A Generalization of Jeffreys' Rule for Nonregular Models

FRANCISCO J. ORTEGA
e-mail: fjortega@us.es
Phone: +34 954556970

JESÚS BASULTO
e-mail: basulto@us.es
Phone: +34 954557539

Address (for both authors):
Departamento de Economía Aplicada I. UNIVERSIDAD DE SEVILLA
Avenida Ramón y Cajal, nº1. 41018-Sevilla.

Author for correspondence: Francisco J. Ortega

A Generalization of Jeffreys' Rule for Nonregular Models

Abstract

We propose a generalization of the one-dimensional Jeffreys' rule in order to obtain noninformative prior distributions for nonregular models, taking into account the coments made by Jeffreys in his article of 1946. These noninformatives are parameterisation-invariant and the Bayesian intervals have good behaviour in Frequentist Inference. In some important cases, we can generate noninformative distributions for multi-parameter models with nonregular parameters. In nonregular models, the Bayesian method offers a satisfactory solution to the inference problem and also avoids the problem that the maximum likelihood estimator has with these models. Finally, we obtain noninformative distributions in Job-Search and Deterministic Frontier Production homogenous models.

Key words: Bayesian Inference, Noninformative Prior Distributions, Nonregular Models.

AMS Classification: 62F15

JEL Classification: C11,C24.

1. Introduction

There is a wide variety of empirical applications where the corresponding models do not verify the regularity conditions of Wald (Azallini, 1996, p.71). Usually, the main problem is that the range of the observed random variable depends on the parameters that are being estimated. In Sareen (2003), Schmidt (1976) and Müller and Wefelmeyer (2010) there are some examples where nonregular models appear in economic applications, as in stationary job-search models, truncated regressions, deterministic frontier production models and sealed-bid auction models.

In models with nonregular parameters, the standard results of parameter consistency and asymptotic normality of the maximum likelihood estimator do not usually hold, so that alternatives schemes to estimate the unknown parameters must be pursued. Researchers have considered some alternatives avenues, as estimating the unknown parameters by using another methods (e.g. the method of moments), or adopting modifications to the original empirical specifications for which the range problem disappears, e.g., by introduction of measurement error as in the literature concerned with the estimation of stochastic production frontiers (Coelli et al. 2005, Tchumtchoua and Dey, 2007), looking sufficient conditions to derive an asymptotic distributions for the maximum likelihood estimator as in Akahira and Takeuchi (1995) and, finally, applying the Bayesian estimation to nonregular models as in Ghosal (1999) and Wiper et al (2008).

The Bayesian method needs to elicit a prior distribution to calculate the posterior distribution from the likelihood function via Bayes' theorem. One procedure of generating prior distributions over parameters spaces is the Jeffreys' rule, which is calculated from the Fisher information. This prior distribution is called, usually, noninformative. Note that Jeffres's rule must verify the regularity conditions to make sense.

The first objective of our article is to give an easy method to obtain prior noninformative distributions, which could be applied to nonregular models. To this end, we will return to the original Jeffreys' works. We will use the interpretation that the author makes of his proposal and the comments about why his proposal is not valid in the case of the uniform model (an important case of nonregular model). Based on this reasoning, we can establish an analogy between regular and nonregular models that allows to obtain a generalization of the meaning of Jeffreys' rule for nonregular models, as we will see in the second section.

The second objective will be to show by mean of two illustrative examples (job-search model and deterministic frontier production model), that using noninformative distributions in Bayesian inference, the solution to the inference problem is also satisfactory when the models are nonregular.

The rest of the article is organized as follows. Section 2 presents our proposal to generate one-dimensional prior distributions. In section 3, we analyze the properties of the prior noninformative distribution, proving that this prior is parameterisation-invariant for some smooth one-to-one transformation and that a Bayesian interval with probability $1-\alpha$, calculated from the posterior distribution, acts as a confidence interval with a level of confidence $1-\alpha$ (exact or approximate). In section 4, we study the multi-parameter case. The two illustrative examples are studied, respectively, in the sections 5 and 6. Finally, section 7 resumes the main conclusions of the present article.

2. The One-Parameter Case.

In this section we shall propose one rule to obtain one-dimensional noninformative prior distributions which is applied to nonregular and regular models.

Our proposal is based on the interpretation that Jeffreys (1946) makes of his rule to obtain prior densities. We will explain this interpretation using more modern notation and applying it to the case of a continuous uniparameter regular model.

