JIRSS (2008)
Vol. 7, Nos. 1-2, pp 73-84

On Identifiablity in Weighted Distributions Using Generalized Maximum Likelihood Estimation

Mohammad Reza Zadkarami

Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University, Iran. (zadkarami_m@scu.ac.ir)

Abstract. In this research, the generalized maximum likelihood estimator (GMLE) is used to investigate the parameters estimation for weighted distributions. There exist situations where the random sample from the population of interest is not available due to the data having unequal probabilities of entering the sample. The method of weighted distributions models the certainty of the probabilities of the events as observed and recorded. It is shown that if the mechanism of sample selection is known up to one unknown parameter, the maximum likelihood estimator (MLE) would be unidentifiable when the conjugate weight function is used. This problem is solved by addition of a prior distribution on model parameters yielding the GMLEs which are identifiable. We also propose the GMLEs for negative exponential, normal and Poisson weighted distributions when MLEs are unidentifiable.

Key words and phrases: conjugate weight, generalized maximum likelihood, identifiablity, maximum likelihood, sample selection, weighted distribution

1 Introduction

Consider a random sample from a distribution function, $F_\theta, \theta \in \Theta$, consistent MLEs of real-valued function of that exist in most of the situations we shall encounter (Lehmann and Casella, 1998). However, there is an important exception when the parameter is not identifiable. It is necessary to establish that the parameters are identifiable before discussing estimation of the parameters. Understanding the sources of identification is essential to determining under what conditions parameters can be recovered and which hypotheses can be tested without imposing arbitrary distributional or functional form assumptions on estimating equations. The parameters identifiablity were investigated for various models and distributions, mixing distributions (Heckman and Singer, 1984, 1985); single-spell duration distribution (Ridder, 1984); Bayesian models (Omlin and Reichert, 1999); competing risk models (Kalbfleisch and Prentice, 2002) and measurement error models (Fuller, 2006).

It is claimed that, the weighted distributions occur naturally in a wide variety of settings with applications in reliability, forestry, ecology, bio-medicine, and many other areas, (Patil, 2002; Oluyede, 2002). In fact, the weighted distribution arises when the observations do not have equal chance of being recorded, due to particular experimenter design or the consequence of behavioral observations. The concept of a weighted distribution can be traced to Fisher (1934) although these models were first formulated in a unified way by Rao (1965). Good surveys on this topic are Rao (1985) and Patil (2002).

If the mechanism of sample selection is completely known, the parameter estimators can be obtained by ML method if the regular conditions are satisfied (Lehmann and Casella, 1998). However, in practice, the functional form of sample selection is known up to unknown parameter. In this case, it is possible that the problem of identifiablity of the parameters occurs. In this regard, Gilbert et al. (1999) investigated parameter identifiablity in the weighted logistic model. But, some authors investigated various models to obtain consistent parameters estimators (Sun and Woodroofe, 1997; Silliman, 1997; Bayarri and Berger, 1998; Cristobal and Alcala, 2001; Ma et al., 2005).

The objective of the present study is to investigate the GMLEs in the weighted distributions when the MLEs are unidentifiable. In fact, when the statistical method is incapable of handling MLE, the GMLE may solve the problem (Berger, 1985; Martins and Stedinger, 2001; Adlouni et al., 2007).

The rest of the paper is organized as follows. In Section 2, we discuss identifiability issues for weighted distributions when the conjugate weight function is used. In section 3, we introduce the GMLE procedure and discuss the conditions on the weight functions to find unique parameter estimations. Conclusion is presented in section 4.

