of $f_j(\cdot, \vartheta_{j,*})$, also denoted for convenience $f_{j,*}$, by

$$\hat{f}_j(x) := f_j(x; \hat{\vartheta}(X_1^n)) = \sum_{k=1}^{K_j} \hat{\pi}_{j,k} f_{j,k}(x|\hat{\theta}_{j,k}), \quad x \in \mathbb{R}. \tag{2}$$

Our goal is to decide, among the $J$ 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. For this purpose, let us introduce the finite collection of pdf $\mathcal{M}_*$ defined by

$$\mathcal{M}_* := \{f_0(x), f_{j,*}(x), \quad x \in \mathbb{R}, \quad j = 1, \ldots, J\}, \tag{3}$$

and the Integrated Cumulative Error (ICE) quantity on $\mathcal{M}_*$ defined for all $(f_1, f_2) \in \mathcal{M}_*^2$ by

$$ICE(f_1, f_2) := \frac{1}{2} \int_{\mathbb{R}} |F_1(x) - F_2(x)| dF_{\mathcal{M}_*}(x), \quad x \in \mathbb{R}, \tag{4}$$

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.

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

**(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 in choosing the most appropriate model, we suggest selecting among the collection of models (1), the one, with label $j_* \in \mathcal{J}$, that minimizes the $ICE$ distance to $f_0$ in the collection $\mathcal{M}_*$, *i.e.*,

$$j_* := \arg\min_{j\in\mathcal{J}} ICE(f_{j,*}, f_0). \tag{5}$$

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

We introduce, for all $x \in \mathbb{R}$ and $j \in \{1, \ldots, J\}$,

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

where $Y_j := (Y_{j,1}, \ldots, Y_{j,n})$ is an i.i.d. 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

$$\begin{aligned} \widehat{ICE}(f_{j,*}, f_0) &:= \frac{1}{n(J+1)} \sum_{l=0}^{J} \sum_{i=1}^{n} |\hat{F}_j(Y_{i,l}) - \bar{F}_0(Y_{i,l})| \\ &= \frac{1}{n(J+1)} \sum_{l=0}^{J} \sum_{i=1}^{n} \left| \hat{F}_j(Y_{l,(i)}) - \frac{n_{X,Y_l}(i)}{n} \right|, \qquad x \in \mathbb{R}, \end{aligned} \tag{6}$$

where for all $l = 0, \ldots, J$, $Y_{l,(1)} < \cdots < Y_{l,(n)}$, and $n_{X,Y_l}(i) = \#\left\{X_q \le Y_{l,(i)};\ q = 1, \ldots, n\right\}$. Finally we estimate $j_*$ by $\hat{j}$ defined by:

$$\hat{j} := \arg\min_{j\in\mathcal{J}} \widehat{ICE}(f_j, f_0). \tag{7}$$

*Remark.* This resampling idea, the so-called parametric-Bootstrap, see *e.g.* Babu [24], has been extensively studied in goodness of fit test problems based on various test statistics such as the Kolmogorov-Smirnov, the Cramer-von Mises or Anderson-Darling statistics. However, to our knowledge, employing the parametric bootstrap technique to rank the fitting-performance among a finite collection of mixture models by using a true distribution-distance estimator (see Theorem 1), and not a test-statistic has not been attempted before.

## 3 Assumptions and asymptotic results

For simplicity, we endow the space $\mathbb{R}^s$, $s \ge 1$, 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.

**(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

$$|\rho_{j,k}(x,\theta) - \rho_{j,k}(x,\theta')| \leq C[|x|+1] \times \|\theta - \theta'\|, \quad x \in \mathbb{R}. \tag{8}$$

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

**(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

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

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 [33]), for general distribution function $f_0$, we have

$$\sqrt{n}\|\bar{F} - F\|_\infty = O_P(1). \tag{9}$$

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

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

*Proof.* Consider the following decomposition

$$\begin{aligned}
|\hat{F}_j(x) - F_{j,*}(x)| &\leq \sum_{k=1}^{K} \left( \hat{\pi}_k |F_{j,k}(x, \hat{\theta}_k) - F_{j,k}(x, \theta_{k,*})| + F_{j,k}(x, \theta_*) |\hat{\pi}_{j,k} - \pi_{j,k}| \right) \\
&\leq \sum_{k=1}^{K} \left( \sup_{x \in \mathbb{R}, \theta_k \in \Theta_k} \|\dot{F}_k(x, \theta_k)\|_k \times \|\hat{\theta}_k - \theta_{k,*}\| + |\hat{\pi}_{j,k} - \pi_{j,k}| \right) \\
&\leq \max(M, 1) \|\hat{\vartheta} - \vartheta_*\|. \qquad (10)
\end{aligned}$$

■

**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 $ICE(f_1, f_2) = 0 \Leftrightarrow f_1 = f_2$ $\lambda$-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).*

*Proof.* 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 $\mathcal{M}_*$. ■

**Theorem 2** *i) If all the parametric mixture models belonging to the collection $\mathcal{M}$ satisfy conditions A1-6 (given in Appendix 1) and assumptions NE, S1 or S2, G, R hold, then*

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

*ii) Under S1, if conditions A1-6 and assumptions NE, G, R hold, then the ICE criterion defined in (7) is quasi-consistent in Probability, i.e.*

$$P\left(\hat{j} = j_*\right) \to 1, \quad as \quad n \to \infty. \qquad (11)$$

*iii) Under S2, if conditions A1-6 and assumptions NE, G, R hold, then the ICE criterion defined in (7) is consistent in Probability, i.e.*

$$P\left(\hat{j} = j_0\right) \to 1, \quad as \quad n \to \infty. \tag{12}$$

*Proof.* 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 $\sigma$-algebra generated by the random variables $(X_1, \ldots, X_n)$.

