Profile Likelihood Estimation of the Vulnerability $P(X{>}v)$ and the Mixing Proportion p Parameters in the Gumbel Mixture Model

Estimación de verosimilitud perfil de los parámetros de vulnerabilidad $P(X{>}v)$ y proporción de mezcla p en el modelo Gumbel de mezclas

José A. Montoya[1,a], Gudelia Figueroa[1,b], Nuša Pukšič[2,c]

[1]Departamento de Matemáticas, División de Ciencias Exactas y Naturales, Universidad de Sonora, Hermosillo, México

[2]Institute of Metals and Technology, Ljubljana, Slovenia

Abstract

This paper concerns to the problem of making inferences about the vulnerability $\theta = P(X{>}v)$ and the mixing proportion p parameters, when the random variable X is distributed as a mixture of two Gumbel distributions and v is a known fixed value. A profile likelihood approach is proposed for the estimation of these parameters. This approach is a powerful though simple method for separately estimating a parameter of interest in the presence of unknown nuisance parameters. Inferences about θ, p or (θ, p) are given in terms of profile likelihood regions and can be easily obtained on a computer. This methodology is illustrated through a real problem where the main purpose is to model the size of non-metallic inclusions in steel.

Key words: Invariance principle, Likelihood approach, Likelihood region, Mixture of distributions.

Resumen

En este artículo consideramos el problema de hacer inferencias sobre el parámetro de vulnerabilidad $\theta = P(X{>}v)$ y la proporción de mezcla p cuando X es una variable aleatoria cuya distribución es una mezcla de dos distribuciones Gumbel y v es un valor fijo y conocido. Se propone el enfoque de verosimilitud perfil para estimar estos parámetros, el cual es un método simple, pero poderoso, para estimar por separado un parámetro de interés en presencia de parámetros de estorbo desconocidos. Las inferencias sobre θ, p o (θ, p) se presentan por medio de regiones de verosimilitud perfil y se

[a]Professor. E-mail: montoya@mat.uson.mx
[b]Professor. E-mail: gfiguero@gauss.mat.uson.mx
[c]Research assistant. E-mail: nusa.puksic@imt.si

pueden obtener fácilmente en una computadora. Esta metodología se ilustra mediante un problema real donde se modela el tamaño de inclusiones no metálicas en el acero.

Palabras clave: enfoque de verosimilitud, mezcla de distribuciones, principio de invarianza, región de verosimilitud.

1. Introduction

Facilities such a electric power, water supplies, communications and transportation are a part of what is named society infrastructure, although in a broad definition this also includes some basic societal functions like education, national defense and financial and health systems. On the other hand, the term critical infrastructure is often used to denote the collection of all large technical systems characterized as public, like electric power, water supply systems, transportation, communications and health systems. All these services are considered a part of a nation critical infrastructure and they are essential for the quality of everyday life. Natural disasters, adverse weather conditions, technical failures, human errors, labor conflicts, sabotage, terrorism and many other situations can disturb the appropriate flow of these services and a severe strain on the society could occur. Hence, national security is directly linked to the vulnerability of critical infrastructure, and problems related with human error or technical failures should be prevented. In particular, it is known that steel inclusions formed during the steel production process degrade the mechanical properties of the steel. Special interest is focused on the control of non-metallic inclusions due to their harmful effect, because their size, amount and chemical composition have a great influence on steel properties and are linked to its vulnerability. Actually, big inclusions can turn out to be dangerous, leading to the failure of the finished steel product. The steel industry fixes some critical limits for these inclusions and those limits depend on the purpose of the steel products. The increasing demand for cleaner steels has led to the continuous improvement of steelmaking practices and modeling the type and distribution of these inclusions has become significant concern in the steel industry.

Murray & Grubesic (2007) define vulnerability of an infrastructure system as the probability that at least one disturbance with negative societal consequence X, could be larger than some (critical) value v, during a given period of time T. Hence, they argued that a simple measure for the vulnerability of an infrastructure system can be formulated as

$$P(X > v) = 1 - F(v)$$

where $F(x)$ denotes the probability distribution function of the random variable X. Skewed distributions such as the exponential, lognormal, log-logistic, and power law distributions have been considered by many authors in a number of different real life situations, like Rosas-Casals, Valverde & Solé (2007) and also by Murray & Grubesic (2007). However, mixture models would be preferable when the random variable X is generated from k distinct random processes. To our knowledge, only

few authors like Zheng (2007) and Barrera-Núñez, Meléndez-Frigola & Herraiz-Jaramillo (2008) have used mixture models to explain vulnerability analysis.

In statistics, a mixture model is a probabilistic model adequate for representing the presence of subpopulations within an overall population, and it is not required for the observed data-set to identify the sub-population to which an individual observation belongs. Formally, given a finite set of probability density functions $f_1(x), \ldots, f_k(x)$ and weights $p_1, \ldots, p_k$ where $p_i \geq 0$ and $\sum p_i = 1$, the density associated with a mixture distribution can be written as $f(x) = \sum p_i f_i(x)$. Mixture distributions arise in a natural way in many areas such as engineering science, medicine, biology, hidrology, geology, as shown in Titterington, Smith & Makov (1985) as well as in Lindsay (1995).

The Gumbel distribution occurs as the limit of maxima of most standard distributions, particularly for the normal distribution. Kotz & Nadarajah (2000) describe in detail this distribution. Actually, the Gumbel distribution has been one of the models used for quantifying the risk associated with extreme rainfalls; it has been also used to model flood levels, the magnitude of earthquakes and even sport records. Some recent applications belong to the engineering area, such as in risk-based engineering, software reliability and structural engineering. Mixture models for Gumbel distributions are of special importance. For example, Cheng & Iles (1990) have used a Gumbel mixture model to estimate the seismic risk of the Chinese mainland. Beretta & Murakami (2001) found that a Gumbel mixture model is useful for modeling two types of steel inclusions.