2 Identifiability and conjugate weight function

Let the original observation Y have probability density function (pdf) $f(y|\theta), \theta \in \Theta$ where it is interesting to perform statistical inference about θ. Let $\omega(y|\tau)$ be the probability of recording the observation y, then the pdf of the recorded observation, Y^w, is

$$f^w(y|\theta,\tau) = \frac{w(y|\tau)f(y|\theta)}{W(\theta,\tau)} \tag{2.1}$$

where $W(\theta,\tau) = E_\theta(w(y|\tau)) = \int w(y|\tau)f(y|\theta)dy$ is just a normalizing constant, and the weight function $w(y|\tau)$ may depend on some perhaps unknown parameter τ. In general, the weight function, $w(y|\tau)$, is a non-negative function with the parameter τ representing the recording mechanism. Clearly, the recorded y is not observation on Y, but on the random variable Y^w. The random variable Y^w is called the weighted version of Y. Note that the weight function $w(y|\tau)$ need not be between zero and one and actually it may exceed unity, as, for example when $w(y|\tau) = y$, in which case Y^w is called the size-biased version of Y.

Definition 2.1. Let Y be distributed according to $f(y|\theta,\tau)$. If there exist pairs (θ_1,τ_1) and (θ_2,τ_2) with $\theta_1 \neq \theta_2$ for which $f(y|\theta_1,\tau_1) = f(y|\theta_2,\tau_2)$, the parameter θ is said to be unidentifiable.

Definition 2.2. Let F denote the class of density functions $f(y|\theta)$ (indexed by θ). A class Π of weight functions is said to be conjugate for F if $f^w(y|\theta,\tau)$ is in the class F. In other words, there exists some real function $g(.)$ such that $f^w(y|\theta,\tau) = f(y|\theta^*)$ where $\theta^* = g(\theta,\tau)$, $\theta^* \in \Theta$. The function $w(y|\tau)$ is called conjugate weight function.

Theorem 2.1. *Let $Y_1, Y_2, ..., Y_n$ be independent identical distributed (i.i.d) according to a distribution $f(y|\theta)$, $\theta \in \Theta$ where $\theta' =$*

$(\theta_1, \theta_2, ..., \theta_k)$, a vector of unknown parameters, and $Y_1^w, Y_2^w, ..., Y_n^w$ be the weighted version of Y's with weight function $w(y|\tau)$. The parameters of the weighted distribution are not identifiable if $w(y|\tau)$ is a conjugate weight function.

Proof. Suppose $w(y|\tau)$ is a conjugate weight function. By definition, there is a function $g(\theta,\tau) = (g_1(\theta,\tau), g_2(\theta,\tau), ..., g_r(\theta,\tau))$, $1 \leq r \leq k$ such that $\theta^* = g(\theta,\tau)$ and $f^w(y|\theta,\tau) = f(y|\theta^*)$, $\theta^* \in \Theta$. If $\hat{\theta^*}(y_1, y_2, ..., y_n)$ is the MLE of θ^*, the MLE of θ and τ are obtained by solving the equation $\hat{\theta_i^*}(y_1, y_2, ..., y_n) = \hat{g}_i(\theta,\tau) = g_i(\hat{\theta},\hat{\tau})$ where $i = 1, 2, ..., r$. But both θ and τ cannot be estimated separately because they are not uniquely defined. Hence, the parameters θ and τ are unidentifiable.

Example 2.1. Let $Y_1, Y_2, ..., Y_n$ be a random sample from the negative exponential distribution with parameter θ and the weight function, $w(y|\tau) = exp(\tau y)$ $\tau > 0$. Then $f^w(y|\theta,\tau) = (\theta - \tau)exp[-(\theta - \tau)y] = \theta^* exp(-\theta^* y) = f(y|\theta^*)$, $\theta^* = (\theta - \tau) > 0$ and $w(y|\tau) = exp(\tau y)$ is a conjugate weight function. The MLE of θ^* is $\hat{\theta^*}(y_1, y_2, ..., y_n) = \frac{n}{\sum_{i=1}^n y_i} = (\hat{\theta} - \hat{\tau})$ and θ and τ are unidentifiable. However, for the weight function $w(y|\tau) = y^\tau$, $\tau > 0$, the weighted density function is given by $f^w(y|\theta,\tau) = \frac{\theta^{\tau+1}}{\Gamma(\tau+1)} y^\tau exp(-\theta y)$ and $w(y|\tau) = y^\tau$ is not a conjugate weight function. The likelihood function is

$$\begin{aligned} L(\theta,\tau) &= \prod_{i=1}^n f^w(y_i|\theta,\tau) = \frac{\theta^{n(\tau+1)}}{[\Gamma(\tau+1)]^n} (\Pi_{i=1}^n y_i)^\tau exp(-\theta \sum_{i=1}^n y_i) \\ \ell(\theta,\tau) &= lnL(\theta,\tau) \\ &= n(\tau+1)ln(\theta) - nln(\Gamma(\tau+1)) + \tau \sum_{i=1}^n ln(y_i) - \theta \sum_{i=1}^n y_i \end{aligned}$$