We note that

$$T_1 \leq D_1 + R_{1,1} + R_{1,2}, \tag{13}$$

where

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

According to the central limit theorem, we have $D_1 = O_P(1/\sqrt{n})$, and from (9) and subsequently Lemma 1, we obtain $R_{1,1} \leq \max(M, 1)\|\vartheta_n - \vartheta_*\| = O_P(1/\sqrt{n})$ and $R_{1,2} \leq \left\|\bar{F}_0 - F_0\right\|_\infty = O_P(1/\sqrt{n})$. Let us bear 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 i.i.d. according to $f_*$, in the following way:

$$\left\{ \begin{array}{l} Y_i = \sum_{k=1}^K \mathbb{I}_{\hat{p}_{k-1} < U \leq \hat{p}_k} \rho_k(Z_{k,i}, \hat{\theta}_k), \\ \check{Y}_i = \sum_{k=1}^K \mathbb{I}_{p_{k-1} < U \leq p_k} \rho_k(Z_{k,i}, \hat{\theta}_k), \\ \tilde{Y}_i = \sum_{k=1}^K \mathbb{I}_{p_{k-1} < U \leq p_k} \rho_k(Z_{k,i}, \theta_{k,*}), \end{array} \right. \tag{14}$$

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:

$$T_2 \leq D_2 + R_{2,1} + R_{2,2} + R_{2,3} + R_{2,4}, \tag{15}$$

where

$$\begin{array}{rcl} D_2 & := & \left| \dfrac{1}{n} \displaystyle\sum_{i=1}^n \left[ \Psi(\tilde{Y}_i) - E_{f_*}(\Psi) \right] \right|, \\ R_{2,1} & := & \dfrac{1}{n} \displaystyle\sum_{i=1}^n \left| \hat{F}(Y_i) - F_*(Y_i) \right|, \quad R_{2,2} := \dfrac{1}{n} \displaystyle\sum_{i=1}^n \left| \bar{F}(Y_i) - F(Y_i) \right|, \\ R_{2,3} & := & \left| \dfrac{1}{n} \displaystyle\sum_{i=1}^n \left[ \Psi(Y_i) - \Psi(\check{Y}_i) \right] \right|, \quad R_{2,4} := \left| \dfrac{1}{n} \displaystyle\sum_{i=1}^n \left[ \Psi(\check{Y}_i) - \Psi(\tilde{Y}_i) \right] \right|, \end{array}$$

The three first terms in the right hand side of (15) being similar to the three first terms in (13) 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 $\Delta\pi := \sum_{k=1}^K |\hat{\pi}_k - \pi_k|$,

$$\mathbb{I}_{Y_i \neq \check{Y}_i} = \sum_{k=1}^K (\mathbb{I}_{\hat{p}_{k-1} \wedge p_{k-1} < U_i < \hat{p}_{k-1} \vee p_{k-1}} + \mathbb{I}_{\hat{p}_k \wedge p_k < U_i < \hat{p}_k \vee p_k}), \quad \text{and} \quad \mathcal{L}\left( \mathbb{I}_{Y_i \neq \check{Y}_i} \,\middle|\, \mathcal{F}_n \right) \sim \mathcal{B}(\Delta\pi).$$

Let us remark that:

$$\Delta\pi \leq \|\hat{\vartheta} - \vartheta_*\|. \tag{16}$$

Term $R_{2,4}$. Using the mean value Theorem and the fact that $\Psi'$ 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 $N_\varepsilon$ such that $P(\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma) \leq \varepsilon$, for all $n \geq N_\varepsilon$. Since $\left\|\hat{\vartheta} - \vartheta_*\right\| = O_P(1/\sqrt{n})$ there exists $\kappa > 0$ such that for all $\delta > 0$ there exists $N_\delta$ ensuring $P(A_n^c) \leq \delta$ for all $n \geq N_\delta$ where $A_{n,\kappa} := \left\{\sqrt{n}\left\|\hat{\vartheta} - \vartheta_*\right\| < \kappa\right\}$. Let us consider $\gamma > 0$ large enough such that:

$$\frac{\kappa}{(\gamma - \kappa)^2} \leq \frac{\varepsilon}{4J}, \quad \frac{V}{\left(\frac{\gamma}{C\kappa} - m\right)^2} \leq \frac{\varepsilon}{4J}, \quad \text{and} \quad \gamma > \max(1, M)\kappa. \tag{17}$$

Then,

$$\begin{aligned} P(\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma) &= P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \cap A_{n,\kappa}\right) + P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \cap A_{n,\kappa}^c\right) \\ &\leq P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \,\middle|\, A_{n,\kappa}\right) + P\left(A_{n,\kappa}^c\right). \end{aligned} \tag{18}$$

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

We now establish an upper bound for the first term in the right hand side of (18) by

$$\begin{aligned} &P\left(\left\{\sqrt{n}\Delta(\hat{f}, f)| \geq \gamma\right\} \,\middle|\, A_{n,\kappa}\right) \\ &\leq P(\sqrt{n}M_1 \geq \gamma | A_{n,\kappa}) \\ &+ \sum_{l=1}^{J} \left(P(\sqrt{n}M_2(l) \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{2,3}(l) \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{2,4}(l) \geq \gamma | A_{n,\kappa})\right). \end{aligned} \tag{19}$$

Using similar reasoning for establishing bounds on the terms involving $M_2$ (as used for terms involving $M_1$ in (19) ), we note, according to the definition of $R_{1,1}$ and the last

condition in (17), that

$$\begin{aligned} P(\sqrt{n}M_1 \geq \gamma | A_{n,\kappa}) &\leq P(\sqrt{n}D_1 \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{1,1} \geq \gamma | A_{n,\kappa}) + P(\sqrt{n}R_{1,2} \geq \gamma | A_{n,\kappa}) \\ &\leq \frac{P(\sqrt{n}D_1 \geq \gamma)}{P(A_{n,\kappa})} + P\left(A^c_{n,\gamma \max(M,1)^{-1}} | A_{n,\kappa}\right) + \frac{P(\sqrt{n}R_{1,2} \geq \gamma)}{P(A_{n,\kappa})} \\ &\leq \frac{2\xi}{1-\delta} \leq \frac{\varepsilon}{4(J+1)}. \end{aligned} \tag{20}$$

The terms involving $R_{2,3}$ in (19) are handled by applying the Tchebychev's inequality:

$$\begin{aligned} &P(\sqrt{n}R_{2,3} \geq \gamma | A_{n,\kappa}) \\ &\leq P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{Y_i \neq \check{Y}_i} \geq \gamma | A_{n,\kappa}\right) \\ &= P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{I}_{Y_i \neq \check{Y}_i} - E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | \mathcal{F}_n)\right) \geq \gamma - \sqrt{n}E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | \mathcal{F}_n) \,\middle|\, A_{n,\kappa}\right) \\ &\leq \frac{\mathrm{Var}\left(\sum_{i=1}^{n} \mathbb{I}_{Y_i \neq \check{Y}_i} \,\middle|\, A_{n,\kappa}\right)}{n(\gamma - \sqrt{n}E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | A_{n,\kappa}))^2} \\ &\leq \frac{\kappa}{(\gamma - \kappa)^2} \leq \frac{\varepsilon}{4J}, \end{aligned} \tag{21}$$