Maximum likelihood estimation for the shape and scale parameters can be found in Evans, Hastings & Peacock (1993) and Johnson, Kotz & Balakrishnan (1994), and parameter estimation for the mixture of two Gumbel distributions is included by Raynal & Guevara (1997), Tartaglia, Caporali, Cavigli & Moro (2005) and Ahmad, Jaheen & Modhesh (2010). However, inferences about the vulnerability parameter $\theta = P(X > v)$ for the Gumbel mixture case has not been carefully studied, despite the actual importance of this kind of analysis. In many applications, inferences about the parameters θ and p, where p is the mixture proportion, can be more relevant than inferences concerning some other model parameters. This will be illustrated with a real data set related to the size of non-metallic inclusions in steel. This data set has two kinds of inclusions, classified as Type 1 and Type 2 inclusions, where p denotes the proportion for the first type of inclusion.

Let the distribution of X be a mixture of two independent Gumbels:

$$f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = pf_1(x; \mu_1, \sigma_1) + (1-p)f_2(x; \mu_2, \sigma_2) \qquad (1)$$

where

$$f_i(x; \mu_i, \sigma_i) = \frac{1}{\sigma_i} \exp\left[-\left(\frac{x - \mu_i}{\sigma_i}\right)\right] \exp\left\{-\exp\left[-\left(\frac{x - \mu_i}{\sigma_i}\right)\right]\right\}$$

$-\infty < \mu_i < \infty$, $\sigma_i > 0$, $i = 1, 2$, $-\infty < x < \infty$, and $0 < p < 1$. A fundamental statistical problem is concerned with making inferences on the vulnerability parameter

$$\theta = P(X > v; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = 1 - \int_{-\infty}^{v} f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p) \, dx \qquad (2)$$

and the mixing proportion p, based on a sample $x_1, \ldots, x_n$ from X. In this paper, we analyze this problem considering $\sigma_1 > 0$, $\sigma_2 > 0$, $0 < p < 1$, $-\infty < \mu_1 < \mu_2 < v$, and v a known fixed value. Our purpose is to estimate these parameters using the profile likelihood function. The profile likelihood approach can be useful in many situations and it is a powerful though simple method for separately estimating a parameter of interest in the presence of unknown nuisance parameters. Inferences about θ, p or (θ, p) can be given in terms of profile likelihood regions which are easily obtained with a computer. This methodology will be illustrated with a real data set concerning the size distribution of non-metallic inclusions in steel.

2. Profile Likelihood Approach

In this section we describe the estimation procedure that will be used to make inferences about the parameters of interest. This approach is based in Sprott (1980), Kalbfleisch (1985), and Sprott (2000). Let $\boldsymbol{x}_o = (x_1, \ldots, x_n)$ be an observed sample from a distribution with likelihood function $L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o)$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{d_\psi})$ represents the parameter of interest and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_\lambda})$ is a nuisance parameter. The profile likelihood function of $\boldsymbol{\psi}$ is

$$L_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right) = L\big[\boldsymbol{\psi}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi}); \boldsymbol{x}_o\big] \qquad (3)$$

The quantity $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi})$ that maximizes $L(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o)$ for a specified value of $\boldsymbol{\psi}$, is called the restricted maximum likelihood estimate of the nuisance parameter $\boldsymbol{\lambda}$.

Usually $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi})$ exists and is unique for each value of $\boldsymbol{\psi}$, so the definition (3) applies. Formally, $L_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right)$ can be defined as

$$L_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right) = \sup_{\boldsymbol{\lambda}} L\left(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o\right) \qquad (4)$$

Since $L\left(\boldsymbol{\psi}, \boldsymbol{\lambda}; \boldsymbol{x}_o\right)$ is proportional to the probability of the observed sample as a function of the parameters of the model, then the supremum exists and it is finite. The profile likelihood function can be used to rank parameter values according to their plausibilities. Now, the relative profile likelihood function of $\boldsymbol{\psi}$ is a standardized version of (4), and takes a value of one at the maximum of the profile likelihood function of $\boldsymbol{\psi}$,

$$R_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right) = \frac{L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}{\sup_{\boldsymbol{\psi}} L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}$$

Hence, the relative profile likelihood varies between 0 and 1. Values of $\boldsymbol{\psi}$ that are supported by the observed sample $\boldsymbol{x}_o$ will result in values of $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o)$ close to one. In contrast, values of $\boldsymbol{\psi}$ with $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o)$ close to zero become implausible, given the sample $\boldsymbol{x}_o$. Morever, if the the maximum likelihood estimate (mle) of

$(\boldsymbol{\psi}, \boldsymbol{\lambda})$ exists and is unique, then the relative profile likelihood function of $\boldsymbol{\psi}$ can be defined as

$$R_P\left(\boldsymbol{\psi}; \boldsymbol{x}_o\right) = \frac{L_P(\boldsymbol{\psi}; \boldsymbol{x}_o)}{L(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}}; \boldsymbol{x}_o)} = \frac{L\left[\boldsymbol{\psi}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\psi}); \boldsymbol{x}_o\right]}{L(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\lambda}}; \boldsymbol{x}_o)}$$

where $\widehat{\boldsymbol{\psi}}$ is the mle of $\boldsymbol{\psi}$. Note that $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\psi}})$ is the ordinary mle of $\boldsymbol{\lambda}$. The relative profile likelihood function is the maximum relative likelihood that ψ can attain when $\boldsymbol{\lambda}$ is unknown and free to vary arbitrarily. Thus, $R_P(\boldsymbol{\psi}; x_o)$ ranks all possible values of $\boldsymbol{\psi}$ according to their maximum plausibilities and supported by the observed data.