Hence, the MLE of θ and τ are obtained by solving the equations $\frac{\partial \ell(\theta,\tau)}{\partial \theta} = 0$ and $\frac{\partial \ell(\theta,\tau)}{\partial \tau} = 0$. In this case, the unique MLE of θ and τ exist, Lehmann and Casella (1998) and θ and τ are identifiable. However, if the density function of Y is $f(y|\theta,\tau) = \frac{\theta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} exp(-\theta y)$ where the weight function is $w(y|\tau) = y^\tau$, $\tau > 0$, then $f^w(y|\alpha,\theta,\tau) = \frac{\theta^{\tau+\alpha}}{\Gamma(\tau+\alpha)} y^{\tau+\alpha-1} exp(-\theta y)$ and θ, α and τ are not identifiable, because the weight function is conjugate. Therefore, the identifiability of parameters depends on the form of both density function and weight function.

Theorem 2.2. *Consider two functions, $W_1(\tau)$ and $W_2(\theta)$ such that $W(\theta,\tau) = E_\theta(w(y|\tau)) = W_1(\tau)W_2(\theta)$. Then the parameters of the weighted distribution are identifiable.*

Proof. Suppose that the expectation of weight function factors into the product of a function of τ alone and a function of θ alone, so that we can write the weighted density function as follows:

$$f^w(y|\theta,\tau) = \frac{w(y|\tau)f(y|\theta)}{W_1(\tau)W_2(\theta)}.$$

The likelihood function is given by

$$L(\theta,\tau) = \prod_{i=1}^{n} f^w(y_i|\theta,\tau) = \frac{\prod_{i=1}^{n} w(y_i|\tau))}{[W_1(\tau)]^n}\frac{\prod_{i=1}^{n} f(y_i|\theta)}{[W_2(\theta)]^n},$$

the log-likelihood function is

$$\begin{aligned}\ell(\theta,\tau) &= [\sum_{i=1}^{n} ln(w(y_i|\tau)) - nln(W_1(\tau))] \\ &\quad +[\sum_{i=1}^{n} ln(f(y_i|\theta)) - nln(W_2(\theta))] \\ &= \ell_1(\tau) + \ell_2(\theta)\end{aligned}$$

If the regular conditions are satisfied, the unique and consistence MLE of θ and τ are obtained by solving the following equations: $\frac{\partial\ell(\theta,\tau)}{\partial\tau} = \frac{\partial\ell_1(\tau)}{\partial\tau} = 0$, $\frac{\partial\ell(\theta,\tau)}{\partial\theta} = \frac{\partial\ell_2(\theta)}{\partial\theta} = 0$. This means that the parameters θ and τ are identifiable.

Some authors assumed that the parameter τ is known and weight function $w(y|\tau) = w(y)$ is free from unknown parameters (Patil, 2002). But, the asymptotic variance of an efficient estimator of θ when τ is unknown can never fall bellow its value when is known unless

$$Cov(\frac{\partial \ln f^w(y|\theta,\tau)}{\partial\theta}, \frac{\partial \ln f^w(y|\theta,\tau)}{\partial\tau}) = 0 \qquad (2.2)$$

This condition states that the Hessian matrix should be diagonal. Therefore, if condition (2.2) is not satisfied and wrongly we assume that τ is known, it results in misleading statistical inference about θ (Lemann and Casella, 1998).

The exponential family is widely used in the applications of weighted distributions (Rao, 1985; Patil, 2002). Theorem 2.3 investigates the identifiablity conditions of the family.