since $E(\mathbb{I}_{Y_1 \neq \check{Y}_1} | A_{n,\kappa}) \leq \kappa/\sqrt{n}$ due to remark (16), and first condition in (17) is supposed. The $R_{2,4}$-depending terms in (19) are handled in a similar manner by using again the Tchebychev's inequality:

$$\begin{aligned} &P(\sqrt{n}R_{2,4} \geq \gamma | A_{n,\kappa}) \\ &\leq P\left(\frac{C' \left\|\hat{\vartheta}_k - \vartheta_{k,*}\right\|}{\sqrt{n}} \sum_{i=1}^{n} W_i \geq \gamma \,\middle|\, A_{n,\kappa}\right) \\ &\leq P\left(\sum_{i=1}^{n} (W_i - m) \geq \gamma \frac{\sqrt{n}}{C' |\hat{\vartheta}_k - \vartheta_{k,*}|} - m \,\middle|\, A_{n,\kappa}\right) \\ &\leq P\left(\sum_{i=1}^{n} (W_i - m) \geq \frac{\gamma n}{C\kappa} - m\right) \\ &\leq \frac{\mathrm{Var}\left(\sum_{i=1}^{n} W_i\right)}{n\left(\frac{\gamma}{C\kappa} - \frac{m}{n}\right)^2} \\ &\leq \frac{V}{\left(\frac{\gamma}{C\kappa} - m\right)^2} \leq \frac{\varepsilon}{4J}, \end{aligned} \tag{22}$$

according to the second condition in (17). The proof of i) is ended by collecting results (17–22) and taking $\delta = \varepsilon/4$.

ii) First we have:

$$\begin{aligned} P(\hat{j} \neq j_*) &= P\left(\cup_{1\leq j\neq j_*\leq J}\left\{\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right\}\right) \\ &\leq \sum_{1\leq j\neq j_*\leq J} P\left(\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right). \end{aligned} \tag{23}$$

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

$$\begin{aligned} &\left\{\widehat{ICE}(\hat{f}_{j_*}, f) > \widehat{ICE}(\hat{f}_j, f)\right\} \\ &= \left\{\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f) + ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f) > \Delta_{j_*,j}\right\} \\ &\supseteq \left\{|\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f)| + |ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f)| > \Delta_{j_*,j}\right\} \\ &\supseteq \left\{|\widehat{ICE}(\hat{f}_{j_*}, f) - ICE(f_{j_*}, f)| > \frac{\Delta_{j_*,j}}{2}\right\} \cup \left\{|ICE(f_j, f) - \widehat{ICE}(\hat{f}_j, f)| > \frac{\Delta_{j_*,j}}{2}\right\} \end{aligned}$$

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 $N_\delta$ 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

$$n_j := \min\left\{n \in \mathbb{N} : \ \frac{\Delta_{j_*,j}}{2} \geq \frac{K}{\sqrt{n}}\right\},$$

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 (23), for all $n \geq N_\varepsilon$:

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

which concludes the proof.

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

### 3.1 A Monte Carlo finite sample size approach

When there is no hope to get large observed datasets $x_1^n := (x_1, \ldots, x_n)$, (*e.g.*, if $n \leq n_0$ for some technical reasons) the spirit of the method previously described can nevertheless be used in a very natural way. 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 model $\hat{f}_j(x, \hat{\vartheta}_j(x_1^n))$ with $f_0$ , the unknown pdf of the observations $(x_1, \ldots, x_n)$, by generating independently $N$ i.i.d. 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.* $mICE_j(x_1^n) := E\left(\widehat{ICE}(\hat{f}_j, f)) \,\middle|\, \{X_1^n = x_1^n\}\right)$, by the empirical mean

$$\widehat{mICE}_j(x_1^n) := \frac{1}{N} \sum_{\ell=1}^{N} \widehat{ICE}_\ell(\hat{f}_j, f)), \tag{24}$$

where $\widehat{ICE}_\ell$ corresponds to expression (6) when taking $Y_{1,\ell}^n$ instead of $Y_1^n$. Moreover since the samples $(Y_{1,\ell}^n)_{1 \leq \ell \leq N}$ are mutually independent and the random variables $0 \leq \widehat{ICE}_\ell(\hat{f}_j, f) \leq 1$ we have the Central Limit Theorem (CLT):

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

which allows us to derive classically parametric bootstrap confidence intervals. In the next section dedicated to a simulation study, we will compare the $\widehat{mICE}$ quantity to its counterpart based on the standard Kolmogorov-Smirnov satistics (KS) and the Shannon-Jensen distance (SJ), defined respectively by

$$\widehat{mKS}(x_1^n) = \frac{1}{N} \sum_{\ell=1}^{N} \widehat{KS}_\ell(\hat{f}_j, f), \quad \widehat{mSJ}(x_1^n) = \frac{1}{N} \sum_{\ell=1}^{N} \widehat{SJ}_\ell(\hat{f}_j, f). \tag{25}$$

Let us mention that the SJ distance is obtained by considering the square root of the SJ divergence whose definition and kernel estimator are given in [34], expressions (13-14).

## 4 Simulation Study

### 4.1 Large sample Monte Carlo study

In order to study the qualitative finite sample properties of criterion ICE, we suggest to test and compare this method to the Kolmogorov-Smirnov (KS) and Shannon-Jensen (SJ)

criteria. Let us use the notation $NG[n_1, n_2](\vartheta_{n_1,n_2})$ to define a generic mixture of $n_1$ Gaussian distributions and $n_2$ Gumbel distributions where $\vartheta_{n_1,n_2} := (\pi_{n_1,n_2}, (\theta_i)_{i=1}^{n_1}, (\phi_j)_{j=1}^{n_2})$ is composed by $\pi_{n_1,n_2} := (\pi_1, \ldots, \pi_{n_1+n_2-1})$ the weights vector of the mixture, and $(\theta_i)_{i=1}^{n_1}$, respectively $(\phi_j)_{j=1}^{n_2}$, the collections of parameters corresponding to the Gaussian and Gumbel distributions respectively. We propose to test our criterion on two benchmark models:

**(M1)**: $NG[1, 2](\vartheta_{1,2})$, with $\pi_{1,2} = (1/2, 1/4)$, $\theta_1 = (2, 0.5)$, $\phi_1 = (0, 2)$, $\phi_2 = (4, 3)$.
**(M2)**: $NG[2, 1](\vartheta_{1,2})$, with $\pi_{1,2} = (1/3, 1/3)$, $\theta_1 = (0, 1)$, $\theta_2 = (2, 0.5)$, $\phi_1 = (3, 2)$.

For these two models we consider three models in competition labeled by $j$: when $j = 1, 2, 3$ we consider respectively $NG[1, 1]$, $NG[2, 1]$, $NG[1, 2]$.