A level c profile likelihood region for $\boldsymbol{\psi}$ is given by

$$\mathcal{R}_P(c) = \{\boldsymbol{\psi} : R_P(\boldsymbol{\psi}; x_o) \geq c\} \tag{5}$$

where $0 \leq c \leq 1$. When $\boldsymbol{\psi}$ is a scalar this region will be an interval if R_P is unimodal and the union of disjoint intervals when R_P is multimodal. Each specific value of $\boldsymbol{\psi}$ within this region has an associated relative profile likelihood $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o) \geq c$, and values outside this region will have a relative profile likelihood $R_P(\boldsymbol{\psi}; \boldsymbol{x}_o) < c$. At level c, this region separates plausible values of $\boldsymbol{\psi}$ from the implausible ones. When varying c from 0 to 1, a complete set of nested likelihood regions is obtained and these converges to the mle $\widehat{\boldsymbol{\psi}}$ as $c \to 1$. Computer algorithms are usually used to find the mle or the borders of a profile likelihood regions given in (5).

In most of the cases, a profile likelihood region $\mathcal{R}_P(c)$ is an approximate confidence region for $\boldsymbol{\psi}$, so it is called a likelihood-confidence region, or a likelihood-confidence interval when $\boldsymbol{\psi}$ is a scalar. Under the null hypothesis $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ the likelihood ratio statistic $-2\ln[R_P(\boldsymbol{\psi}_0; x)]$ usually converge, in distribution, to a chi-squared distribution with d_ψ degrees of freedom (Serfling 1980). When this is true, the set $\mathcal{R}_P(c)$ becomes a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\psi}$, where $c = \exp(-\chi^2_{d_\psi, 1-\alpha}/2)$ and $\chi^2_{d_\psi, 1-\alpha}$ represents the quantile of probability $1 - \alpha$ of a chi-squared distribution with d_ψ degrees of freedom. For example, if $d_\psi = 1$, $\boldsymbol{\psi}$ is a scalar parameter, then the profile likelihood region at level $c = 0.15$ becomes a confidence region for $\boldsymbol{\psi}$, with an approximate 95% confidence level. In a similar way, if $d_\psi = 2$, the level $c = 0.05$ profile likelihood region for $\boldsymbol{\psi}$ will be a confidence region with an approximate 95% confidence level.

Some authors like Montoya, Díaz-Francés & Sprott (2009) and Figueroa (2012) suggest to include the precision of the measuring instrument to avoid unbounded likelihoods, which usually occurs when the continuous approximation to the likelihood function is used and regularity conditions are not satisfied. The unboundedness and also the flatness of a profile likelihood function have been used to propose alternative approaches to estimate nuisance parameters, like in Smith & Naylor (1987) and Green, Roesch, Smith & Strawderman (1994), who criticized the profile likelihood function for being flat and uninformative, overlooking that it can be indicative that a simpler (limiting) model might be a good alternative to explain the data (Cheng & Iles 1990). Although here there is no problem of unbounded likelihoods and to our knowledge, a flat profile likelihood can also be obtained even when including the precision of the measuring instrument, there are some others

circumstances where incorporating this information is reasonable; for example, when the instrument measures with different precision or when it produces many repeated observations, like in the example included in Section 4.

3. Inferences about θ and p Using the Profile Likelihood in the Gumbel Mixture Model

Let X be a random variable from a two-component Gumbel mixture model with density function $f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p)$ given in (1), where $\sigma_1 > 0$, $\sigma_2 > 0$, $0 < p < 1$, $-\infty < \mu_1 < \mu_2 < v$ and v is a known fixed value. In this case, the parameters of interest are the vulnerability parameter $\theta = P(X > v)$ and the mixing proportion p. Although the parametrization of the Gumbel mixture model involves the five unknown parameters μ_1, σ_1, μ_2, σ_2, and p, the parameter θ has been left out. In order to make profile likelihood inferences about θ and p, it is convenient to use a one to one reparametrization in such a way that θ becomes one of the new parameters and p is included as well. Hence, the vulnerability parameter θ can be written, explicitly, as a function of μ_1, σ_1, μ_2, σ_2 and p,

$$
\begin{aligned}
\theta &= P(X > v; \mu_1, \sigma_1, \mu_2, \sigma_2, p) \\
&= 1 - P(X \leq v; \mu_1, \sigma_1, \mu_2, \sigma_2, p) \\
&= 1 - \left[p \Phi_G(\delta_1) + (1-p) \Phi_G(\delta_2) \right]
\end{aligned}
$$

where $\Phi_G(\cdot)$ is the standard Gumbel distribution and $\delta_i = (v - \mu_i)/\sigma_i$, $i = 1, 2$. Here, δ_i is introduced for algebraic and computational simplicity. Note that $\delta_i > 0$ when $\sigma_i > 0$ and $-\infty < \mu_1 < \mu_2 < v$.