Let X be a random variable with density $f(x|\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$ is a unknown parameter. Given two values $\theta_1, \theta_2 \in \Theta$, Jeffreys consider the discrepancy (or information) measures :

$$I_1(\theta_1,\theta_2) = \int \left(\sqrt{f(x|\theta_2} - \sqrt{f(x|\theta_1} \right)^2 dx \tag{1}$$

$$I_2(\theta_1,\theta_2) = \int \log\left(\frac{f(x|\theta_2)}{f(x|\theta_1)} \right) \left(f(x|\theta_2) - f(x|\theta_1) \right) dx \tag{2}$$

Then he analyzes the behaviour of $I_1(\theta,\theta+\Delta\theta)$ and $I_2(\theta,\theta+\Delta\theta)$, concluding that "$I_1$ and I_2 are apparently the only ones that are ordinarily of the second order in the differences of the parameters in the laws when these differences are small" (p.455). This property is the one that finally led him to establish his rule, looking for the invariance property check.

We will use the information measure

$$J(\theta_1,\theta_2) = -8\log\int \sqrt{f(x|\theta_2)}\sqrt{f(x|\theta_1)}dx \tag{3}$$

whose local behaviour is of second order too, as we can see in the following proposition, whose proof can be found in Akahira and Takeuchi (1991).

Proposition 1. In regular models, $J(\theta,\theta+\Delta\theta) = I(\theta)(\Delta\theta)^2 + o((\Delta\theta)^2)$, where $I(\theta)$ is the Fisher information amount of the regular model, defined as

$$I(\theta) = \sqrt{-E\left[\frac{\partial^2 \log\left(f(x|\theta)\right)}{\partial\theta^2} \right]} \tag{4}$$

Thus, we can interpret that the Jeffreys' rule take a prior distribution that give more density of probability to those values of parameter space in which local information is greater, considering a distribution proportional to $\sqrt{I(\theta)}$ in order to verify the invariance property.

Remark 1. In regular models, from the proposition 1, we can write $I(\theta) = \lim_{\Delta\theta\to 0} \frac{J(\theta,\theta+\Delta\theta)}{(\Delta\theta)^2}$ and therefore the Jeffreys' rule can be expressed as

$$\pi(\theta) \propto \sqrt{I(\theta)} = \sqrt{\lim_{\Delta\theta\to 0} \frac{J(\theta,\theta+\Delta\theta)}{(\Delta\theta)^2}} \tag{5}$$

However, the behaviour of information measures in nonregular models for small variations in the parameter is different. Jeffreys (1946) points out that: "The requirement that the chances shall be differentiable with respect to the parameters is not always satisfied. One case is the rectangular distribution [...]. For α' slightly greater than α, I_1 is of the first order in $\alpha' - \alpha$ instead of the second" (p.458).

This difference in behaviour can also be interpreted as a certain analogy. This allows us to define a prior distribution for nonregular models "similar" to the Jeffreys' disribution. Indeed, if $J(\theta,\theta+\Delta\theta)$ is of the first order in $|\Delta\theta|$ when $|\Delta\theta| \to 0$, there is a function $C(\theta)$ (not identically null) such that $J(\theta,\theta+\Delta\theta) = C(\theta)|\Delta\theta| + o(|\Delta\theta|)$ and therefore $\lim_{\Delta\theta\to 0} \left[J(\theta,\theta+\Delta\theta) / |\Delta\theta| \right] = C(\theta)$. Following the interpretation of Jeffrey's rule, we will define a prior distribution that give more density of probability to those values of parameter space in which local information is greater; we will propose

$$\pi(\theta) \propto C(\theta) = \lim_{\Delta\theta\to 0} \frac{J(\theta,\theta+\Delta\theta)}{|\Delta\theta|} \tag{6}$$

We want to emphasize that now we do not consider the square root of $C(\theta)$ because of the local behaviour is of the first order instead of the second and, as we shall see in section 3, this proposal is parameterization-invariant.

Remark 2. Note that the two situations described above can be written as

$$\pi(\theta) \propto \left(\lim_{\Delta\theta \to 0} \frac{J\left(\theta, \theta + \Delta\theta\right)}{|\Delta\theta|^k} \right)^{1/k} \tag{7}$$

where $k = 2$ for regulars models and $k = 1$ for nonregular models.

Considering nonregular models, we obtain noninformative prior distribution for simple examples.