The motivation for prescribing the aforementioned mixtures is two fold. First, the mixtures specified are a combination of symmetric (Gaussian) and asymmetric (Gumbel) distributions. Also, the parameters have been specified to ensure overlap of individual populations of components of the mixture i.e. the heterogeneity of the mixed symmetric and asymmetric densities is not trivial to detect compared to datasets where individual populations are location separate. Second, we expect a similar qualitative nature of the mixture to be applicable to the real dataset considered in Section 4. For each mixture, an initial random variable of size $n$ is generated with the corresponding parameters. The parameters of each of the three models in competition are estimated using the iterative EM algorithm. The initial values for the parameters, for the EM algorithm, of each of the three models in competition are assumed to coincide with the parameters of the mixtures being prescribed. In the absence of a closed form solution for updating the parameters during each iteration of the maximization step, we employ a method in which a discretized search in the neighborhood of the current parameters is performed. 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 $\pm 0.1$ with increments of 0.01 for the location parameter and $\pm 0.01$ 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. Note that the maximization of the global log-likelihood implies the maximization of the likelihood of the individual components of the mixture.

The resulting estimate of the mixture model is used to generate a sample of size $n$, whose distance from the initial random variable is indicative of the accuracy of the estimated parameters. An example of the nonparametric density distributions of the initial random variable and a sample generated by estimating parameters of different models from the EM algorithm is shown in Fig. 1.

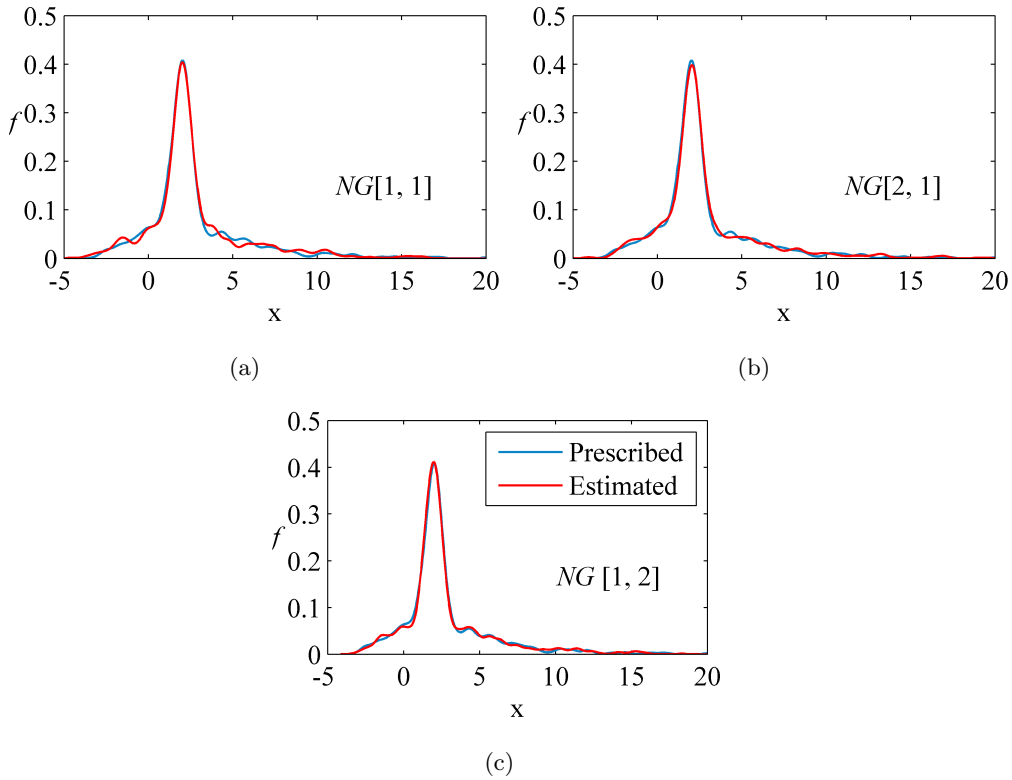



Figure 1: Nonparametric densities of random variables generated using parameters of mixture M1 and parameters estimated using the EM algorithm (a) model 1, (b) model 2 and (c) model 3. The mixture corresponds to model 3.

Subsequent to estimation of the model parameters using the EM algorithm, multiple (100) samples are generated using these parameters.The objective of generating multiple samples with the same estimated parameter set for a given model is to account for the inherent inaccuracies in the random variable generation. Each of these samples is compared to the initial random variable using the three metrics - KS, SJ and ICE criteria. The average of these realizations is assumed to smear out the uncertainties associated with one realization of random variable generation. For each metric (KS, SJ and ICE), the average metric is compared for each of the models. A metric is deemed to have selected a model correctly if the average of the metric for the correct model, known a priori, is the lowest. This exercise is repeated for 100 such initially generated random variables of the same assumed mixture model. The success rate i.e. thus number of times each metric identifies the correct model for different initial random variable sizes ($n$) is plotted in Fig. 2. It can be seen that the success rate of all metrics seems to increase with the sample size. However, the ICE criterion identifies the correct model most of the times.

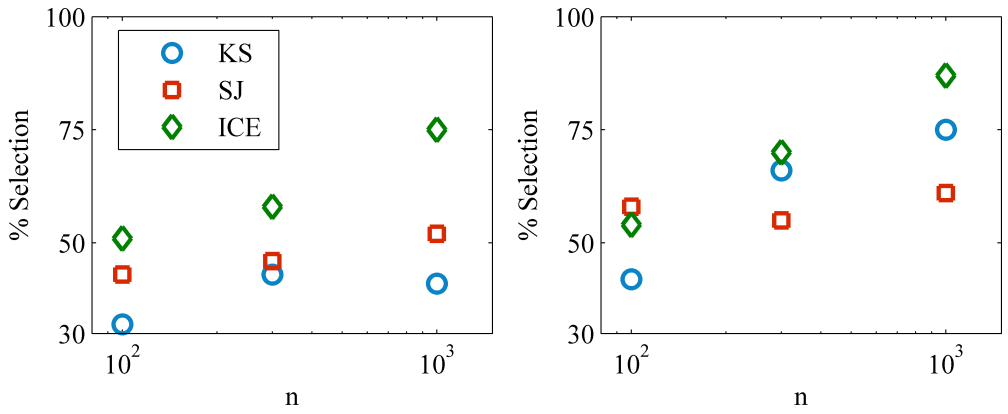



Figure 2: Success rates of the metrics in identifying the correct model for (a) Mixture **(M1)** and (b) Mixture **(M2)** for sample size $n$.