3.1. Reparametrizations

Let $\sigma_i = (v - \mu_i)/\delta_i$, $i = 1, 2$. This produces the one to one parametrization $(\mu_1, \sigma_1, \mu_2, \sigma_2, p) \leftrightarrow (\mu_1, \delta_1, \mu_2, \delta_2, p)$ with a Jacobian

$$
J_1 = \frac{(v - \mu_1)(v - \mu_2)}{\sigma_1^2 \sigma_2^2} > 0
$$

The Gumbel mixture model can be reparametrized in terms of $(\mu_1, \delta_1, \mu_2, \delta_2, p)$ when substituting $\sigma_i = (v - \mu_i)/\delta_i$, $i = 1, 2$,

$$
f^*(x; \mu_1, \delta_1, \mu_2, \delta_2, p) = p f_1^*(x; \mu_1, \delta_1) + (1-p) f_2^*(x; \mu_2, \delta_2) \tag{6}
$$

where

$$
f_i^*(x; \mu_i, \delta_i) = \frac{\delta_i}{v - \mu_i} \exp\left[-\delta_i \left(\frac{x - \mu_i}{v - \mu_i} \right) \right] \exp\left\{ -\exp\left[-\delta_i \left(\frac{x - \mu_i}{v - \mu_i} \right) \right] \right\}
$$

with $-\infty < \mu_1 < \mu_2 < v$, $0 < p < 1$ and $\delta_i > 0$, $i = 1, 2$.

The parameter δ_1 can now be written as

$$\delta_1 = \delta_1\left(\theta, p, \delta_2\right) = \Phi_G^{-1}\left[\left(\frac{1-\theta}{p}\right) - \left(\frac{1-p}{p}\right)\Phi_G\left(\delta_2\right)\right] \tag{7}$$

where $\Phi_G^{-1}\left(\cdot\right)$ denotes the inverse of the standard Gumbel distribution. Again, this produces the one to one parametrization $\left(\mu_1, \delta_1, \mu_2, \delta_2, p\right) \leftrightarrow \left(\mu_1, \theta, \mu_2, \delta_2, p\right)$ with a Jacobian given by

$$J_2 = p\exp\left(-\delta_1\right)\Phi_G\left(\delta_1\right) > 0$$

Thus, the Gumbel mixture model (6) can be reparametrized in terms of $\left(\theta, p, \mu_1, \mu_2, \delta_2\right)$ by substituting in (6) the expression $\delta_1\left(\theta, p, \delta_2\right)$ given in (7). Here, θ and p are the parameters of interest and the remaining ones are nuisance parameters.

3.2. Likelihood

In this section a likelihood function for the parameters of the reparametrized mixture model in terms of the parameters of interest θ and p, and the vector of nuisance parameters $\boldsymbol{\lambda} = \left(\mu_1, \mu_2, \delta_2\right)$ is presented. This likelihood includes the precision of the measuring instrument because it could provide valuable information that should be included into the analysis. As Lindsey (1999) explains to include the precision of the measuring instrument into the analysis requires no additional computational effort nowadays.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with density function given in (6) and $\boldsymbol{x}_o = \left(x_1, \ldots, x_n\right)$ its observed sample. Since all measuring instruments have finite precision, that is, data can only be recorded to a finite number of decimals, then $\boldsymbol{x}_o$ must *always* be discrete. Thus the observation $X_i = x_i$ can be interpreted as $x_i - h/2 \leq X_i \leq x_i + h/2$, where h is the precision of the measuring instrument, and so is a fixed positive number, as is described in Sprott (2000, p. 10), Montoya et al. (2009) and Figueroa (2012). Therefore, for $\boldsymbol{x}_o = \left(x_1, \ldots, x_n\right)$, the resulting likelihood function of $\left(\theta, p, \boldsymbol{\lambda}\right)$ is proportional to the probability of the observed sample,

$$
\begin{aligned}
L\left(\theta, p, \boldsymbol{\lambda}; \boldsymbol{x}_o\right) \quad &\propto \quad \prod_{i=1}^{n}\int_{x_i-h/2}^{x_i+h/2} f^*\left[x_i; \mu_1, \delta_1\left(\theta, p, \delta_2\right), \mu_2, \delta_2, p\right] \\
&= \quad \prod_{i=1}^{n}\left\{p\left[F_1\left(x_i + \frac{h}{2}; \theta, p, \mu_1, \delta_2\right) - F_1\left(x_i - \frac{h}{2}; \theta, p, \mu_1, \delta_2\right)\right] + \right.\\
&\qquad\quad \left. \left(1-p\right)\left[F_2\left(x_i + \frac{h}{2}; \mu_2, \delta_2\right) - F_2\left(x_i - \frac{h}{2}; \mu_2, \delta_2\right)\right]\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
F_1\left(z; \theta, p, \mu_1, \delta_2\right) \quad &= \quad \exp\left\{-\exp\left[-\delta_1\left(\theta, p, \delta_2\right)\left(\frac{z-\mu_1}{v-\mu_1}\right)\right]\right\} \\
&= \quad \Phi_G\left\{\Phi_G^{-1}\left[\left(\frac{1-\theta}{p}\right) - \left(\frac{1-p}{p}\right)\Phi_G\left(\delta_2\right)\right]\left(\frac{z-\mu_1}{v-\mu_1}\right)\right\}
\end{aligned}
$$

and

$$F_2(z; \mu_2, \delta_2) \quad = \quad \exp\left\{ -\exp\left[-\delta_2\left(\frac{z - \mu_2}{v - \mu_2} \right) \right] \right\} = \Phi_G\left[\delta_2\left(\frac{z - \mu_2}{v - \mu_2} \right) \right]$$

with $0 < \theta < 1$, $0 < p < 1$, $\delta_2 > 0$, $0 \le [(1 - \theta)/p] - [(1 - p)/p]\,\Phi_G(\delta_2) \le 1$, $-\infty < \mu_1 < \mu_2 < v$ and v is a known fixed value.

