Advances and Applications in Statistics
Volume 10, Number 1, 2008, Pages 131-147
This paper is available online at http://www.pphmj.com

A PARAMETRIC ESTIMATION OF WEIBULL DENSITY

EZZ H. ABDELFATTAH

Faculty of Science
Helwan University
Cairo, Egypt

Abstract

Abdelfattah et al. [2] had estimated the parameters of the 2-Weibull distribution using Bayesian estimation. Abdelfattah [1] had estimated the Weibull density using goodness of fit tests. The estimation of the Weibull density is now introduced by parametric methods. Some numerical results were obtained through a simulation study to obtain the critical values for some well known statistics, beside the power function for these tests.

1. Introduction

Suppose that $p(x \mid \theta)$ is the distribution of a random sample $\underline{x}$. Suppose further that a random sample of n observations $\underline{x} = (x_1, x_2, \ldots, x_n)$ is available from this distribution. Let y be a future observation coming from the parametric density function $p(y \mid \theta)$ which is unknown. Then, there are two available methods for estimating $p(y \mid \theta)$. The first method is the classical estimative which uses

$$p(y \mid \theta) = p(y \mid \theta = \hat{\theta}), \tag{1.1}$$

where $\hat{\theta}$ is an estimate of θ based on the sample $\underline{x}$.

2000 Mathematics Subject Classification: 46N30.

Keywords and phrases: Bayesian predictive method, Weibull density, simulation, predictive fit.

Received June 26, 2008

The second method is the Bayesian predictive method which uses

$$p(y \mid \underline{x}) = \int_{\Theta} p(y \mid \theta) p(\theta \mid \underline{x}) d\theta, \tag{1.2}$$

where $p(\theta \mid \underline{x})$ is the posterior distribution of θ given $\underline{x}$, based on the prior distribution $p(\theta)$ and the sample $\underline{x}$. Geisser [7] proposed some conditions on $p(y \mid \underline{x})$ before it can be used as a surrogate density for $p(y \mid \theta)$. Aitchison [3] introduced a criterion for evaluating the closeness of the estimative density and the predictive density to the true density of a future observation. Further theoretical justification for the use of $p(y \mid \underline{x})$ was provided by Murray [10, 11] and Ng [12] in terms of invariance. Dunsmore [6] and Amaral and Dunsmore [4] have shown that in the case of large samples the predictive density $p(y \mid \underline{x})$ and the estimative density $p(y \mid \theta = \hat{\theta})$ are approximately related.

2. Aitchison's Criterion

The Aitchison's criterion focuses on the small sample sizes where it uses the Kullback and Leibler [8] measure of the divergence. The divergence of any estimate $q(y \mid \underline{x})$ to the true unknown density $p(y \mid \theta)$ is given by

$$\int_{Y} p(y \mid \theta) \log\left(\frac{p(y \mid \theta)}{q(y \mid \underline{x})}\right) dy > 0 \tag{2.1}$$

unless $q(y \mid \underline{x})$ coincides with $p(y \mid \theta)$.

If we have two estimates for $p(y \mid \theta)$, say $q(y \mid \underline{x})$ and $r(y \mid \underline{x})$, then $q(y \mid \underline{x})$ is closer to $p(y \mid \theta)$ if

$$\int_{Y} p(y \mid \theta) \log\left(\frac{q(y \mid \underline{x})}{r(y \mid \underline{x})}\right) dy > 0 \tag{2.2}$$

and it depends on θ and $\underline{x}$.

The measure is then represented by the expectation of (2.2) with respect to $p(\underline{x} \mid \theta)$ as follows:

$$\int_{\underline{x}} p(\underline{x} \mid \theta) \int_Y p(y \mid \theta) \log\left(\frac{q(y \mid \underline{x})}{r(y \mid \underline{x})}\right) dy d\underline{x}. \tag{2.3}$$

In general the criterion in (2.3) will depend on θ, but in some cases does not.

The expectation of criterion (2.3) over different values of θ with respect to the prior distribution $p(\theta)$ and replacing $q(y \mid \underline{x})$ by $p(y \mid \underline{x})$ is expressed as

$$\int_\Theta p(\theta) \int_{\underline{X}} p(\underline{x} \mid \theta) \int_Y p(y \mid \theta) \log\left(\frac{p(y \mid \underline{x})}{r(y \mid \underline{x})}\right) dy d\underline{x} d\theta. \tag{2.4}$$