It is to be noted that the $\widehat{ICE}$ and $\widehat{SJ}$ estimators are domain average-type estimators which are less sensitive than the $\widehat{KS}$ statistics to abrupt changes in the cdf, which are in general difficult to track with empirical cdfs. Further, since the non-parametric density estimates used for the $\widehat{SJ}$ estimator are bandwidth dependent, it is likely that the $\widehat{SJ}$ estimator suffers from excess smoothing in areas with abrupt changes in the cdf i.e.,

the statistics generated could be bandwidth dependent. However, the $\widehat{ICE}$ estimator is completely free from any bandwidth dependence, contrary to the $\widehat{SJ}$ estimator.

### 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 material anisotropy displayed at the scale of individual grains. More precisely, we characterize the distribution of the extreme values of the shear stress resolved along specific crystallographic (slip) directions, averaged over a grain. For simplicity, we have considered an idealized grain structure (cubes) subject to monotonic deformation with linear elastic material properties. A schematic of the microstructure generated is shown in Fig. 3. In the computational cell shown in Fig. 3, the different color codes of each grain indicate that different orientations are assigned to each grain. The orientation of each grain is specified by a set of Euler angles, using the Bunge notation, which determine the orientation of the crystallographic directions of each grain with respect to a reference frame. Here, the reference frame coincides with the axes describing the computational cell and the Euler angles are sampled from a uniform orientation distribution function. Since linear elastic behavior is being considered, only the elastic constants for the material need to be specified to define the 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, i.e. the material has three independent elastic constants defined in the orientation frame of each grain. The grain averaged resolved shear stresses are found along each of the possible 12 octahedral slip directions. The single crystal elastic constants have been adopted from [26].

Previous studies in quantifying the distribution of fatigue indicator parameters [12] have considered volume averages of driving forces that include microscale cyclic plasticity in the microstructure obtained from computational experiments. Also, the distribution of the microscale cyclic plasticity in a computational volume has also 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, since crack initiation mechanisms that occur on favorably oriented crystallographic directions

(along which the grain averaged resolved shear stresses are calculated) are being considered, the extremal value of this grain averaged resolved shear stress would be an indicator of the potency of fatigue crack formation in a particular grain. 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. A detailed analysis of the effect of grain size variation and deformation amplitude on the nature of the distributions indicating fatigue crack potency is left to a later study. It is to be noted that the computational cell, shown in Fig. 3, serves as a statistical volume element (SVE) i.e. a given computational cell is not large enough to capture all the statistical variations of the fatigue crack formation potency for all combinations of orientation descriptions for the grains and inter-granular interactions. Thus, multiple realizations are required to obtain the distribution of the extremal values indicative of the fatigue crack formation potency.

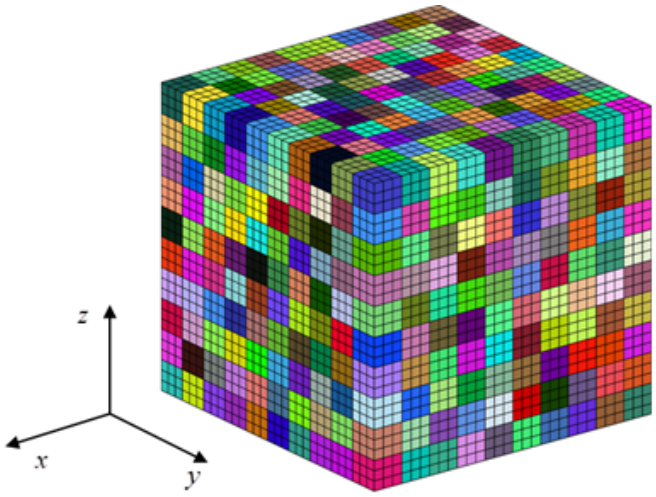



Figure 3: Realization of the idealized microstructure used for the finite element simulations.

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. For the free surface boundary conditions, the top and bottom faces of the

computational cell shown in Fig. 3 remain horizontal with displacement imposed on the top face along the $z$ direction and the bottom face fixed. No restrictions are placed on the lateral faces i.e. these faces act as free surfaces. For the periodic boundary conditions, displacement is imposed along the $z$ direction on the top face of the computational cell with the constraint that opposing faces of the computational cell remain parallel to each other. The computational cell size is varied from 5 grains in each dimension to 10 grains in each dimension with 27 finite elements for each grain. The elements used are 8-node brick elements with linear interpolation and reduced integration (C3D8R). The grain averaged values of resolved 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 data set in the subsequent subsection.

### 4.3 Application to 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 $\partial B$. The input $U(r)$ is considered random since each computational realization has randomly assigned grain orientations. We denote by

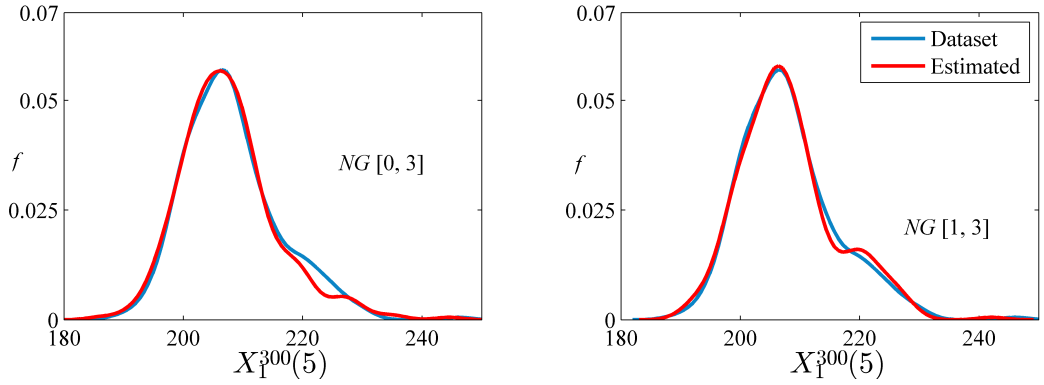
$$X(r) := \max_{1\le k,l,m\le r} R_{k,l,m}.$$

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

$\partial B$. 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.