Since the mle $(\widehat{\theta}, \widehat{p}, \widehat{\boldsymbol{\lambda}})$ cannot be obtained analytically, it must be calculated numerically. For computational convenience, maximization of the log-likelihood function of the original parametrization can be used to obtain $(\widehat{\mu}_1, \widehat{\sigma}_1, \widehat{\mu}_2, \widehat{\sigma}_2, \widehat{p})$ the maximum likelihood estimators of $(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$, and by the invariance property of the likelihood function, the mle of θ is

$$\widehat{\theta} \quad = \quad 1 - \left[\widehat{p}\Phi_G(\widehat{\delta}_1) + (1 - \widehat{p})\Phi_G(\widehat{\delta}_2) \right]$$

where $\widehat{\delta}_i = (v - \widehat{\mu}_i)/\widehat{\sigma}_i$.

3.3. Profile likelihood

Profile likelihood inference about the vector (θ, p) in the presence of the vector of nuisance parameters $\boldsymbol{\lambda} = (\mu_1, \mu_2, \delta_2)$ is based on the relative profile likelihood function of (θ, p). The profile likelihood function of (θ, p) is

$$L_P(\theta, p; \boldsymbol{x}_o) = \sup_{\lambda} L(\theta, p, \boldsymbol{\lambda}; x_o)$$

and its associated relative profile likelihood function is

$$R_P(\theta, p; \boldsymbol{x}_o) = \frac{L_P(\theta, p; \boldsymbol{x}_o)}{\sup\limits_{\theta, p} L_P(\theta, p; \boldsymbol{x}_o)} = \frac{L_P(\theta, p; \boldsymbol{x}_o)}{L(\widehat{\theta}, \widehat{p}, \widehat{\boldsymbol{\lambda}}; \boldsymbol{x}_o)} \tag{8}$$

where $(\widehat{\theta}, \widehat{p}, \widehat{\boldsymbol{\lambda}})$ is the mle of (θ, p, λ). Note that $R_P(\theta, p; \boldsymbol{x}_o)$ will be a surface sitting above the parameter space $(0, 1) \times (0, 1)$ and its maximum value 1 occurs at $(\theta, p) = (\widehat{\theta}, \widehat{p})$. A convenient way to visualize $R_P(\theta, p; \boldsymbol{x}_o)$ in two dimensions is by plotting contours of level c, obtained by solving $R_P(\theta, p; \boldsymbol{x}_o) = c$. The regions obtained from this contour plot are likelihood-based confidence regions for (θ, p). For instance, if $c = 0.05$ the region delimited by the contour plot is a likelihood-based confidence region for (θ, p), with an approximate 95% confidence level.

On the other hand, profile likelihood inferences about the scalar parameter θ are based on the relative profile likelihood function of θ. In this case the profile likelihood and relative profile likelihood functions are

$$
\begin{aligned}
L_P(\theta; \boldsymbol{x}_o) \quad &= \quad \sup_{p, \lambda} L(\theta, p, \boldsymbol{\lambda}; \boldsymbol{x}_o) = \sup_{p} L_P(\theta, p; \boldsymbol{x}_o) \\
R_P(\theta; \boldsymbol{x}_o) \quad &= \quad \frac{L_P(\theta; \boldsymbol{x}_o)}{\sup\limits_{\theta} L_P(\theta; \boldsymbol{x}_o)} = \frac{\sup\limits_{p} L_P(\theta, p; \boldsymbol{x}_o)}{\sup\limits_{\theta, p} L_P(\theta, p; \boldsymbol{x}_o)}
\end{aligned}
\tag{9}
$$

Similarly, the profile likelihood and the relative profile likelihood for parameter p are given by

$$L_P(p; \boldsymbol{x}_o) = \sup_{\theta, \lambda} L(\theta, p, \lambda; \boldsymbol{x}_o) = \sup_{\theta} L(\theta, p; \boldsymbol{x}_o)$$

$$R_P(p; \boldsymbol{x}_o) = \frac{L_P(p; \boldsymbol{x}_o)}{\sup_p L_P(p; \boldsymbol{x}_o)} = \frac{\sup_\theta L(\theta, p; \boldsymbol{x}_o)}{\sup_{\theta, p} L_P(\theta, p; \boldsymbol{x}_o)} \tag{10}$$

3.4. Computational implementation

Relative profile likelihoods given in (8), (9) and (10) can be obtained through the computation of $L_P(\theta, p; \boldsymbol{x}_o)$, which can be easily implemented using the R 'stats' package function constrOptim, for each specified value of (θ, p). A feasible region is defined for $AB - C \geq 0$, where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix}, \ B = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \delta_2 \end{bmatrix}, \ \text{and } C = \begin{bmatrix} 0 \\ -v \\ 0 \\ \Phi_G^{-1}\left(\dfrac{1-\theta-p}{1-p}\right) \\ -K \end{bmatrix}$$

K is an upper bound for user-defined δ_2, serving as a tuning value. Based on that, the following algorithm is used to display contour lines and profiles of $L_P(\theta, p; x_o)$.

1. Create a full grid from two monotonically increasing grid vectors: $\theta \in S_1 \subset (0, 1)$ and $p \in S_2 \subset (0, 1)$.

2. Map these points to the function $L_P(\theta, p; \boldsymbol{x}_o)$ and store the values in a matrix $M[\theta, p]$.

3. Calculate the matrix $R[\theta, p] = M[\theta, p]/L^*$, the vector $R[\theta] = \max_p M[\theta, p]/L^*$ and the vector $R[p] = \max_\theta M[\theta, p]/L^*$, where $L^* = \max_{\theta, p} M[\theta, p]$.