Theorem 2.3. *Let $Y_1, Y_2, ..., Y_n$ be i.i.d form multi-parameter exponential family with the density function*

$$f(y|\theta) = a(\theta)b(y)exp[\sum_{j=1}^{J}\theta_j T_j(y)], \theta' = (\theta_1, \theta_2, ..., \theta_J)$$

Let $Y_1^w, Y_2^w, ..., Y_n^w$ be the weighted version of Y's when the weight function is given by $w(y|\theta) = exp[\sum_{j=1}^{J} C_j(\theta,\tau)T_j(y)]$ for some functions $C_j(\theta,\tau)$. Then the weight function is conjugate and the parameters are unidentifiable.

Proof. It is easy to show that $W(\theta,\tau) = E_\theta(w(y|\tau)) = a(\theta)[k(\theta^*)]^{-1}$ where $[k(\theta^*)]^{-1} = \int b(y)exp[\sum_{j=1}^{J}\theta_j^* T_j(y)]dy$, $\theta'^* = (\theta_1^*, \theta_2^*, ..., \theta_J^*)$ and $\hat{\theta_j^*} = \theta_j + C_j(\theta,\tau)$ $j = 1, 2, ..., J$. The weighted density function is $f^w(y|\theta,\tau) = b(y)k(\theta^*)exp[\sum_{j=1}^{J}\theta_j^* T_j(y)]$. Hence, the MLE of θ_j^* exists, unique and is denoted by $\hat{\theta_j^*}(y_1, y_2, ..., y_n)$ $j = 1, 2, ..., J$. The MLE of θ_j's and τ are obtained by solving the equations:

$$\hat{\theta_j^*}(y_1, y_2, ..., y_n) = \theta_j + C_j(\theta,\tau), \quad j = 1, 2, ..., J \tag{2.3}$$

Equations (2.3) do not have unique solution and both θ and τ are unidentifiable.

3 The GMLE for the weighted distributions

When the MLEs are unidentifiable, the extra information about the parameters of the weighted distribution is required to achieve the parameter identifiablity. Let Y^w be the weighted version of Y with density function $f^w(y|\theta,\tau) = \frac{w(y|\tau)f(y|\theta)}{W(\theta,\tau)}$ where $w(y|\tau)$ is the weight function. Once the prior function $h(\theta,\tau)$ is chosen, the posterior function (or the generalized likelihood function) is computed as

$$\begin{aligned} GL(\theta,\tau) &= \Pi(\theta,\tau|y) \\ &= f^w(y|\theta,\tau)h(\theta,\tau)m^{-1}(y) \\ &= L(\theta,\tau|y)h(\theta,\tau)m^{-1}(y) \end{aligned}$$

where $m(y) = \int\int f^w(y|\theta,\tau)h(\theta,\tau)d\tau d\theta$. Thus $ln[GL(\theta,\tau)]$ equals $ln[L(\theta,\tau)]$ plus $ln[h(\theta,\tau)] - ln[m(y)]$. This extra information helps to identify the parameters. The GMLE of θ and τ can be obtained by maximizing the generalized log-likelihood function, which corresponds to mode of the Bayesian posterior distribution of the parameters (Berger, 1985).

Theorem 3.1. *Let $f^w(y|\theta,\tau) = f(y|\theta^*)$, $\theta^* = g(\theta,\tau)$ be the weighted version of the density function $f(y|\theta)$ with a conjugate weight function $w(y|\tau)$. The parameters of weighted distribution are identified if the prior function $h(\theta,\tau)$ factors as $h(\theta,\tau) = h_0(\theta^*,\tau^*)h_1(\theta,\tau)$ where the function $h_1(\theta,\tau)$ is non-constant function of θ and τ, does not involve θ^* and τ^*, and the function $h_0(\theta^*,\tau^*)$ is nonnegative which depends only on θ^* and τ^*.*