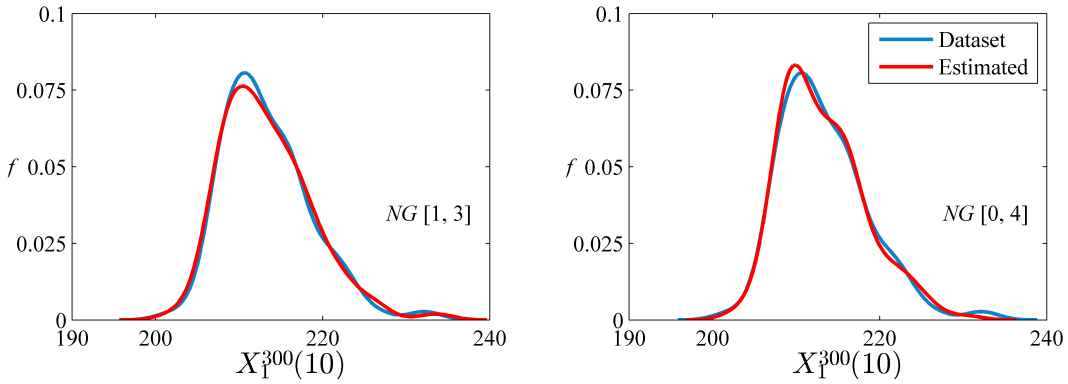
The method of identifying the best model to estimate the underlying mixture of $X_1^n(r)$ is similar to the one outlined in the previous section. We start by assuming various possible models that would describe the underlying mixture and estimate the parameters of the model using the EM algorithm. The limits and increments of the location and scale parameters used to define the grid of points over which the search is performed to optimize the weighted log-likelihood remain the same. However, the tolerance used as criterion for the convergence of the method, based on the variation of the global log-likelihood of the model being estimated for fifty successive iterations of the EM algorithm, 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).

To explain the existence of a mixture of distributions, we recall that the dataset is a collection of the extreme values of the grain averaged resolved shear stresses for multiple realizations of numerical experiments. Consider a computational cell (assembly of grains) where the entire assembly is deformed homogeneously i.e. the deformation of each grain is equal to the imposed macroscopic deformation. In this case, the extreme value of the resolved shear stress would be a function of the grain orientations alone and a collection of such extreme values would converge to a unique distribution. However, in the present case, the interaction of grains, more specifically, the kinematic constraints resulting from the difference in the grain orientations, coupled with the boundary conditions introduce perturbations from homogeneous deformation of the assembly. This results in the convergent extreme value distribution being corrupted that manifests as a mixture of distributions. We present a conjecture later for multiple distributions acting as domains of attraction.




(a) (b)

Figure 4: Comparison of the nonparametric density estimates for the dataset $X_1^{300}(5)$, the set of extreme values of the grain averaged resolved shear stress from 300 numerical experiments for the 125 grain assembly with free surface BC and the random sample with parameters estimated from EM algorithm. Assumed model (a) $NG[0, 3]$ and (b) $NG[1, 3]$.




(a) (b)

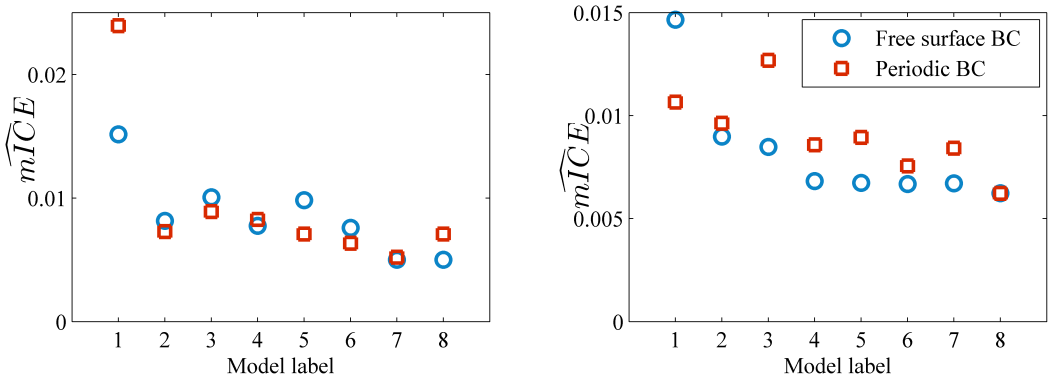
Figure 5: Comparison of the nonparametric density estimates for the dataset $X_1^{300}(10)$, the set of extreme values of the grain averaged resolved shear stress from 300 numerical experiments for the 1000 grain assembly with periodic BC and the random sample with parameters estimated from EM algorithm. Assumed model (a) $NG[1, 3]$ and (b) $NG[0, 4]$.

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 number of models are considered, in competition, to obtain the best fit to the dataset, $X_1^n(r)$. The models being considered for the estimation of the mixture 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 outlined earlier. Subsequent to determining the parameters of each model, a random sample of size $n$ is generated to find the distance ($\widehat{ICE}$) from the dataset, $X_1^n(r)$. This process is repeated for 100 samples generated to yield the average distance, $\widehat{mICE}$, for each model. The rationale for using multiple sample to arrive at an averaged metric is similar to the one presented in the earlier study - to smear out the inherent randomness of one realization of sample generation. This approach assumes significance in the current context of a small sample size ($n = 300$). The model that most accurately fits the dataset is assessed based on the minimum $\widehat{mICE}$ for all the models considered. The accuracy of $\widehat{mICE}$ in identifying the correct model that describes the underlying mixture of datasets has already been demonstrated. This approach of considering models in competition allows the ranking of the different models considered. The $\widehat{mICE}$ for all the models considered for different sizes of the computational cell is plotted in Fig. 6. The values of the $\widehat{mICE}$ for all the models and combinations of computational cell size and boundary conditions are listed in Table 3.

From Fig. 6, it is apparent that a better fit to the data is obtained by using a higher number of distributions. In almost all cases, a single Gumbel distribution provides the least accurate description of the observed distribution. This is particularly important since the fatigue life distributions are often fit to a single extreme value distribution [16, 17, 18, 19]. The accuracy of a single Gumbel distribution in describing the distribution is dependent on the nature of the boundary conditions and the computational cell size. The periodic BC show a marked increase in the accuracy of a single Gumbel distribution describing the distribution, compared to free surface BC, for an increased computational cell size. We theorize that the extent of perturbation from a homogeneously deformed assembly, and the consequent deviation from a single extreme value distribution describing the data, is dependent on both the inter-granular interactions and the boundary conditions. For the case of the 1000 grain assembly with periodic BC, we have noticed that in more than half of the realizations, the location of the grains where the extreme values of the resolved shear




(a) (b)

Figure 6: Variation of $\widehat{mICE}$ of different assumed mixture models for computational cells consisting of (a) 125 grains and (b) 1000 grains.

stresses occur, do not lie on any of the boundaries. It can be argued that the increase in the computational cell size for the periodic BC reduces the deviation of the observed extreme value from the idealized scenario (one extreme value distribution describing the dataset) by mitigating one of causes - boundary conditions.