Since $p(\theta)p(\underline{x} \mid \theta) = p(\underline{x})p(\theta \mid \underline{x})$ and $p(y \mid \underline{x}) = \int_\Theta p(y \mid \theta) p(\theta \mid \underline{x}) d\theta$, the criterion (2.4) after a change of the order of the integration becomes

$$\int_{\underline{X}} p(\underline{x}) \int_Y p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{r(y \mid \underline{x})}\right) dy d\underline{x}. \tag{2.5}$$

Since $\int_Y p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{r(y \mid \underline{x})}\right) dy > 0$, equation (2.5) is always positive for any $r(y \mid \underline{x})$ different from $p(y \mid \underline{x})$. This would mean the superiority of the predictive method to the estimative method if averaging over θ is possible. But if criterion (2.3) does not depend on θ and it is positive, then $q(y \mid \underline{x})$ is closer to $p(y \mid \theta)$ than $r(y \mid \underline{x})$. If replacing $q(y \mid \underline{x})$ by $p(y \mid \underline{x})$ makes equation (2.3) free from θ, then it must be positive according to equation (2.5). Therefore, $p(y \mid \underline{x})$ will be better than $r(y \mid \underline{x})$.

3. The Bayesian Predictive Method for Estimation of Weibull Density

In this section we will compare the estimative method and the Bayesian predictive method for estimating the Weibull density by applying the Aitchison criterion.

Suppose a future observation with probability density function

$$f(x;\, \alpha,\, \beta) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], \quad x \geq 0 \tag{3.1}$$

and suppose also that a random sample $\underline{x} = (x_1, x_2, \ldots, x_n)$ is available from the above distribution. We shall estimate the Weibull density when α is unknown and β is known and when both α and β are unknown.

3.1. Estimating the Weibull density when β is known

First we will assume that the distribution of a future observation (say) y is given by equation (3.1) and assume also that the distribution for n observations $\underline{x} = (x_1, x_2, \ldots, x_n)$ is available from the same above distribution. The maximum likelihood estimate of α when β is known is $\hat{\alpha}$. By replacing α by $\hat{\alpha} = \left(\frac{t}{n}\right)^{1/\beta}$ in $p(y \mid \alpha, \beta)$ the estimative density function becomes

$$p(y \mid \alpha = \hat{\alpha}, \beta) = \frac{\beta}{\left(\left(\frac{t}{n}\right)^{1/\beta}\right)^{\beta}} y^{\beta-1} \exp\left[-\left(\frac{y}{\left(\frac{t}{n}\right)^{1/\beta}}\right)^{\beta}\right]$$

$$= \frac{n\beta}{t} y^{\beta-1} \exp\left[-\frac{ny}{t}\right], \quad y > 0, \ \alpha, \beta > 0 \tag{3.2}$$

since $t = \sum_{i=1}^{n} x_i^{\beta}$.

If we combined the prior distribution of α which is in the form

$$p(\alpha) = b_o^{a_o}\beta \exp\left(-\frac{b_o}{\alpha^{\beta}}\right) \Big/ [\Gamma(a_o)\alpha^{\beta a_o+1}]; \quad \alpha \geq 0$$

and the likelihood function, then we will obtain the posterior distribution of α given $\underline{x}$ which takes the expression

$$p(\alpha \mid \underline{x}) = \frac{\beta(b_o + t)^{n+a_o}}{\Gamma(n + a_o)\alpha^{(n+a_o)\beta+1}} \exp\left[-\left(\frac{b_o + t}{\alpha^{\beta}}\right)\right]$$

$$= \frac{\beta b^{a}}{\Gamma(a)\alpha^{a\beta+1}} \exp\left[-\frac{b}{\alpha^{\beta}}\right], \quad \alpha \geq 0, \tag{3.3}$$

where $a = n + a_o$ and $b = b_o + t$.

The predictive density function from equation (1.2) is

$$p(y \mid \underline{x}) = \int_0^\infty p(y \mid \alpha, \beta) p(\alpha \mid \underline{x}) d\alpha = \frac{\beta b^a y^{\beta-1}}{\beta(1, a)(y^\beta + b)^{a+1}}, \quad y > 0 \qquad (3.4)$$

which is the inverse beta distribution written Inb(1, a, b).