Example 1. If we consider the uniform model $U\left(0,\theta\right)$, $\theta \in \left(0,+\infty\right)$, with density function $f(x\,|\,\theta) = \theta^{-1}$, $0 \le x \le \theta$, we obtain $J(\theta,\theta+\Delta\theta) = -4\log\left(\theta/(\theta+\Delta\theta)\right)$ for $\Delta\theta > 0$; while for $\Delta\theta < 0$, we obtain $J(\theta,\theta+\Delta\theta) = -4\log\left((\Delta\theta+h)/\Delta\theta\right)$, and so

$$\lim_{\Delta\theta \to 0^+} \left(J(\theta+\Delta\theta)/|\Delta\theta|\right) = \lim_{\Delta\theta \to 0^-} \left(J(\theta+\Delta\theta)/|\Delta\theta|\right) = 4\theta^{-1} \tag{8}$$

Thus the prior distribution is $\pi\left(\theta\right) \propto \theta^{-1}$.

Example 2. If we consider the uniform model $U\left(\theta - 1/2, \theta + 1/2\right)$, $\theta \in \mathbb{R}$, with density function $f(x,\theta) = 1$, $\theta - 1/2 \le x \le \theta + 1/2$, we obtain $J(\theta,\theta+h) = -8\log(1-\Delta\theta)$ if $\Delta\theta > 0$ (for $\Delta\theta$ sufficiently small), while that for $\Delta\theta < 0$ the amount of information is $J(\theta,\theta+h) = -8\log(1+\Delta\theta)$. It is very easy to prove that $\lim_{\Delta\theta \to 0}\left(J(\theta,\theta+\Delta\theta)/|\,\Delta\theta\,|\right) = 8$, and, consequently, the prior distribution is $\pi(\theta) \propto 1$.

In the two nonregular previous examples, the rate of convergent of $J(\theta,\theta+\Delta\theta)$ to zero is of order $|\,\Delta\theta\,|$, while in the regular models the rate is of order $(\Delta\theta)^2$, as we pointed out earlier. It is interesting to recall that the usual characterization of a regular model $f(x;\theta)$ is that the parameter θ does not appear on endpoints of the interval which $f(x;\theta)>0$. But a general definition of a regular model is that $\lim_{\Delta\theta\to 0}\left(J(\theta,\theta+\Delta\theta)/(\Delta\theta)^2\right)=I(\theta)$, and so the usual characterization is not a necessary condition (Pitman, 1979, pág.13). Note that Pitman (1979) works with $\tilde{J}(\theta_1,\theta_2)=\int\left(f(x\mid\theta_1)^{1/2}-f(x\mid\theta_2)^{1/2}\right)^2 dx$, but is easy to prove that both information measures have the same local behaviour, that is, $\lim_{\Delta\theta\to 0}\left(J(\theta,\theta+\Delta\theta)/\tilde{J}(\theta,\theta+\Delta\theta)\right)=4$ and consequently is the same to use either one or another amount of information.

In the following proposition, we prove that in nonregular models, under certain conditions, the rate of convergent is of order $|\Delta\theta|$ and we also obtain one alternative expression to calculate the noninformative prior distribution.

Proposition 2. Let $X_1,...,X_n$ be independent and identically distributed observations from a density f(x|θ), where $\theta\in\Theta$, and Θ is an open interval(bounded or unbounded) in $\mathbb{R}$. We make the following assumptions:

i) $\forall\theta\in\Theta$, $f(\bullet|\theta)$ is strictly positive in a closed interval (bounded or unbounded) $S(\theta)=\left[a_1(\theta),a_2(\theta)\right]$,which depends on θ, and is zero outside $S(\theta)$. It is allowed that one of the endpoints is free of θ and may be plus or minus infinity.

ii) The sets $S(\theta)$ are either increasing or decreasing in θ (that is, $\forall\ \theta_1<\theta_2$ we have $S(\theta_1)\subseteq S(\theta_2)$ or $S(\theta_1)\supseteq S(\theta_2)$).

iii) We assume that $a_1(\theta)$ and $a_2(\theta)$ are strictly monotone and continuously differentiable functions unless they are infinite or free from θ.

iv) On the set $R(x,\theta)=\left\{(x,\theta)\in\mathbb{R}^2 : x\in S(\theta)\right\}$, either $f(x\mid\theta)$ and $\partial f(x\mid\theta)/\partial\theta$ are jointly continuous in (x,θ).

Then:

$$\lim_{h \to 0} \frac{J(\theta, \theta + \Delta\theta)}{|\Delta\theta|} = \left| 4E\left[\frac{\partial \log f(x|\theta)}{\partial\theta} \right] \right| \qquad (9)$$

The proof of this proposition is given in the appendix.