4. Contour plots can be created from θ, p, $R[\theta, p]$ coordinates using the contour function included in the R 'graphics' package. The profile plot can be obtained by plotting θ versus $R[\theta]$ (or p versus $R[p]$).

Note that $R[\theta]$ and $R[p]$ can be used to obtain profile likelihood intervals for θ and p, respectively.

4. Illustrative Example

Quality of steel is strongly influenced by the presence of non-metallic inclusions. Although inferences about the size of large inclusions have largely been

based on the assumption that inclusions are all of a single type and methods of classic extreme value statistics are appropriate, as shown in Murakami (1994), the need to analyze for the presence of multiple inclusions has been noticed by Lund, Johansson & Olund (1998) through experiments in the bearing industry. In particular, Beretta & Murakami (2001) have found that a Gumbel mixture model can be appropriate when studying only two types of inclusions.

The methodology proposed in this paper is implemented in a set of data obtained from an experiment conducted at the Institute of Metals and Technology, Ljubljana, Slovenia and concerns modeling the size distribution of non-metallic inclusions in steel, where mainly two types of inclusions are investigated: (a) Type 1, these are soft elongated inclusions composed mainly of manganese sulfide and (b) Type 2, hard round inclusions comprising mainly aluminum oxides. These two types of inclusions can be distinguished by their shape when seen under a microscope, and their composition has been confirmed by spectroscopic analyses. Round inclusions are much more harmful and sometimes can lead to premature failure of the steel piece.

As a part of this experiment, a standard inspection area S_0 of 0.27 mm^2 is defined. The area of the maximum inclusion in S_0, defined as $A_{\max}$, is measured in $n = 544$ sample areas, from a single steel slab. All the inclusions were measured using automatic image analysis and only those with a cross-section area larger than 3 μm^2 were considered real inclusions in this analysis. Cross-section areas smaller than 3 μm^2 are not clearly visible by light microscopy at 100x magnification; it is not only the matter of sufficient resolution, but it is difficult to verify what may be inclusions and what could be artifacts from image contrast adjustments. It is worthwhile to mention that none of the $A_{\max}$ measurements was smaller than 3 μm^2.

The square root of the measured area $x = \sqrt{A_{\max}}$ is taken; this is called the inclusion size and these are the measurements used in the statistical analysis. The minimum observation is $x_{(1)} = 1.9339$, the maximum $x_{(544)} = 20.8835$, the sample mean $\bar{x} = 7.3184$, the sample standard deviation $s_x = 2.2158$, and quantiles 25, 50 and 75 are $Q_{25} = 5.8583$, $Q_{50} = 7.1091$, $Q_{0.75} = 8.5621$, respectively. The data set formed with these 544 inclusion sizes has many repeated values. This suggests that the precision of the measuring instrument should be included into the analysis. Actually, this set has only 185 different values, some of them repeated even eight times. Now, since the size of each inclusion is not directly measured by the instrument, the precision associated with each of these values must be computed through error propagation techniques. This implies that each measurement has an associated precision h_i, that should be considered in the analysis.

A Gumbel mixture model is proposed to study the size of non-metallic inclusions in steel and the adequacy of this model can be seen in Figures 1 and 2. Although an observation in Figure 2 is apparently an outlier, it turns out to be a possible observation, when the cloud formed by the Q-Q plots of many simulated samples with the estimated model includes the points of the Q-Q plot of the observed sample; this cloud can be seen in Figure 3. The parameter estimates

for the Gumbel mixture model given in (1) were obtained by maximum likelihood estimation and are shown in Table 1.
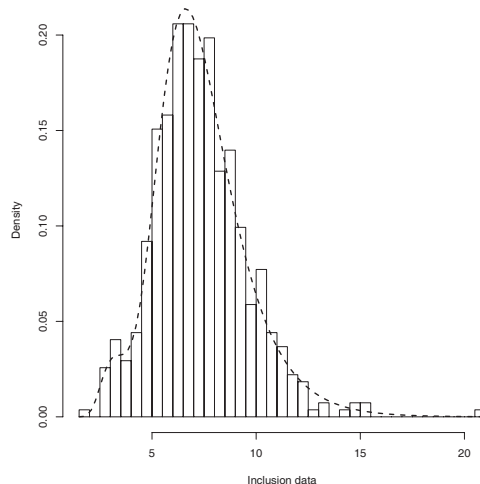

FIGURE 1: Histogram and estimated density for non-metallic inclusion sizes.

$\widehat{\mu}_1 = 3.1943$	$\widehat{\sigma}_1 = 0.7195$	$\widehat{\mu}_2 = 6.6251$	$\widehat{\sigma}_2 = 1.6229$	$\widehat{p} = 0.0597$

TABLE 1: Parameter estimates
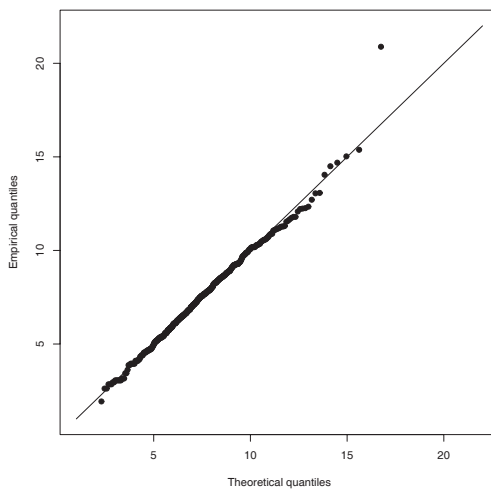
FIGURE 2: Q-Q plot: Theoretical versus Empirical quantiles.