If we take $a_o = b_o = 0$ when $a = n + a_o$ and $b = b_o + t$, then

$$p(y \mid \underline{x}) = \frac{\beta t^n y^{\beta-1}}{\beta(1, n)(y^\beta + t)^{n+1}} = \frac{n\beta t^n y^{\beta-1}}{(y^\beta + t)^{n+1}}, \quad y > 0 \qquad (3.5)$$

and if we compare it with equation (3.1), we will get

$$\log\left(\frac{p(y \mid \underline{x})}{p(y \mid \hat{\alpha}, \beta)}\right) = \log\left[\frac{t^{n+1}\exp\left[\frac{ny^\beta}{t}\right]}{(y^\beta + t)^{n+1}}\right] = \frac{ny^\beta}{t} - (n+1)\log\left(1 + \frac{y^\beta}{t}\right), \qquad (3.6)$$

where the above expression depends on y^β and t only through the ratio y^β / t. The distribution of $z = y^\beta / t$ is Inb(1, n, 1) and it is independent of α. Since $E(z) = \frac{1}{n-1}$ and:

$$\int_0^\infty p(z)\log(1+z)dz = \Psi(n+1) - \Psi(n),$$

where $\Psi(h) = \frac{d}{dh}\log\Gamma(h)$,

$$\int_{\underline{X}} \int_Y p(\underline{x} \mid \alpha, \beta) p(y \mid \alpha, \beta) \log\left(\frac{p(y \mid \underline{x})}{p(y \mid \hat{\alpha}, \beta)}\right) dy d\underline{x}$$

$$= (n+1)\log t + \frac{n}{n-1} - (n+1)\log(\Psi(n+1) - \Psi(n)) \qquad (3.7)$$

is free from α. This indicates that the predictive estimate is better than the estimative density based on Kullback-Leibler measure of divergence.

3.2. Estimating the Weibull density when β is unknown

When we assume that both α and β are unknown then they will be replaced by their estimates $\hat{\alpha}$ and $\hat{\beta}$ (where β is estimated by using Newton-Raphson method). In this case the estimative density function will be

$$p(y \mid \hat{\alpha}, \hat{\beta}) = \frac{\hat{\beta}}{\hat{\alpha}^{\hat{\beta}}} y^{\hat{\beta}-1} \exp\left[-\frac{y^{\hat{\beta}}}{\hat{\alpha}^{\hat{\beta}}}\right], \quad y > 0, \ \alpha, \beta > 0. \tag{3.8}$$

To get the predictive density function for y given $\underline{x}$, we will use the posterior distribution of α and β which is given by

$$p(\alpha, \beta \mid \underline{x}) = \frac{\dfrac{\beta^n u^{\beta-1} e^{-\frac{t}{\alpha^\beta}}}{\alpha^{n\beta+1}}}{\Gamma(n) I_{10}}, \quad \text{where } u = \prod_{i=1}^{n} x_i.$$

Therefore

$$p(y \mid \underline{x}) = \int_0^\infty \int_0^\infty \frac{\beta}{\alpha^\beta} y^{\beta-1} \exp\left[-\frac{y^\beta}{\alpha^\beta}\right] \frac{\dfrac{\beta^n u^{\beta-1} e^{-\frac{t}{\alpha^\beta}}}{\alpha^{n\beta+1}}}{\Gamma(n) I_1} d\alpha d\beta$$

$$= \frac{\Gamma(n+1) I_2}{\Gamma(n) I_1} = \frac{n I_2}{I_1}, \tag{3.9}$$

where $I_1 = \int_0^\infty \frac{\beta^{n-1} u^{\beta-1}}{t^n} d\beta$ and $I_2 = \int_0^\infty \frac{\beta^n y^{\beta-1} u^{\beta-1}}{(y^\beta + t)^{n+1}} d\beta.$

It will be difficult (even numerically) to compare the closeness of the estimative density (3.8) and the predictive density (3.9) to the true density $p(y \mid \alpha, \beta)$ based on Kullback-Leibler measure. Therefore we will use the mean square error as a measure of the divergence. Then the divergence of $p(y \mid \hat{\alpha}, \hat{\beta})$ from $p(y \mid \alpha, \beta)$ will be given by

$$M_1 = \int_0^\infty [p(y \mid \alpha, \beta) - p(y \mid \hat{\alpha}, \hat{\beta})]^2 p(y \mid \alpha, \beta) dy. \tag{3.10}$$

We note that M_1 depends on the sample $\underline{x}$, therefore we will use its expected value with respect to $\underline{x}$. This will give

$$E(M_1) = \int_{\underline{X}} M_1 p(\underline{x} \mid \alpha, \beta) d\underline{x}. \tag{3.11}$$

Also the divergence of $p(y \mid \underline{x})$ from $p(y \mid \alpha, \beta)$ is also given by

$$M_2 = \int_0^{\infty} [p(y \mid \alpha, \beta) - p(y \mid \underline{x})]^2 p(y \mid \alpha, \beta) dy. \tag{3.12}$$

If we take the $E(M_2)$ as above, then we will get

$$E(M_2) = \int_{\underline{X}} M_2 p(\underline{x} \mid \alpha, \beta) d\underline{x}. \tag{3.13}$$