It is also noteworthy that the same mixture model predicts 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 since the extreme values are being sampled from a smaller population of outliers, as explained later. Note that the models involving the normal distribution are identified as the best descriptors of the distribution for the smaller computational cell size (125 grains).

*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* $\mathcal{A}_{Gumbel} = \{\mathcal{C}_i\}_{i=1,\dots,L_1}$ and *Normal attraction domain* $\mathcal{A}_{Normal} = \{\mathcal{C}_i\}_{i=L_1+1,\dots,L}$, justified as follows:

- given $\{U(r) \in \mathcal{C}_i\}$, $i = 1,\dots,L_1$, the set of resolved shear stresses from a numer-

ical 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 [29] 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 a (possibly) slower rate, *i.e.* $O(\log(n)^{-1})$, see Han and Ferreira [30], 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:

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

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 $(\mathcal{C}_j)_{1\leq j\leq L}$, 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 datasets containing a mixture of distributions. The parameters of the model are identified using the EM algorithm. A generalized approach has been presented for the maximization step of the EM algorithm through a neighborhood search method that is valid for models with a mixture of distributions from different families, for which closed form solutions to update the parameters do not exist. It is to be noted that the approach is also valid for

mixtures from the same family of distributions. To test the accuracy of the models in describing the mixture, a novel metric, the Integrated Cumulative Error ($\widehat{ICE}$) has been defined. The $\widehat{ICE}$ metric has been shown to be more efficient in identifying the correct model that describes the underlying mixture than commonly used approaches, such as the Kolmogorov-Smirnov statistic.

The approach developed is used to identify the underlying mixture of the distribution of indicators of fatigue crack formation potency (grain averaged resolved shear stresses), based on linear elastic analysis for polycrystals with idealized grain structure and elastic anisotropy. The observations indicate that a mixture model characterizes the distribution more accurately than a single extreme value distribution, which is commonly followed. 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. Further, refining the computational models to (a) better describe the variation of microstructure and (b) account for damage accumulation through cyclic plasticity might provide more insight into the fatigue life distribution of a material. Nevertheless, the use of the methods developed in this work would better characterize the tails of the distributions 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.

# 6 Acknowledgments

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

# 7 Appendix 1. Behavior of the MLE when the model is possibly misspecified

In this section, we briefly recall some basic material, from White [39], 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 i.i.d. 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 $\vartheta$ in $\Theta$ a compact subset of $\mathbb{R}^p$, and continuous in $\vartheta$ for all $x \in \mathbb{R}$. Next, we define the quasi-log-likelihood of the sample as

$$L_n(X_1^n, \vartheta) := \frac{1}{n} \sum_{i=1}^{n} \log f(X_i, \vartheta), \tag{26}$$

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

$$\hat{\vartheta}_n = \arg\max_{\vartheta \in \Theta} L_n(X_1^n, \vartheta). \tag{27}$$

In Theorem 2.1, White [39] 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 $\vartheta_0$ in the interior of $\Theta$), the MLE is consistent for $\vartheta_0$ under suitable conditions, see *e.g.* Theorem 2 in Wald [38], Theorem 5.a in LeCam [32]. 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 $\vartheta_*$ that minimizes the Kullback Leibler [31] divergence ($\mathcal{K}$), *i.e.*

$$\vartheta_* := \arg\min_{\vartheta \in \Theta} \mathcal{K}(f, f_\vartheta), \quad \text{where} \quad \mathcal{K}(f, f_\vartheta) := E\left(\log\left[\frac{f(X_1)}{f(X_1, \vartheta)}\right]\right). \tag{28}$$

To support the Akaike's observation that $\hat{\vartheta}_n$ is a natural estimator for $\vartheta_*$, White [39] impose the additionnal 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 repect to $f_0$.
iii) $K(f, f_\vartheta)$ has a unique minimum at point $\vartheta_*$ in $\Theta$.

When assumption (A3) ii) holds, $\vartheta_*$ is globally identifiable. In Theorem 2.2, White [39] establishes, under assumptions A1–A3, the strong $\vartheta_*$-consistency of the QMLE defined in (27), *i.e.*

$$\hat{\vartheta}_n \xrightarrow{a.s.} \vartheta_*, \quad \text{as} \quad n \to \infty. \tag{29}$$

With additionnal conditions (given below), White [39] 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 $\vartheta$ for each $x$ in $\mathbb{R}$.

**Assumption (A5)**. The two collections $\left\{\left|\partial^2 \log((f(x,\vartheta))/\partial\vartheta_k\partial\vartheta_l\right|,\ k,l=1,\ldots,p\right\}$ 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 $\vartheta_*$ is an interior point of $\Theta$.

ii) The $p \times p$ matrix $B(\vartheta_*)$ is nonsingular.

iii) The parameter $\vartheta_*$ is a regular point of $A(\vartheta)$.

Under assumptions A1–6, White ([39], Theorem 3.2) establishes the asymptotic normality of the QMLE, *i.e.*

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_*) \xrightarrow{d} \mathcal{N}(0, C(\vartheta_*)), \quad \text{as} \quad n \to +\infty. \tag{30}$$

*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 (27) and (30) when replacing $\vartheta_*$ by $\vartheta_0$.

# 8 Appendix 2. Mixtures of Gaussian and Gumbel distributions

## 8.1 Identifiability

In this section, 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:

$$\mathcal{L} := \{L(x;\theta) :\ x \in \mathbb{R}, \theta \in \Theta\}, \quad \mathcal{G} := \{G(x;\phi) :\ x \in \mathbb{R}, \phi \in \Psi\} \tag{31}$$

where $\Theta$ and $\Psi$ 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:

$$\mathcal{H} := \left\{ H(x) = \sum_{i=1}^{n_1} c_i L_i(x) + \sum_{j=n_1+1}^{n_1+n_2} c_j G_j(x),\ \sum_{i=1}^{n_1+n_2} c_i = 1,\ c_i > 0, (n_1+n_2) \in \mathbb{N}^* \times \mathbb{N}^* \right\}. \tag{32}$$

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 $D_\alpha$ and $D_\beta$ (the domains of definition of $\alpha$ and $\gamma$) such that the mappings $L \to \alpha$ and $G \to \gamma$ are linear and one-to-one. We denote by $I_\alpha$ and $I_\gamma$, the largest interval contained respectively in $D_\alpha$ and $D_\beta$. Let us denote for all $F \in \mathcal{F} \cup \mathcal{G}$ by $\rho_F$, 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 $\rho_{F_2}$) 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 (1963). ■