The likelihood approach described in Section 3.2 was used to estimate the probability that the maximum size of a non-metallic inclusion could be greater than a maximum allowed inclusion size v. The problem about fixing a critical
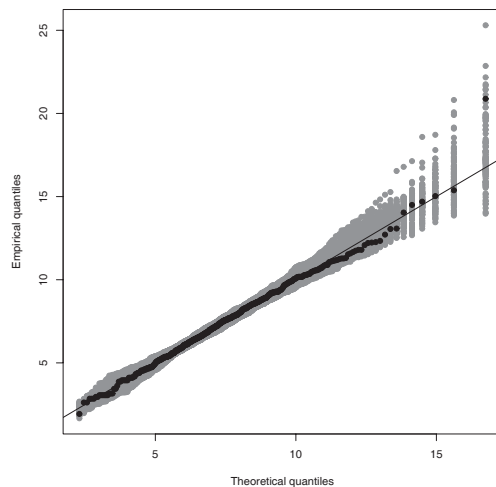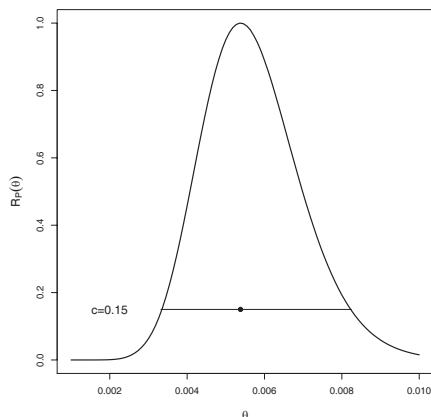
FIGURE 3: Q-Q plot simulation.

size v is that it may differ for different kind of steels, depending on their use or purpose. Here, we fixed $v = 15$ μm, just to show how to model vulnerability.

FIGURE 4: Relative profile likelihood of parameter θ.

The profile likelihood function of parameter $\theta = P(X > 15$ μm$)$ is plotted in Figure 4, and is almost symmetric around its mle $\widehat{\theta} = 0.005380$ which is marked with a dark circle; the 15% profile likelihood interval for θ is $(0.003364, 0.008182)$. It is worthwhile to mention that this information can be used for quality control, because better steel grades can be obtained by reducing its inclusion content. For example, for a plausible value of θ, like $\widehat{\theta} = 0.005380$, a non-metallic inclusion larger than 15, a value merely illustrative, could be associated with a return period of approximately 200 sample areas. This could be low or high depending of the steel and its use. Actually, the information contained in the profile likelihood function of parameter θ is very useful when comparing candidates for the improved steel

grade, since the amount and size of non-metallic inclusions in steel are directly linked with many of its properties.

Besides the importance of having information about parameter θ, it is very informative to characterize parameter p. We are considering two types of non-metallic inclusions and knowledge about the proportion of each of these types is crucial. Here, p denotes the proportion of Type 1 inclusions within the mixture model. Figure 5(a) shows the relative profile likelihood of parameter p, where $\widehat{p} = 0.0597$ is marked with a dark circle and the 15% profile likelihood interval for p results $(0.0294, 0.3422)$. This interval is wide with respect to 1, the length of its parameter space and it does not contain the value $p = 0$. The proportion of Type 1 inclusions can be considered small or large depending on the application of the steel. By the invariance property of the likelihood function, point and interval estimates for Type 2 inclusions can be easily obtained and these are $1 - \widehat{p} = 0.9403$ and $(0.6578, 0.9706)$, respectively. The relative profile likelihood for parameter $1 - p$ is shown in Figure 5(b). Plots in Figure 5 show that these likelihoods are strongly asymmetric around their maximum. It is important to mention that Type 2 inclusions are much more harmful since they do not work when the steel is deformed, so they serve as a crack nucleation sites and can lead to premature failure of the steel piece.
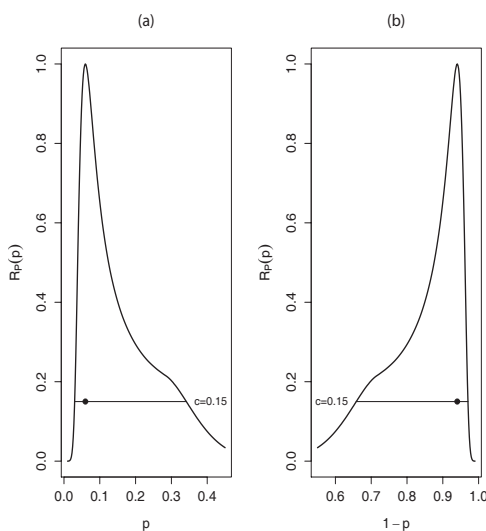

FIGURE 5: Relative profile likelihood of parameter (a) p and (b) $1 - p$.

Inferences about parameters θ and p can be obtained by constructing a likelihood contour plot for these parameters. Figure 6(a) shows the likelihood confidence regions for parameters θ and p at different levels of c. Using the invariance property of the likelihood function, a contour plot for θ and $1 - p$ is easily obtained; this is shown in Figure 6(b). This kind of plot allows us to make simultaneous inferences about the proportion of Type 1 or Type 2 inclusions and the probability of exceeding the maximum allowed inclusion size v. These plots can play an important role in the improvement of the steelmaking practices, for example, they

can be used to compare candidates for the improved steel grade through the analysis of parameters θ and p. We consider that the approach used in this statistical analysis adds another dimension to the overall characterization of the steel grade.
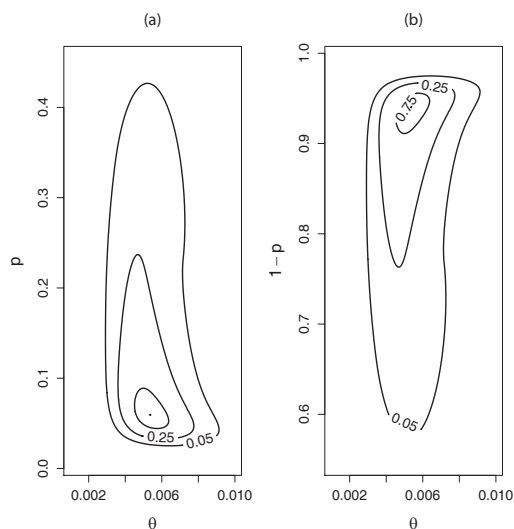