From the proposition 2, the prior distribution that we propose is

$$\pi(\theta) \propto \left| E\left[\frac{\partial \log f(x|\theta)}{\partial\theta} \right] \right| \qquad (10)$$

The conditions stated in proposition 2 are the same as those established by Ghosal (1999). Ghosal has obtained a matching prior in nonregular cases which has the same functional form of (10). Therefore, if conditions of proposition 2 hold, both prior distributions remain the same, but if not, we have the advantage of being able to obtain the prior distribution applying (6). For example, the model $U(0, \theta)$ fall in all conditions of proposition 2, and so its prior distribution can be obtained from the expression

$$\pi(\theta) \propto \left| E\left[\left(\partial \log f(x|\theta) \right) / \partial\theta \right] \right| \propto \left| E\left[-\left(\partial \log\theta / \partial\theta \right) \right] \right| \propto \theta^{-1}, \qquad (11)$$

but the model $U(\theta - 1/2, \theta + 1/2)$ does not verify condition ii) of proposition 2, so, in this case, it is not possible to get the matching prior of Ghosal applying the formula (10), but we have obtained the prior $\pi(\theta) \alpha 1$ applying the formula (6).

3. Properties of the noninformative prior distribution.

There are two procedures to evaluate the methods for obtaining noninformative prior distributions. One is the parameterisation-invariant for some smooth one-to-one transformation, and the other one is the so-called matching priors. In this section we will see that the noninformative prior that we have proposed has a good behaviour in both procedures.

The invariance property is proved in the following proposition.

Proposition 3. The noninformative prior distribution is parameterisation-invariant for some smooth one-to-one transformation.

Proof. The measure of information $J(\theta_1, \theta_2)$ does not depend on the choice of parameterisation. For the values θ_1, θ_2 and transformation $\lambda = \lambda(\theta)$ one-to-one (continuous and differentiable), we have $J(\theta_1, \theta_2) = J(\lambda_1, \lambda_2)$, where $\lambda_i = \lambda(\theta_i), i = 1, 2.$

Now, if we consider the values θ and $\Delta\theta$, and its respective $\lambda = \lambda(\theta)$ and $\Delta\lambda = \lambda(\theta + \Delta\theta) - \lambda(\theta)$,we obtain that

$$J(\theta, \theta + \Delta\theta) = J(\lambda(\theta), \lambda(\theta + \Delta\theta)) = J(\lambda, \lambda + \Delta\lambda) .$$

Hence,

$$\frac{J(\theta, \theta + \Delta\theta)}{\Delta\theta} = \frac{J(\lambda, \lambda + \Delta\lambda)}{\Delta\lambda}\frac{\Delta\lambda}{\Delta\theta} = \frac{J(\lambda, \lambda + \Delta\lambda)}{\Delta\lambda}\frac{\lambda(\theta + \Delta\theta) - \lambda(\theta)}{\Delta\theta}$$

and for $\Delta\theta \to 0$, we obtain $\pi(\theta) = \pi(\lambda)\left|\frac{\partial\lambda}{\partial\theta}\right|$ and this proves the proposition.

■

The matching priors procedure is a method of generating prior distributions over parameter spaces based on the confidence properties of sets arising from the corresponding posterior distributions. For example, for a single parameter, we can generate with this method a noninformative prior so that a Bayesian interval with probability $1\text{-}\alpha$, calculated from the posterior distribution, acts as a confidence interval with a level of confidence $1\text{-}\alpha$ (exact or approximate).

The notion of matching, which first appeared in Welch and Peers (1963), may be considered as an attempt to reconcile the two different schools, classical and Bayesian, as a validation of a noninformative prior, or as a useful method for constructing confidence sets. For one-dimensional smooth families, matching leads to Jeffreys' prior (Welch and Peers, 1963).

As we have pointed out in the previous section, if conditions of proposition 2 hold, then the distribution proposed is equal to the Ghosal's matching prior. Specifically, in Ghosal(1999) is proved that the Bayesian one-sided interval with probability $1-\alpha$ acts as a confidence interval with a level of confidence equal to $1-\alpha+O(n^{-2})$.

If the conditions of proposition 2 do not hold, there are not general results about matching priors. However, there are several examples where the posterior Bayesian interval, with probability $1-\alpha$, acts as an exact confidence interval with a level of confidence $1-\alpha$ (Basulto, 1997; Ortega and Basulto, 2003).

4. The Multi-Parameter Case.

When there is more than one parameter, that is, when $\theta \in \Theta \subseteq \mathbb{R}^m$, the general Jeffreys' rule, in the regular case, is proportional to the square root of the determinant of information matrix, that is, $\pi(\theta) \propto \sqrt{|I(\theta)|}$, where $|I(\theta)|$ is the determinant of the information matrix. This prior is invariant to arbitrary parameterisation. But, this choice presents important deficiencies so it is not the usual option (Jeffreys, 1961, Ortega and Basulto, 2003). To avoid this problem, Jeffreys proposed one modification of his rule for multi-parameter case which he had applied to models with location-scale parameters. This modified rule is equivalent to obtain: (i) the distribution for each parameter treating the remaining as fixed parameters and (ii) to calculate the multi-parameter prior distribution as product of the corresponding one-dimensional (Jeffreys, 1961, p.182-183).