Proof. The generalized likelihood function is given by

$$\begin{aligned} GL(\theta,\tau) &= f^w(y|\theta,\tau)h(\theta,\tau)m^{-1}(y) \\ &= f(y|\theta^*,\tau^*)h_0(\theta^*,\tau^*)h_1(\theta,\tau)m^{-1}(y) \\ ln[GL(\theta,\tau)] &= ln[f(y|\theta^*,\tau^*)h_0(\theta^*,\tau^*)] + ln[h_1(\theta,\tau)] - ln[m(y)] \end{aligned}$$

where $m(y) = \int\int f^w(y|\theta,\tau)h(\theta,\tau)d\tau d\theta$. The unique and consistence GMLEs of θ and τ are obtained by solving the equations

$$\frac{\partial \ln[GL(\theta,\tau)]}{\partial\theta} = \frac{\partial \ln[f(y|\theta^*,\tau^*)h_0(\theta^*,\tau^*)]}{\partial\theta} + \frac{\partial \ln[h_1(\theta,\tau)]}{\partial\theta} = 0 \quad (3.1)$$

$$\frac{\partial \ln[GL(\theta,\tau)]}{\partial\tau} = \frac{\partial \ln[f(y|\theta^*,\tau^*)h_0(\theta^*,\tau^*)]}{\partial\tau} + \frac{\partial \ln[h_1(\theta,\tau)]}{\partial\tau} = 0 \quad (3.2)$$

when the regular conditions are satisfied. The extra information about θ and τ in the second statements of equations (3.1) and (3.2) provides the unique solution for θ and τ which has resulted in the identifiabilty of θ and τ.

In general, GMLEs will have the desired asymptotic optimal properties if both the likelihood and the prior satisfy a few regularity conditions, and they may even have these properties in other cases (Lemann and Casella, 1998).

3.1 The GMLE in the conjugate weighted for exponential family

Let $Y_1, Y_2, ..., Y_n$ be a random sample form one parameter exponential family with density function, $f(y|\theta) = a(\theta)b(y)exp[C(\theta)T(y)]$. Let $Y_1^*, Y_2^*, ..., Y_n^*$ be the weighted version of Y's with the conjugate weight function, $w(y|\tau,\theta) = exp[Q(\theta,\tau)T(y)]$. The weighted density function is given by $f^w(y|\theta,\tau) = b(y)k(\theta^*)exp[\theta^* T(y)]$, where $\theta^* = Q(\theta,\tau) + C(\theta)$. The generalized likelihood function for prior function $h(\theta,\tau)$ is

$$\begin{aligned} GL(\theta,\tau) &= L(\theta,\tau|y)h(\theta,\tau)m^{-1}(y) = L(\theta^*|y)h(\theta,\tau)m^{-1}(y) \\ ln[GL(\theta,\tau)] &= ln[L(\theta^*|y)] + ln[h(\theta,\tau)] - ln[m(y)] \end{aligned}$$

If $h(\theta,\tau)$ provides extra information about θ and τ, according to theorem 3.1, the unique GMLEs of θ and τ are obtained by solving the equations $\frac{\partial \ln[GL(\theta,\tau)]}{\partial\theta} = 0$ and $\frac{\partial \ln[GL(\theta,\tau)]}{\partial\tau} = 0$.

Example 3.1. Suppose the random variables $Y_1, Y_2, ..., Y_n$ are i.i.d from $f(y|\theta) = \theta e^{-\theta y}$, $\theta > 0$ with the weight function, $w(y|\tau) = exp(\tau y)$, $\tau > 0$. The weighted density function can be written as $f^w(y|\theta,\tau) = f(y|\theta^*) = \theta^* exp(-\theta^* y)$ where $\theta^* = (\theta - \tau)$. Let the prior function be $h(\theta,\tau) = h_1(\theta|\tau)h_2(\tau) = \tau e^{-\theta\tau}e^{-\tau}$, $\theta > 0$ and $\tau > 0$. The generalized likelihood function is

$$GL(\theta,\tau) = (\theta-\tau)^n exp[-(\theta-\tau)T(y)]\tau exp[-(\theta+1)\tau]m^{-1}(y)$$

where $T(y) = \sum_{i=1}^n y_i$ and

$$m(y) = \int\int (\theta-\tau)^n \tau exp[-(\theta-\tau)T(y) - (\theta+1)\tau]d\tau d\theta.$$