The predictive density function will be closer to the true density than the estimative density if

$$E(M_2) < E(M_1) \tag{3.14}$$

and we will check this by the following numerical calculations. $E(M_1)$ can be estimated by

$$\frac{1}{N} \frac{1}{M} \sum_{i=1}^{N} \sum_{\underline{x}=1}^{M} [p(y_i \mid \alpha, \beta) - p(y_i \mid \hat{\alpha}, \hat{\beta})]^2, \tag{3.15}$$

where y_i will take very large number N from its valid values, while there will be M samples $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M$ with $M = 1000$. Also $E(M_2)$ can be estimated by

$$\frac{1}{N} \frac{1}{M} \sum_{i=1}^{N} \sum_{x=1}^{M} [p(y_i \mid \alpha, \beta) - p(y_i \mid \underline{x})]^2, \tag{3.16}$$

where $p(y \mid \alpha, \beta)$, $p(y \mid \hat{\alpha}, \hat{\beta})$ and $p(y \mid \underline{x})$ are given by equations (2.7), (3.8) and (3.9), respectively.

4. Numerical Calculations

Now, we will compare the closeness of the estimative density $p(y \mid \hat{\alpha}, \hat{\beta})$ and the predictive density $p(y \mid \underline{x})$ to the true density $p(y \mid \alpha, \beta)$, when both α and β are unknown, for the Weibull distribution. So we used the expected mean square error (MSE) as a measure of the divergence of $p(y \mid \hat{\alpha}, \hat{\beta})$ from $p(y \mid \alpha, \beta)$ which is expressed by $E(M_1)$ and the divergence of $p(y \mid \underline{x})$ from $p(y \mid \alpha, \beta)$ which is given by $E(M_2)$, then we will compare between $E(M_1)$ and $E(M_2)$ to see which density will estimate the true density better.

As example, the results in Table 4.1 for $n = 10$ and certain values of α and β.

Table 4.1. The results of $E(M_1)$ and $E(M_2)$ for $n = 10$

α \ β	1		2	
	$E(M_1)$	$E(M_2)$	$E(M_1)$	$E(M_2)$
2	.0177	.0138		
3	.0234	.0131	.0178	.0098
4	.0302	.0134	.0234	.0104

It is noticed from Table 4.1 that $E(M_1) > E(M_2)$ for the values calculated which indicates that the predictive density is closer than the estimative density to the true density. This shows that $p(y \mid \underline{x})$ is a better estimator than $p(y \mid \hat{\alpha}, \hat{\beta})$ for $p(y \mid \alpha, \beta)$ based on the mean square error as a criterion.

5. Asymptotic Predictive Distribution

We will concentrate in this section on the work of Dunsmore [6] who derived an approximation for the Bayesian predictive distribution as defined in equation (1.2) when n is large.

The asymptotic form of $p(\theta \mid \underline{x})$ has been broadly discussed (see Lindley [9]). However, for reasonable conditions, the posterior density function is asymptotically normal with mean $\hat{\theta}$ and variance $(nC_{\hat{\theta}})^{-1}$, where $\hat{\theta}$ is the maximum likelihood estimation and $C_{\hat{\theta}}$ is the analogue of Fisher's measure of information, that is,

$$C_{\hat{\theta}} = -\frac{\partial^2 \log p(\underline{x} \mid \theta)}{n \partial \theta^2}.$$

By using Tayler's theorem Dunsmore [5] expanded $p(y \mid \theta)$ about $\hat{\theta}$ as follows:

$$p(y \mid \underline{x}) = \int_\Theta \left[p(y \mid \hat{\theta}) + (\theta - \hat{\theta}) \frac{\partial p(y \mid \hat{\theta})}{\partial \theta} + \frac{1}{2} (\theta - \hat{\theta})^2 \frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2} + \cdots \right] p(\theta \mid \underline{x}) d\theta$$

$$= p(y \mid \hat{\theta}) + \frac{\partial p(y \mid \hat{\theta})}{\partial \theta} \int_\Theta (\theta - \hat{\theta}) p(\theta \mid \underline{x}) d\theta$$

$$+ \frac{1}{2} \frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2} \int_\Theta (\theta - \hat{\theta})^2 p(\theta \mid \underline{x}) d\theta + \cdots,$$

since $\int_\Theta \theta p(\theta|\underline{x}) d\theta \simeq \hat{\theta}$ is almost the posterior mean and $\int_\Theta (\theta - \hat{\theta})^2 p(\theta|\underline{x}) d\theta = (nC_{\hat{\theta}})^{-1}$ is almost the posterior variance,

$$p(y \mid \underline{x}) = p(y \mid \hat{\theta}) + \frac{1}{2} (nC_{\hat{\theta}})^{-1} \frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2} + \cdots, \tag{5.1}$$

where Dunsmore [5] has retained terms up to $O\left(\frac{1}{n}\right)$.