The most usual method to generate multi-parameter prior distribution is to obtain them from certain one-dimensional distributions (marginal or conditional distributions).

The final prior distribution depends on considering that all parameters are of interest (Bernardo and Smith, 1994, Nicolau, 1993) or considering of interest some of them being the rest nuisance parameters.

For two-dimensional parameter $\theta = (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2$, where the two parameters are of interest, we can calculate the one-dimensional conditional prior distributions $\pi_{1|2}(\theta_1|\theta_2)$ and $\pi_{2|1}(\theta_2|\theta_1)$ and then, look for a join prior distribution $\pi(\theta_1, \theta_2)$ compatible with them (Arnold et al, 1999). An important problem about this calculus is that the distribution $\pi_{1|2}(\theta_1|\theta_2)$ is determined except for an arbitrary function which can dependent of the θ_2, and this function has influence in the join posterior distribution.

Thus, applying one-dimensional rule to parameter θ_1, for fixed θ_2, we find that $\pi_{1|2}(\theta_1 | \theta_2) \propto g_{1|2}(\theta_1, \theta_2) C_{1|2}(\theta_2)$, where $C_{1|2}(\theta_2)$ is an arbitrary function. Similarity, we will obtain $\pi_{2|1}(\theta_2 | \theta_1) \propto g_{2|1}(\theta_1, \theta_2) C_{2|1}(\theta_1)$. From these conditional distributions, we propose the following definition.

Definition. If it is certain that $g_{1|2}(\theta_1, \theta_2) C_{1|2}(\theta_2) = h_{1|2}(\theta_1) h^*_{1|2}(\theta_2)$ and $g_{2|1}(\theta_1, \theta_2) C_{2|1}(\theta_1) = h_{2|1}(\theta_2) h^*_{2|1}(\theta_1)$, then the prior distribution will be

$$\pi(\theta_1, \theta_2) \propto h_{1|2}(\theta_1) h_{2|1}(\theta_2) \tag{12}$$

This definition is taken from the papers of Nicolau(1993) and Ghosal (1999). We observe that in this case, as point out Nicolau, 1993, we can consider that the parameters are a priori independent. Though the previous solution is partial, the method is of interest and is frequently used in the practice (Ghosal, 1999), principally in two-parameter models where one parameter is regular and the other is nonregular.

Example 3. Pareto Model (η, φ)

The density of a Pareto model is given by $f(x|\eta,\varphi)=\varphi\eta^{\varphi}x^{-(1+\varphi)},\ x>\eta,\ \eta,\varphi>0.$ Let $\ell(\eta,\varphi|x)=\log f\left(x|\eta,\varphi\right)$. We can obtain the join prior distribution applying the one-dimensional rule to each individual conditional distribution, so:

A) For each fixed η, the model is regular on the parameter φ, that is, φ is regular, and the conditional noninformative prior distributions for $\varphi|\eta$ is proportional to the square root of the Fisher information of model (fixed η). From $E\left[\partial^2\ell/\partial\varphi^2\right]=-\varphi^{-2}$, the conditional noninformative prior is $\pi\left(\varphi|\eta\right)\propto\varphi^{-1}C_1(\eta)$.

B) For each fixed φ, the model is nonregular on the parameter η and the conditions of the proposition 2 are true. From $E\left[\partial\ell/\partial\eta\right]=\varphi/\eta$, we obtain $\pi(\eta|\varphi)\propto\eta^{-1}C_2(\varphi)$.

From (12), the conditional noninformative priors are the product of the functions $C_2\left(\eta\right)$ and $C_1\left(\varphi\right)$, with different parameters, and hence the join noninformative prior is $\pi\left(\eta,\varphi\right)\propto\eta^{-1}\varphi^{-1}$.