FIGURE 6: Contour plot for parameters (a) θ and p, (b) θ and $1 - p$.

All the parameter inferences were obtained through computational techniques using the R software and the R source code for this example is available upon request.

5. Conclusions

The precision of a measuring instrument turns out to be an important aspect in some applications, like the one included here, where the lack of precision of the measuring instrument caused many repeated observations. The likelihood function allows to include, in a natural way, the precision associated with each of the observations. The profile likelihood function is a simple method devised to handle the estimation of parameters of interest in the presence of unknown nuisance parameters, and it inherits all the information contained in the likelihood function. Estimation statements about parameters $\theta = P(X > v)$, p or (θ, p) in the Gumbel mixture case can be given in terms of profile likelihood confidence regions that were easily obtained through computational techniques. This approach was particularly useful when analyzing the vulnerability of steel.

References

Ahmad, K. E., Jaheen, Z. F. & Modhesh, A. A. (2010), 'Estimation of a discriminant function based on small sample size from a mixture of two Gumbel distributions', *Communications in Statistics-Simulation and Computation* **39**, 713–725.

Barrera-Núñez, V., Meléndez-Frigola, J. & Herraiz-Jaramillo, S. (2008), A survey on voltage sag events in power systems, *in* 'Transmission and Distribution Conference and Exposition: Latin America, 2008 IEEE/PES', pp. 1–3.

Beretta, S. & Murakami, Y. (2001), 'Largest-extreme-value distribution analysis of multiple inclusion types in determining steel cleanliness', *Metallurgical and Materials Transactions B* **32**, 517–523.

Cheng, R. C. H. & Iles, T. C. (1990), 'Embedded models in three-parameter distributions and their estimation', *Journal of the Royal Statistical Society. Series B* **52**(1), 135–149.

Evans, M., Hastings, N. & Peacock, B. (1993), *Statistical Distributions*, John Wiley & Sons, New York.

Figueroa, P. G. (2012), Las funciones de verosimilitud discretizada y restringida perfil en la inferencia científica, Ph.D. Thesis, Universidad de Sonora, Hermosillo, Sonora, México.

Green, E. J., Roesch, F. A. J., Smith, A. F. M. & Strawderman, W. E. (1994), 'Bayesian estimation for the three-parameter Weibull distribution with tree diameter data', *Biometrics* **50**(1), 254–269.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 2, John Wiley & Sons, New York.

Kalbfleisch, J. G. (1985), *Probability and Statistical Inference*, Vol. 2, Springer-Verlag, New York.

Kotz, S. & Nadarajah, S. (2000), *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London.

Lindsay, B. (1995), *Mixture Models: Theory, Geometry and Applications*, Institute for Mathematical Statistics, Hayward.

Lindsey, J. K. (1999), 'Some statistical heresies', *The Statistician* **48**(1), 1–40.

Lund, T., Johansson, S. & Olund, L. (1998), Nucleation of fatigue in very low oxygen bearing steels, *in* 'Bearing Steels: Into the 21st Century, STP 1327', American Society for Testing and Materials, West Conshohocken, pp. 124–130.

Montoya, J. A., Díaz-Francés, E. & Sprott, D. A. (2009), 'On a criticism of the profile likelihood function', *Statistical Papers* **50**, 195–202.

Murakami, Y. (1994), 'Inclusion rating by statistics of extreme values and its application to fatigue strength prediction and quality control of materials', *Journal of Research of the National Institute of Standards and Technology* **99**, 345–351.

Murray, A. T. & Grubesic, T. H. (2007), *Critical Infrastructure: Reliability and Vulnerability*, Springer-Verlag, Berlin.

Raynal, J. & Guevara, J. (1997), 'Maximum likelihood estimators for the two populations Gumbel distribution', *Hydrological Science and Technology Journal* **13**, 47–56.

Rosas-Casals, M., Valverde, S. & Solé, R. V. (2007), 'Topological vulnerability of the European power grid under errors and attacks', *International Journal of Bifurcation and Chaos* **17**(7), 2465–2475.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

Smith, R. L. & Naylor, J. C. (1987), 'Statistics of the three-parameter Weibull distribution', *Annals of Operations Research* **9**, 577–587.

Sprott, D. A. (1980), 'Maximum likelihood and small samples: Estimation in the presence of nuisance parameters', *Biometrika* **67**, 515–523.

Sprott, D. A. (2000), *Statistical Inference in Science*, Springer-Verlag, New York.

Tartaglia, V., Caporali, E., Cavigli, E. & Moro, A. (2005), 'L-moments based assessment of a mixture model for frequency analysis of rainfall extremes', *Advances in Geosciences* **2**, 331–334.

Titterington, D., Smith, A. & Makov, U. (1985), *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York.

Zheng, H. (2007), Investigation of power system blackouts and reliability improvement for power distribution systems, Ph.D. Thesis, The University of Texas, Arlington.