Similarly when we have two unknown parameters (say) α and β, then $p(y \mid \underline{x})$ is given by $p(y \mid \underline{x}) = \int_\beta \int_\alpha p(y \mid \alpha, \beta) p(\alpha, \beta \mid \underline{x}) d\alpha d\beta$.

Using Taylor's expansion of 2nd order, we get

$$p(y \mid \underline{x}) = \int_\beta \int_\alpha \left[p(y \mid \hat{\alpha}, \hat{\beta}) + (\alpha - \hat{\alpha}) \frac{\partial p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \alpha} + (\beta - \hat{\beta}) \frac{\partial p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \beta} \right.$$

$$+ (\alpha - \hat{\alpha})^2 \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \alpha^2} + 2(\alpha - \alpha)(\beta - \hat{\beta}) \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \alpha \partial \beta}$$

$$\left. + (\beta - \hat{\beta})^2 \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \beta^2} + \cdots \right] p(\alpha, \beta \mid \underline{x}) d\alpha d\beta,$$

since $\int_\beta \int_\alpha p(\alpha, \beta|\underline{x}) d\alpha d\beta = 1$ and since $\int_\alpha \alpha p(\alpha, \beta|\underline{x}) d\alpha \simeq \hat{\alpha}$, $\int_\beta \beta p(\alpha, \beta|\underline{x}) d\beta$

$\simeq \hat{\beta}$, we get directly

$$p(y \mid \underline{x}) = p(y \mid \hat{\alpha}, \hat{\beta}) + \frac{1}{2}(nC_{11})^{-1}_{\hat{\alpha},\hat{\beta}} \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \alpha^2}$$

$$+ (nC_{12})^{-1}_{\hat{\alpha},\hat{\beta}} \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \alpha \partial \beta} + \frac{1}{2}(nC_{22})^{-1}_{\hat{\alpha},\hat{\beta}} \frac{\partial^2 p(y \mid \hat{\alpha}, \hat{\beta})}{\partial \beta^2} + \cdots,$$

where

$$(C_{11})_{\hat{\alpha},\hat{\beta}} = -\frac{\partial^2 \log p(\underline{x} \mid \hat{\alpha}, \hat{\beta})}{n \partial \alpha^2}, \quad (C_{12})_{\hat{\alpha},\hat{\beta}} = -\frac{\partial^2 \log p(\underline{x} \mid \hat{\alpha}, \hat{\beta})}{n \partial \alpha \partial \beta},$$

$$(C_{22})_{\hat{\alpha},\hat{\beta}} = -\frac{\partial^2 \log p(\underline{x} \mid \hat{\alpha}, \hat{\beta})}{n \partial \beta^2}$$

and we have retained terms up to $O\left(\frac{1}{n}\right)$.

In the case of the Weibull distribution when α is unknown and β is known, we have

$$\ln p(\underline{x} \mid \alpha) = n \ln \beta - n \ln \alpha + (\beta - 1)\left[\sum_{i=1}^{n} \ln x_i - n \ln \alpha\right] - \left(\frac{1}{\alpha}\right)^{\beta} \sum_{i=1}^{n} x_i^{\beta},$$

then

$$\frac{d \ln p}{d\alpha} = -\frac{n}{\alpha} - \frac{n(\beta - 1)}{\alpha} + \frac{\sum_{i=1}^{n} x_i^{\beta}}{\alpha^{\beta+1}}$$

and

$$\frac{d^2 \ln p}{d\alpha^2} = \frac{n}{\alpha^2} + \frac{n(\beta - 1)}{\alpha^2} - \frac{\beta(\beta + 1)\sum_{i=1}^{n} x_i^{\beta}}{\alpha^{\beta+2}} = \frac{n\beta}{\alpha^2} - \frac{\beta(\beta + 1)\sum_{i=1}^{n} x_i^{\beta}}{\alpha^{\beta+2}}.$$

Then

$$nC_{\hat{\alpha}} = \frac{n^2\beta}{\hat{\alpha}} - \frac{n\beta(\beta + 1)\sum_{i=1}^{n} x_i^{\beta}}{\alpha^{\beta+2}}.$$