$$\begin{aligned} ln[GL(\theta,\tau)] &= nln[(\theta-\tau)] - (\theta-\tau)T(y) + ln[\tau] \\ &\quad -(\theta+1)\tau - ln[m(y)] \\ \frac{\partial \ln[GL(\theta,\tau)]}{\partial\theta} &= \frac{n}{(\theta-\tau)} - T(y) - \tau = 0 \Rightarrow \hat{\theta} = \hat{\tau} + \frac{n}{T(y)+\hat{\tau}} \\ \frac{\partial \ln[GL(\theta,\tau)]}{\partial\tau} &= -\frac{n}{(\theta-\tau)} + T(y) + \tau^{-1} - (\theta+1) = 0 \\ &\Rightarrow 2\tau^3 + [2T(y)+1]\tau^2 + [T(y)+n-1]\tau - T(y) = 0. \end{aligned}$$

The function $g(\tau) = 2\tau^3 + [2T(y)+1]\tau^2 + [T(y)+n-1]\tau - T(y)$ is continuous over the real line, $g(0) = -T(y) < 0$ and $g(1) = 2[T(y)+1] + n > 0$. We also have $g\prime(\tau) = \frac{\partial\, g(\tau)}{\partial\tau} = 6\tau^2 + 2[2T(y)+1]\tau + (T(y)+n-1) > 0$, $\tau > 0$, then $g(\tau)$ is a strictly increasing function over interval $(0,1)$ and consequently the unique solution $\hat{\tau} = (\frac{d}{6} - 6cd^{-1} - \frac{T(y)}{3} - \frac{1}{6}) \in (0,1)$ exists and both parameters θ and τ are identifiable. The calculation of $\hat{\tau}$ is done by option

(Solve, linear) in software Maple 7, where $d = [a + 3b^{0.5}]^{\frac{1}{3}}$, $a = 6[T(y)]^2 + 39[T(y)] + 18n[T(y)] + 9n - 10 - 8[T(y)]^3$, $b = -54[T(y)] + 81[T(y)]^2 + 24n^3 + 10[T(y)]^3 + 78n - 18n[T(y)] - 108[T(y)]^4 - 75n^2 + 288n[T(y)]^2 + 60n^2[T(y)] - 24n[T(y)]^3 - 12n^2[T(y)]^2 - 27$ and $c = \frac{T(y)}{18} + \frac{n}{6} - \frac{7}{36} - \frac{[T(y)]^2}{9}$. Then, $\hat{\theta}$ and $\hat{\tau}$ are the unique GMLE of θ and τ, respectively.

Example 3.2. Consider a random sample, $Y_1, Y_2, ..., Y_n$, of $N(\theta, \sigma^2)$ where $\sigma^2 > 0$ known, with weight function $w(y|\tau) = exp(-\tau y)$. Then $Y_i^w \sim N(\theta^*, \sigma^2)$, $i = 1, 2, ..., n$ where $\theta^* = (\theta - \tau\sigma^2)$ and both parameters θ and τ are unidentifiable. When one chooses the prior, $h(\theta, \tau) = h_1(\theta|\tau)h_2(\tau)$, where $\theta|\tau \sim N(\tau, \delta^2)$ and $\tau \sim N(\eta, \gamma^2)$, then

$$\begin{aligned} GL(\theta, \tau) &= (2\pi\sigma^2)^{-\frac{n}{2}}(2\pi)^{-1}exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\theta - \tau\sigma^2)]^2 \\ &\quad -0.5[(\theta-\tau)^2 + \tau^2]]m^{-1}(y) \end{aligned}$$

where, for simplicity we assume $\delta^2 = \gamma^2 = 1$, $\eta = 0$, and

$$\begin{aligned} m(y) &= \int\int (2\pi\sigma^2)^{-\frac{n}{2}}(2\pi)^{-1}exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\theta - \tau\sigma^2)]^2 \\ &\quad -0.5[(\theta-\tau)^2 + \tau^2]]d\tau d\theta. \end{aligned}$$