**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

$$\gamma(t) = e^{\mu t} \Gamma(1 - \beta t), \qquad \beta t \notin \mathbb{N}^*,$$

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 [37] 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. ■

## 8.2 Checking the assumptions

*Assumption G.* For the Normal distribution and the Gumbel distribution, it is enough to simulate random variables according to $\mathcal{N}(0, 1)$ and respectively, $\mathcal{G}(0, 1)$ distribution, and consider the transformation $\rho(y, m, \sigma) := (y - m)/\sigma$ and $\rho(y, \mu, \beta) := (y - \mu)/\beta$. Moreover the condition (8) 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 can propose a similar upper-bound whose construction is detailed hereafter. 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 $\exp(x/\beta) > \exp(x/\overline{\beta}) > 1$ 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

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

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 8 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.

| Model $j$ | $n$ | Mean | Std Dev | %-selection |
|---|---|---|---|---|
| $j=1$ | 100 | $(0.0141, 0.0980, 0.0392)$ | $[0.0034, 0.0046, 0.0071]$ | $\{13, 25, 1\}$ |
| $j=2$ | 100 | $(0.0120, 0.0967, 0.0344)$ | $[0.0022, 0.0043, 0.0071]$ | $\{36, 43, 56\}$ |
| $j=3$ | 100 | $(0.0116, 0.0969, 0.0351)$ | $[0.0021, 0.0044, 0.0065]$ | $\{51, 32, 43\}$ |
| $j=1$ | 300 | $(0.0083, 0.0582, 0.0108)$ | $[0.0021, 0.0031, 0.0016]$ | $\{13, 21, 2\}$ |
| $j=2$ | 300 | $(0.0069, 0.0571, 0.0094)$ | $[0.0015, 0.0025, 0.0018]$ | $\{29, 37, 52\}$ |
| $j=3$ | 300 | $(0.0064, 0.0568, 0.0096)$ | $[0.0011, 0.0025, 0.0016]$ | $\{58, 43, 46\}$ |
| $j=1$ | 1000 | $(0.0050, 0.0326, 0.0039)$ | $[0.0014, 0.0014, 0.0005]$ | $\{8, 14, 2\}$ |
| $j=2$ | 1000 | $(0.0039, 0.0317, 0.0034)$ | $[0.0008, 0.0014, 0.0005]$ | $\{17, 45, 46\}$ |
| $j=3$ | 1000 | $(0.0035, 0.0316, 0.0034)$ | $[0.0006, 0.0015, 0.0005]$ | $\{75, 41, 52\}$ |

Table 1: **(M1)** Mean and Std. Dev. of 100 estimates of $\widehat{ICE}$, $\widehat{KS}$, $\widehat{SJ}$ and rate of selection.

| Model $j$ | $n$ | Mean | Std Dev | %-selection |
|---|---|---|---|---|
| $j=1$ | 100 | $(0.0187, 0.1070, 0.0313)$ | $[0.0065, 0.0134, 0.0068]$ | $\{2, 10, 5\}$ |
| $j=2$ | 100 | $(0.0103, 0.0951, 0.0250)$ | $[0.0013, 0.0037, 0.0048]$ | $\{54, 42, 58\}$ |
| $j=3$ | 100 | $(0.0106, 0.0953, 0.0254)$ | $[0.0018, 0.0035, 0.0050]$ | $\{44, 48, 37\}$ |
| $j=1$ | 300 | $(0.0151, 0.0706, 0.0094)$ | $[0.0068, 0.0141, 0.0027]$ | $\{1, 6, 2\}$ |
| $j=2$ | 300 | $(0.0056, 0.0553, 0.0064)$ | $[0.0009, 0.0021, 0.0011]$ | $\{70, 66, 55\}$ |
| $j=3$ | 300 | $(0.0061, 0.0566, 0.0064)$ | $[0.0011, 0.0025, 0.0010]$ | $\{29, 28, 43\}$ |
| $j=1$ | 1000 | $(0.0146, 0.0512, 0.0053)$ | $[0.0068, 0.0169, 0.0020]$ | $\{0, 1, 0\}$ |
| $j=2$ | 1000 | $(0.0031, 0.0309, 0.0023)$ | $[0.0005, 0.0012, 0.0003]$ | $\{87, 75, 61\}$ |
| $j=3$ | 1000 | $(0.0039, 0.0323, 0.0024)$ | $[0.0007, 0.0017, 0.0003]$ | $\{13, 24, 39\}$ |

Table 2: **(M2)** Mean and Std. Dev. of 100 estimates of $\widehat{ICE}$, $\widehat{KS}$, $\widehat{SJ}$ and rate of selection.

| 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 |

Table 3: $\widehat{mICE}$ of 100 estimates of $ICE$ of different mixtures for various computational cell sizes and boundary conditions. FS denotes free surface and PBC denotes periodic boundary conditions. The number denotes the number of grains in the computational cell.

## 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.

[2] Izenman, A. J. and Sommer, C. (1988) Philatellic 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.

[4] Lindsay, B. G. (1983). Moment matrices: application in mixtures. *Ann. Statist.*, **17**, 722–740.

[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.

[6] Keribin, C. (2000) Consistent Estimation of the Order of Mixture Models *Sankhya Series A*, **62**, 49–66.

[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.

[8] Vuong, Q. H.. (1989) Likelihood ratio test for model selection and non-nested hypothesis. *Econometrica*. **57**, 307–333.

[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 Rene 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.

[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.

[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.

[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. Budapest: Akadémiai Kiado.

[24] Babu, G. J. (2011). Resampling method for model fitting and model selection. *J. Biopharma. Statist.,* **21**, 1177–1186.

[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 Grard, 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.

[28] ABAQUS FEA, V6.7.1. *D S Simulia, Dassault Systmes, Providence, RI.*

[29] Dabrowski, A. R. (1990) Extremal Point Processes and Intermediate Quantile Functions. *Probab. Theory Related Fields*, **85**, 365–386.

[30] Han, L. and Ferreira, A. (2006). *Extreme Value Theory.* New-York, Springer.

[31] Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Ann. Math. Statist.*, **22**, 79–86.

[32] LeCam, L. (1953). On some Asymptotic Properties of Maximum Likelihood estimates and related Bayes' Estimates. *University of California Publications in statistics*, **1**, 277–330.

[33] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[34] 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.

[35] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester.

[36] VAN DER VAART. A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New-York.

[37] TEICHER, H. (1963). Identifiability of finite mixtures. *Ann. Math. Stat.*, 34, 1265–1269.

[38] WALD, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *Ann. Math. Statist.*, **60**, 595–603.

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