Example 4. Location-scale model

The density of location-scale model is given by

$$f(x|\theta,\varphi)=\varphi^{-1}f_0((x-\theta)/\varphi),\ \theta\in\mathbb{R},\varphi>0 \tag{13}$$

where $f_0\left(z\right)$ is a density function in the interval $\left[0,+\infty\right)$. Suppose that for each fixed θ, the Fisher information $I\left(\varphi/\theta\right)=E\left[\left(\partial\ell/\partial\varphi\right)^2\right]=-E\left[\left(\partial^2\ell/\partial\varphi^2\right)\right]$ exists and is finite. That is, the parameter φ is regular. Suppose that $f_0(0+)>0$ and is finite, that is, the parameter θ is nonregular. In these cases, we find that $E[\partial\ell/\partial\theta]=\varphi^{-1}f_0(0+)$ and $-E\left[\partial^2\ell/\partial\varphi^2\right]=k\varphi^{-2}$, where $k=\int\left[1+xf_0^{'}(x)/\ f_0(x)\right]^2 f_0(x)dx$, and consequently $\pi\left(\theta|\varphi\right)\propto 1\cdot C_1(\varphi)$ and $\pi\left(\varphi|\theta\right)\propto\varphi^{-1}\cdot C_2(\theta)$. Therefore the conditional noninformative priors are, respectively, $C_1(\varphi)$ and $C_2(\theta)$, so, they are functions with different parameters and hence the join noninformative prior is $\pi(\theta,\varphi) \propto \varphi^{-1}$.

If we cannot apply (12), then we can try to looking for a reparameterisation which allows to apply it.

Example 5. Uniform Model (α, β)

The density of a uniform model (α, β) is given by

$$f(x \mid \alpha, \beta) = (\beta - \alpha)^{-1}, \ \alpha \le x \le \beta, \ \alpha < \beta \in \mathbb{R} \tag{14}$$

In this case, the parameter vector (α, β) is nonregular, and the condition of proposition 2 is true when one parameter is fixed. From the expression $\ell(\alpha, \beta \mid x) = \log f(x \mid \alpha, \beta)$, we obtain $\left|E\left[\partial\ell/\partial\alpha\right]\right| = \left|E\left[\partial\ell/\partial\beta\right]\right| = (\beta - \alpha)^{-1}$, consequently $\pi(\alpha \mid \beta) \propto (\beta - \alpha)^{-1} C_1(\beta)$ and $\pi(\beta \mid \alpha) \propto (\beta - \alpha)^{-1} C_2(\alpha)$, that is, the conditional noninformative prior densities are not the product of two functions with different parameter.

We consider the reparameterisation $\mu = \alpha, \ \sigma = \beta - \alpha$, that is, the parameter μ is the minimum value and σ the width the interval (α, β) (μ is a parameter of location and σ of scale). Now, $\ell(\mu, \sigma \mid x) = \log(\sigma^{-1}), \ \mu \le x \le \mu + \sigma$. For each fixed σ, we can applied the situation of example 2, and then the conditional noninformative prior distribution is $\pi(\mu \mid \sigma) \propto 1 \bullet C_1(\sigma)$, while for each fixed μ, we can applied the situation of example 1, and then the conditional noninformative prior distribution is $\pi(\sigma \mid \mu) \propto \sigma^{-1} C_2(\mu)$. Hence, the join noninformative prior distribution is $\pi(\mu, \sigma) \propto \sigma^{-1}$ (for parameter vector (α, β), the join noninformative prior distribution is $\pi(\alpha, \beta) \propto (\beta - \alpha)^{-1}$).

5. Optimal job-search models with homogenous agents

Lancaster (1997) had proposed a job-search model for homogenous unemployed population when the observed data are the duration (T) of search and the accepted wage (W), the resultant n person likelihood for a sample $(\mathbf{t},\mathbf{w}) = ((t_1, w_1),...,(t_n, w_n))$ is

$$L(\theta,\lambda,\xi \mid \mathbf{t},\mathbf{w}) = \lambda^n \exp\left\{-\lambda\bar{F}(\xi \mid \theta)T\right\}\prod_{i=1}^{n} f\left(w_i \mid \theta\right), \quad b < \xi < w_i, i = 1,2,...,n \quad (15)$$

where λ is the rate of arrival of wage offers; $T = \sum_{i=1}^{n} t_i$; $f(w \mid \theta)$, the density function of wage offers W (θ is a parameter vector); $F\left(w|\theta\right)$ the distribution function of wage offers W; $\bar{F}\left(w|\theta\right) = 1 - F\left(w|\theta\right)$; ξ is the reservation wage, where if an offer is greater than ξ then it is accepted, and b is the rate of unemployment benefit.

A hypothesis, H_0, that is central to the optimal job-search model, it is that the agents make utility-maximizing choices. The mathematical effect of this hypothesis is to impose the restriction $\xi = b + (\lambda/\rho)\int_{\xi}^{+\infty} \bar{F}(w \mid \theta)dw = g(\theta,\lambda)$, so ξ is functionally dependent of θ and λ (the values b and the discount rate, ρ, are assumed know to the researcher) and the likelihood function parameters are restricted to $b < g(\theta,\lambda) < w_i, i = 1,2,...,n$. Now the parameter vector (θ,λ) is nonregular because the support of W dependent on them.