Then

$$\frac{\partial \ln[GL(\theta,\tau)]}{\partial\theta} = 0 \quad\Rightarrow\quad \hat{\theta} = \sum_{i=1}^{n} y_i - n(\theta - \tau\sigma^2) - \sigma^2(\theta - \tau) = 0$$

$$\frac{\partial \ln[GL(\theta,\tau)]}{\partial\tau} = 0 \quad\Rightarrow\quad \hat{\theta} = \sum_{i=1}^{n} y_i - n(\theta - \tau\sigma^2) - (2\tau - \theta) = 0$$

and the unique GMLEs of θ and τ are $\hat{\theta} = \frac{\sigma^2+2}{2n-\sigma^2(\sigma^2-1)}\sum_{i=1}^{n} y_i$ and $\hat{\tau} = \frac{(\sigma^2+1)}{(\sigma^2+2)}\hat{\theta} = \frac{(\sigma^2+1)}{2n-\sigma^2(\sigma^2-1)}\sum_{i=1}^{n} y_i$, respectively. Therefore, the extra information about θ and τ, from prior function $h(\theta, \tau)$, results in the identifiability of θ and τ.

Example 3.3. Let random variables $Y_1, Y_2, ..., Y_n$ be i.i.d from $f(y|\theta) = \frac{e^{-\theta}\theta^y}{y!}$, $\theta > 0$ with the weight function, $w(y|\tau) = \tau^y$, $\tau > 0$. The weighted density function can be written as $f^w(y|\theta, \tau) = f(y|\theta^*) = \frac{e^{-\theta^*}(\theta^*)^y}{y!}$ where $\theta^* = \theta\tau$ and both parameters θ and τ are unidentifiable. Now consider the prior function $h(\theta, \tau) = h_1(\theta|\tau)h_2(\tau)$

$= e^{-\theta}e^{-\tau}$, $\theta > 0$ and $\tau > 0$. The generalized likelihood function is given by

$$GL(\theta,\tau) = exp(-n\theta\tau)(\theta\tau)^{T(y)}exp[-(\theta+\tau)]m^{-1}(y)$$

where $T(y) = \sum_{i=1}^{n} y_i$ and

$$m(y) = \int\int(\theta\tau)^{T(y)}exp[-\tau(n\theta+1)-\theta)]d\tau d\theta.$$

$$\begin{aligned}
ln[GL(\theta,\tau)] &= -\tau(n\theta+1)-\theta+T(y)ln[\theta]+T(y)ln[\tau]-ln[m(y)] \\
\frac{\partial \ln[GL(\theta,\tau)]}{\partial\theta} &= -n\tau-1+\frac{T(y)}{\theta}=0 \Rightarrow \hat{\theta}=\frac{T(y)}{n\hat{\tau}+1} \\
\frac{\partial \ln[GL(\theta,\tau)]}{\partial\tau} &= -(n\theta+1)+\frac{T(y)}{\tau}=0 \\
&\Rightarrow g(\tau)=n\tau^2+\tau-T(y)=0.
\end{aligned}$$

The unique positive solution of $g(\tau)$ is $\hat{\tau} = \frac{\sqrt{1+4nT(y)}-1}{2n}$. The unique GMLEs of θ and τ are $\hat{\theta}$ and $\hat{\tau}$, respectively and, both θ and τ are identifiable.

Remark 3.1. The conjugate property is an essential condition for non-identifiably whether the weight function is free from θ or not. Consider negative exponential distribution with weight function $w(y|\tau) = exp(\frac{\tau y}{\theta})$. The weighted density function is given by $f^w(y|\theta,\tau) = (\theta-\frac{\tau}{\theta})exp(-(\theta-\frac{\tau}{\theta})y)$, $\theta^2 \geq \tau$. Hence, the parameters θ and τ are unidentifiable because we have no unique solution for equation $\hat{\theta}^*(y_1,y_2,...,y_n) = \frac{n}{\sum_{i=1}^{n} y_i} = (\theta-\frac{\tau}{\theta})$.