Now, we have $p(y|\alpha) = \frac{\beta y^{\beta-1}}{\alpha^{\beta}} \exp\left[-\frac{y^{\beta}}{\alpha^{\beta}}\right]$, then

$$\frac{dp(y|\alpha)}{d\alpha} = \beta y^{\beta-1}\left[\frac{\alpha^{\beta}\left(\frac{\beta y^{\beta}}{\alpha^{\beta+1}}\exp\left(-\frac{y^{\beta}}{\alpha^{\beta}}\right) - \beta\alpha^{\beta-1}\exp\left(-\frac{y^{\beta}}{\alpha^{\beta}}\right)\right)}{\alpha^{2\beta}}\right]$$

$$= p(y|\alpha)\left(\frac{\beta y^{\beta}}{\alpha^{\beta+1}} - \frac{\beta}{\alpha}\right).$$

Therefore

$$\frac{d^2 p(y|\alpha)}{d^2\alpha} = p(y|\alpha)\left(-\frac{\beta(\beta-1)y^{\beta}}{\alpha^{\beta+2}} + \frac{\beta}{\alpha^2}\right) + p(y|\alpha)\left(\frac{\beta y^{\beta}}{\alpha^{\beta+1}} - \frac{\beta}{\alpha}\right)^2$$

$$= p(y|\alpha)\left(\frac{\beta(\beta+1)}{\alpha^2} - \frac{\beta(3\beta+1)y^{\beta}}{\alpha^{\beta+2}} + \frac{\beta^2 y^{2\beta}}{\alpha^{2\beta+2}}\right).$$

Then substituting in (5.1), we get

$$p(y|\underline{x}) = p(y|\hat{\alpha}) + \frac{1}{2}p(y|\hat{\alpha})\left[\frac{\frac{\beta(\beta+1)}{\hat{\alpha}^2} - \frac{\beta(3\beta+1)y^{\beta}}{\hat{\alpha}^{\beta+2}} + \frac{\beta^2 y^{2\beta}}{\hat{\alpha}^{2\beta+2}}}{\frac{n^2\beta}{\hat{\alpha}^2} - \frac{n\beta(\beta+1)}{\hat{\alpha}^{\beta+2}}\sum_{i=1}^{n} x_i^{\beta}}\right] + \cdots$$

$$\simeq p(y|\hat{\alpha})\left[1 - \frac{1}{2n^2}\left(\frac{\beta+1}{\beta} - \frac{(3\beta+1)y^{\beta}}{\beta\hat{\alpha}^{\beta}} + \frac{y^{2\beta}}{\hat{\alpha}^{2\beta}}\right)\right]. \tag{5.2}$$

Equation (5.2) may be written as

$$p(y|\underline{x}) = p(y|\hat{\alpha})\left[1 + O\left(\frac{1}{n^2}\right)\right]. \tag{5.3}$$

This means that the predictive density is almost equal to the estimative density, when n is large.

In the case when both α and β are unknown and where $p(y|\hat{\alpha}, \hat{\beta})$ is given by equation (3.8), then $p(y|\underline{x})$ takes an expression similar to (5.3) which also indicates that, the predictive density is almost equal to the estimative density when n is large.

6. Optimal Approximation for the Predictive Density

The Kullback and Leibler [8] measure of divergence is given here

$$\int_Y p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{w(y \mid \underline{x})}\right) dy. \tag{6.1}$$

It is positive unless $w(y \mid \underline{x})$ is greater than $p(y \mid \underline{x})$. If $w(y \mid \underline{x})$ is constrained to be within the same family of $p(y \mid \theta)$, then $w(y \mid \underline{x}) = p(y \mid \theta = \theta^*)$ is the optimal estimate of $p(y \mid \underline{x})$, where θ^* an estimate of θ is chosen to minimize

$$\int_Y p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{p(y \mid \theta^*)}\right) dy. \tag{6.2}$$

Equivalently we need to select θ^* to maximize

$$\int_Y p(y \mid \underline{x}) \log p(y \mid \theta^*) dy. \tag{6.3}$$

Under regularity conditions, the value θ^* which maximizes (6.3) is obtained by setting

$$\frac{\partial}{\partial \theta^*} \int_Y p(y \mid \underline{x}) \log p(y \mid \theta^*) dy = 0. \tag{6.4}$$

Assume that $p(y \mid \theta)$ is one parameter exponential family, that is,

$$p(y \mid \theta) = \exp(a(\theta) b(y) + c(\theta) + d(y)) \tag{6.5}$$

and using equation (6.4) the solution will be

$$E_{Y/\underline{X}}(b(y)) = -\frac{c'(\theta^*)}{a'(\theta^*)}. \tag{6.6}$$

As an example, we consider the case of the Weibull distribution when

β is known where we assume that $p(y \mid \alpha)$ defined by equation (1.5) is one parameter exponential family, that is,

$$p(y \mid \alpha) = \exp(A(\alpha)B(y) + C(\alpha) + D(y)),$$

where

$$A(\alpha) = -\frac{1}{\alpha^{\beta}}, \quad B(y) = y^{\beta} \text{ and } C(\alpha) = -\beta \ln \alpha.$$

Therefore, by using the equation

$$\frac{\partial}{\partial \theta^*} \int_Y p(y \mid \underline{x}) \log p(y \mid \alpha^*) dy = 0,$$

the solution becomes

$$E_{Y/\underline{X}}(B(y)) = -\frac{C'(\theta^*)}{A'(\theta^*)}$$

which gives $\alpha^* = \left(\frac{t}{n-1}\right)^{\frac{1}{\beta}}$.