If the optimality is no enforced, H_1, the parameter ξ is functionally independent of the remaining parameters, (θ,λ). Now, the parameter vector (θ,λ) is regular and ξ is nonregular.

While the density function of wage offers was assumed Log-Normal in Lancaster (1997), we assume a uniform density truncated in (α,β), where the parameters α and

β are unknown to the researcher. This model is mainly interesting from an operative point of view, because it leads to an easier posterior distribution which is even analytically tractable under the hypothesis H_0. On the other hand, if the population observed is homogenous (for example, people without qualification that is searching their first employment) we think that it would be reasonable to suppose that all wage offers will be similar among them and will be uniformly distributed in a short or moderate length interval.

In this case, the accepted wages W is uniform in (ξ, β), and it is independent of α. If we make the transformation $\sigma = \beta - \xi$, then the accepted wages W are uniform in $(\xi, \xi + \sigma)$.

Hypothesis H_0. The parameter ξ is independent of σ

In this case, $f(w \mid \xi, \sigma) = \sigma^{-1}$, $\xi < w < \xi + \sigma$, $\bar{F}(\xi \mid \sigma) = 1$ and the likelihood function is

$$L(\lambda, \xi, \sigma \mid \mathbf{t}, \mathbf{w}) = \lambda^n \exp\{-\lambda T\} \sigma^{-n}, \quad b < \xi < w_{(1)}, \ \xi + \sigma > w_{(n)} \tag{16}$$

where $w_{(1)}$ and $w_{(n)}$ are the minimum and maximum, respectively, of $\{w_1, ..., w_n\}$.

The log-likelihood function for n=1 is

$$\ell(\lambda, \xi, \sigma \mid \mathbf{t}, \mathbf{w}) = Log\lambda - \lambda t - Log\sigma, \ \xi < w < \xi + \sigma \tag{17}$$

and for the regular parameter λ, conditional to (ξ, σ), we have $-\partial^2 \ell / \partial \lambda^2 = \lambda^{-2}$ and so $\pi(\lambda \mid \xi, \sigma) \propto \lambda^{-1} C_1(\xi, \sigma)$ is the conditional noninformative prior distribution.

The parameter vector (ξ, σ) is nonregular, and from example 5, we have $\pi(\xi \mid \lambda, \sigma) \propto 1 \cdot C_2(\lambda, \sigma)$ and $\pi(\sigma \mid \lambda, \xi) \propto \sigma^{-1} \cdot C_3(\lambda, \xi)$. Then we can assume the independent among the parameters and so the join noninformative prior distribution is $\pi(\lambda, \xi, \sigma) \propto \lambda^{-1} \sigma^{-1}$.

Now, the join posterior distribution is

$$\pi(\lambda,\xi,\sigma\,|\,\mathbf{t},\mathbf{w}) \propto \lambda^{n-1}\exp\{-\lambda T\}\,\sigma^{-(n+1)},\quad b<\xi<w_{(1)},\ \xi+\sigma>w_{(n)} \qquad (18)$$

The constant of integration is $K = nT^n\left((w_{(n)}-b)^{-(n-1)} - r^{-(n-1)}\right)\Big/(n-2)!$ and $r = w_{(n)} - w_{(1)}$ is the range of accepted wages.

The marginal posterior distributions are: for the parameter λ is a gamma distribution with parameters *(n,T);* for ξ is $\pi(\xi\,|\,\vec{t},\vec{w}) \propto (w_{(n)}-\xi)^{-n},\quad b<\xi<w_{(1)}$ and, finally, the marginal posterior density of σ is

$$\pi(\sigma\,|\,\vec{t},\vec{w}) \propto \begin{cases} \sigma^{-(n+1)}(\sigma-r) & si \quad r\le\sigma\le w_{(n)}-b \\ \sigma^{-(n+1)} & si \quad w_{(n)}-b\le\sigma<+\infty \end{cases} \qquad (19)$$

Hypothesis H_1. The parameter ξ is dependent of σ

From the restriction, $\xi = b + (\lambda/\rho)\int_{\xi}^{+\infty}\bar{F}(w\,|\,\theta)dw$ (*b* and ρ are known by the researcher) and we suppose that W is uniform in $(\xi,\xi+\sigma)$, we obtain that $\int_{\xi}^{+\infty}\bar{F}(w\,|\,\theta)dw = \sigma/2$ and so $\xi = b+\lambda\sigma/(2\rho)$. Now the Log-Likelihood function for $n=1$ is

$$\ell(\lambda,\sigma\,|\,t,w) = Log\lambda - \lambda t - Log\sigma,\ \ b+\lambda\sigma/(2\rho) < w < b+\lambda\sigma/(2\rho)+\sigma \qquad (20)$$