4 Conclusion

Based on the above discussion, it is found that if the conjugate weight function is used, the MLEs of weighted distribution are unidentifiable. However, when the extra information is available, for example the convenient prior function, the identifiablity of parameters is possible through the GMLEs. An advantage of GMLEs is that they can efficiently employ historical information. The cases with historical information are not as easily or as adequately addressed by MLEs (Jin and Stedinger, 1989).

Acknowledgments

The author likes to thank the referees and editor for their many constructive comments, which improved the quality of this paper significantly.

References

Adlouni, S. E., Ouarda, T. B. M. J., Zhang, X. et al. (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resource Res., **43**, WO3410, doi:10.1029/2005WR004545, 2007.

Bayarri, M. J. and Berger, J. (1998), Robust bayesian analysis of selection models. Ann. Stat., **26**, 645-659.

Berger, J. O. (1985), Statistical Decision theory and Bayesian Analysis, 2nd Edn. New York: Springer.

Cristobal, J. A. and Alcala, J. T. (2001), An overview of nonparametric contributions to the problem of functional estimation from biased data. Test, **10**, 309-332.

Fisher, R. A. (1934), The effect of methods of ascertainment upon the estimation of frequencies. Ann. Eng., **6**, 13-25.

Fuller, Wayne, A. (2006), Measurement Error Models. New Jersey: Wiley.

Gilbert, P. B., Lele, S. R. and Vardi, Y. (1999), Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. Biometrika, **86**, 27-43.

Heckman, J. J. and Singer, B. (1984), The identifiablity of the proportional hazard model. Rev. Econ. Study, **51**, 231-241.

Heckman, J. J. and Singer, B. (1985), Longitudinal Data Analysis of Labour Market, Chapter seven. Cambridge University Press.

Kalbfleisch, J. D. and Prentice, R. L. (2002), The Statistical Analysis of Failure Time Data. New York: Wiley.

Jin, M. and Stedinger, J. R. (1989), Flood frequency analysis with regional and historical information. Water Resource Res., **35**, 925-936.

Lehmann, E. L. and Casella, G. (1998), Theory of Point Estimation. 2nd Edn. New York: Chapman and Hall.

Ma, Y., Genton, M. G. and Tsiatis, A. A. (2005), Locally efficient semiparametric estimators for generalized skew-elliptical distributions. J. Am. Stat. Assoc., **100**, 980-989.

Martins, E. S. and Stedinger, J. R. (2000), Generalized maximum likelihood GEV quantile estimators for hydrologic data. Water Resource Res., **36**, 737-744.

Martins, E. S. and Stedinger, J. R. (2001), Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. Water Resource Res., **37**, 2551-2557.

Oluyede, B. O. (2002), Some inequalities and bounds for weighted reliability measures. J. Inequality Pure App. Math., **3**, Article 60.

Omlin, M. and Reichert, P. (1999), A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Mod., **115**, 45-59.

Patil, G. P. (2002), Weighted distributions. In Encyclopedia of Environmetrics, Eds. A.H. EI-Shaarawi and W.W. Piegorsch. Vol. 4, John Wiley & Sons, Ltd. Chichester, pp 2369-2377.

Rao, C. R. (1965), On discrete distributions arising out of methods of ascertainment. In Classical and Contagious Discrete Distributions, Eds. G.P. Patil, Pergamon Press and Statistical Publishing Society. Calcutta, pp 320-332.

Rao, C. R. (1985), Weighted distributions arising out of methods of ascertainment. In A Celebration of Statistics, Eds. A. C. Atknson and S. E. Fienberg. New York: Springer, pp 543-569.

Ridder, J. (1984), The distribution of single-spell duration data. In Studies in Labour Market Dynamics, Eds. G.R. Neumann and Westergaard-Nielson. Berlin: Springer.

Silliman, N. P. (1997), Nonparametric classes of weight functions to model publication bias. Biometrika, **84**, 909-918.

Sun, J. and Woodroofe, M. (1997), Semi-parametric estimates under biased sampling. Stat. Sin., **7**, 545-575.