7. Asymptotic Goodness of Predictive Fit

The main idea of this section is to find an asymptotic measure of the relative closeness of $p(y \mid \underline{x})$ and $p(y \mid \hat{\theta})$ to the true density $p(y \mid \theta)$. By using Taylor's theorem for expanding $p(y \mid \theta)$ about $\hat{\theta}$, we get

$$p(y \mid \theta) = p(y \mid \hat{\theta}) + (\theta - \hat{\theta}) \frac{\partial p(y \mid \hat{\theta})}{\partial \theta} + (\theta - \hat{\theta})^2 \frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2} + \cdots.$$

Retaining the terms of order $(\theta - \hat{\theta})^2$, we find that

$$p(y \mid \theta) \log \frac{p(y \mid \theta)}{p(y \mid \hat{\theta})} = (\theta - \hat{\theta}) \frac{\partial p(y \mid \hat{\theta})}{\partial \theta} + \frac{1}{2} (\theta - \hat{\theta})^2 \frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2}$$

$$+ \frac{1}{2} \frac{(\theta - \hat{\theta})^2}{p(y \mid \hat{\theta})} \left(\frac{\partial p(y \mid \hat{\theta})}{\partial \theta} \right)^2 + \cdots. \tag{7.1}$$

The posterior density function $p(\theta \mid \underline{x})$ for large samples is asymptotically normal with mean $\hat{\theta}$ and variance $(nC_{\hat{\theta}})^{-1}$, where $C_{\hat{\theta}} = -\frac{\partial^2 \log p(\underline{x} \mid \theta)}{n \partial \theta^2}$.

Integrating equation (7.1) first with respect to θ and then with respect to y will result in the following:

$$\int_Y \int_\Theta p(\theta \mid \underline{x}) p(y \mid \theta) \log\left(\frac{p(y \mid \theta)}{p(y \mid \hat{\theta})}\right) d\theta dy$$

$$\simeq \frac{1}{2}(nC_{\hat{\theta}})^{-1} \int_Y p(y \mid \hat{\theta}) \left(\frac{\partial \log p(y \mid \hat{\theta})}{\partial \theta}\right)^2 dy = \frac{1}{2}(nC_{\hat{\theta}})^{-1} I_Y(\hat{\theta}),$$

where $I_Y(\hat{\theta})$ is the amount of Fisher's information in a single observation from $p(y \mid \theta)$. Then

$$I = \int_{\underline{X}} \int_Y \int_\Theta p(\theta) p(\underline{x} \mid \theta) p(y \mid \theta) \log\left(\frac{p(y \mid \theta)}{p(y \mid \hat{\theta})}\right) d\theta dy d\underline{x}$$

$$= \int_{\underline{X}} p(\underline{x}) d\underline{x} \left[\frac{1}{2}(nC_{\hat{\theta}})^{-1} I_Y(\hat{\theta})\right]. \tag{7.2}$$

It has been noticed in many cases that the value between braces in equation (7.2) does not depend on $\underline{x}$, then we have

$$I \simeq \frac{1}{2}(nC_{\hat{\theta}})^{-1} I_Y(\hat{\theta}) = IA \quad \text{(say)}. \tag{7.3}$$

Also the divergence of $p(y \mid \hat{\theta})$ from $p(y \mid \underline{x})$ is given by

$$\int_Y p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{p(y \mid \hat{\theta})}\right) dy \simeq \frac{1}{8}(nC_{\hat{\theta}})^{-2} \int_Y \frac{1}{p(y \mid \hat{\theta})} \left(\frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2}\right)^2 dy.$$

Therefore

$$J = \int_x \int_Y \int_\Theta p(\theta) p(\underline{x} \mid \theta) p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{p(y \mid \hat{\theta})}\right) d\theta dy d\underline{x}$$

$$= \int_{\underline{X}} p(\underline{x}) d\underline{x} \left[\frac{1}{8}(nC_{\hat{\theta}})^{-2} \int_Y \frac{1}{p(y \mid \hat{\theta})} \left(\frac{\partial^2 p(y \mid \hat{\theta})}{\partial \alpha^2}\right) dy\right]. \tag{7.4}$$

Also in many cases the value between braces in equation (7.4) does not depend on $\underline{x}$ and as a result

$$J \simeq \frac{1}{8}(nC_{\hat{\theta}})^{-2} \int_Y \frac{1}{p(y \mid \hat{\theta})} \left(\frac{\partial^2 p(y \mid \hat{\theta})}{\partial \theta^2} \right)^2 dy = JA \quad \text{(say)} \tag{7.5}$$

which is of order $\frac{1}{n^2}$.

In the case of the Weibull distribution when α is unknown and β is known the divergence of $p(y \mid \hat{\alpha}, \beta)$ from $p(y \mid \alpha, \beta)$ is as follows:

$$I = \int_{\underline{X}} \int_Y \int_\alpha p(\underline{x}) p(\alpha \mid \underline{x}) p(y \mid \alpha, \beta) \log\left(\frac{p(y \mid \alpha, \beta)}{p(y \mid \hat{\alpha}, \beta)} \right) d\alpha dy d\underline{x} = \frac{n+1}{2n(n-1)}$$

and

$$IA = \frac{1}{2}(nC_{\hat{\alpha}})^{-1} \int_Y p(y \mid \hat{\alpha}, \beta) \left(\frac{\partial \log p(y \mid \hat{\alpha}, \beta)}{\partial \alpha} \right)^2 dy = \frac{1}{2}(nC_{\hat{\alpha}})^{-1} I_Y(\hat{\alpha}) = \frac{1}{2n}.$$

Also the divergence of $p(y \mid \hat{\alpha}, \beta)$ from $p(y \mid \underline{x})$ is given by

$$J = \int_{\underline{X}} \int_Y p(\underline{x}) p(y \mid \underline{x}) \log\left(\frac{p(y \mid \underline{x})}{p(y \mid \hat{\alpha}, \beta)} \right) dy d\underline{x} = \frac{1}{n(n-1)} + \frac{1}{4n^2},$$

$$JA = \frac{1}{8}(nC_{\hat{\theta}})^{-2} \int_Y \frac{1}{p(y \mid \hat{\alpha}, \beta)} \left(\frac{\partial^2 p(y \mid \hat{\alpha}, \beta)}{\partial \alpha^2} \right)^2 dy = \frac{1}{n^2}.$$

Now we will compare between I and IA and between J and JA for different values of n as in Table 7.1.

Table 7.1. Values of I, IA, J and JA for selected values of n

n	I	IA	J	JA
10	.061	.050	.0110	.0100
15	.038	.033	.0060	.0040
20	.028	.025	.0030	.0030
25	.022	.020	.0020	.0020
30	.018	.017	.0010	.0010
40	.013	.013	.0008	.0006
50	.010	.010	.0005	.0004

It is clear that the values of IA and JA are better than the values of I and J for different values of n, which indicates that the approximations IA and JA of I and J, respectively, are good asymptotic measures of the closeness of the predictive density and the estimative density to the true density.

References

[1] E. H. Abdelfattah, Goodness of fit tests for the two parameter Weibull distribution, J. Statist. Theo. Appl. 7(3) (2008).

[2] Ezz H. Abdelfattah, Galal M. Elsayyad and Muna I. Alkadi, A Bayesian estimation of the Weibull distribution parameters, Far East J. Theo. Stat. 21(1) (2007), 57-64.

[3] J. Aitchison, Goodness of prediction fit, Biometrika 64 (1975), 547-549.

[4] M. A. Amaral and I. R. Dunsmore, Optimal estimates of predictive distribution, Biometrika 67 (1980), 685-689.

[5] I. R. Dunsmore, The Bayesian predictive distribution in life testing models, Technometrics 16 (1974), 455-460.

[6] I. R. Dunsmore, Asymptotic predictive analysis, Biometrika 63 (1976), 627-630.

[7] S. Geisser, The Inferential Use of Predictive Distribution in Foundations of Statistical Inference, V. P. Godambe and D. A Sports, eds., Holt, Rinehart and Winston, Toronto, 1971, pp. 456-469.

[8] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Statist. 22(1) (1951), 79-86.

[9] D. V. Lindley, Bayesian Statistics, a Review, SIAM, Philadelphia, 1971.

[10] G. Murray, A note on the estimation of probability density function, Biometrika 64 (1977), 150-152.

[11] G. Murray, The estimation of multivariate normal density functions using incomplete data, Biometrika 66 (1979), 375-380.

[12] V. M. Ng, On the estimation of parametric density function, Biometrika 67 (1980), 505-506.

■