For the conditional noninformative prior distributions of $\lambda\,|\,\sigma$ and $\sigma\,|\,\lambda$ is not possible to apply proposition 2 (neither to obtain the matching prior of Ghosal) because the condition ii) is not satisfied. So, for each fixed σ, we obtain $\lim_{\Delta\lambda\to 0}\left(I(\lambda,\lambda+\Delta\lambda)/|\Delta\lambda|\right) = 4\rho^{-1}$, and so $\pi(\lambda\,|\,\sigma) \propto 1\cdot C_1(\sigma)$; for each fixed λ, we obtain $\lim_{\Delta\sigma\to 0}\left(I(\sigma,\sigma+\Delta\sigma)/|\Delta\sigma|\right) = 4\lambda\rho^{-1}\sigma^{-1}$, so the conditional noninformative prior distribution is $\pi\left(\sigma \mid \lambda\right) \propto \sigma^{-1} \cdot C_2(\lambda)$. The join noninformative prior is $\pi\left(\lambda, \xi, \sigma\right) \propto \sigma^{-1}$. Now the join posterior distribution is

$$\pi(\lambda, \sigma \mid \mathbf{t}, \mathbf{w}) \propto \lambda^n \exp\left\{-\lambda T\right\} \sigma^{-(n+1)}, \quad b + \frac{\lambda\sigma}{2\rho} < w_{(1)},\, b + \frac{\lambda\sigma}{2\rho} + \sigma > w_{(n)} \qquad (21)$$

This last distribution is a bit more complex because it depends on the range of vector $\left(\lambda, \sigma\right)$, so its constant of integration has not closed-form solution.

6. Deterministic frontier production models.

Examples of old models with nonregular parameters are the deterministic frontier production models of Aigner and Chu (1968) and Schmidt (1976). For homogenous production units, the output, Y, for any individual firm, is equal to $Y = \mu + \varepsilon$, where μ is the maximum output and ε is a normal variable $\left(0, \sigma\right)$ truncated in $\left(-\infty, 0\right]$. The density function of Y is

$$f(y \mid \mu, \sigma) = \frac{2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \; y \le \mu \qquad (22)$$

In this model, for each fixed σ, the parameter μ is nonregular, thus we need to obtain the conditional noninformative prior $\mu \mid \sigma$. We can use the proposition 2 to calculate $\left|E[\partial\ell/\partial\mu]\right|$; from $\partial\ell/\partial\mu = \sigma^{-2}(y-\mu)$ and $E[Y] = \mu - \sqrt{2\pi^{-1}}\sigma$, we obtain that $\left|E[\partial\ell/\partial\mu]\right| = \sqrt{2\pi^{-1}}\sigma^{-1}$, and the conditional noninformative prior is $\pi\left(\mu \mid \sigma\right) \propto 1 \cdot C_1(\sigma)$, where $C_1(\sigma)$ is a function of σ. For each fixed μ, the parameter σ is regular, thus the conditional noninformative prior for $\sigma \mid \mu$ is obtained from to apply Jeffreys' rule. From $-E[\partial^2\ell/\partial\mu^2] = \sigma^{-2}$, we obtain the conditional noninformative prior $\pi\left(\sigma \mid \mu\right) \propto \sigma^{-1} \cdot C_2(\mu)$, where $C_2(\mu)$ is a function of μ. Now from (12), we obtain the join noninformative prior $\pi(\mu,\sigma) \propto \sigma^{-1}$, which is similar to the noninformative prior for a normal model.

The likelihood function for a sample $\mathbf{y} = (y_1,..., y_n)$ is

$$L(\mu,\sigma \mid \mathbf{y}) \propto \sigma^{-n} \exp\left(\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right), \quad y_{(n)} \le \mu, \sigma > 0 \tag{23}$$

then the join posterior distributions is

$$\pi(\mu,\sigma \mid \mathbf{y}) \propto \sigma^{-(n+1)} \exp\left(\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right), \quad y_{(n)} \le \mu, \sigma > 0 \tag{24}$$

From the statistics $\bar{y} = \sum_{i=1}^{n} y_i \big/ n$, $S^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \big/ (n-1)$, and the identity $\sum_{i=1}^{n} (y_i - \mu)^2 = (n-1)S^2 + n(\bar{y} - \mu)^2$, other expression for the join posterior distribution is

$$\pi(\mu,\sigma \mid \mathbf{y}) \propto \sigma^{-(n+1)} \exp\left(-\frac{(n-1)S^2}{2\sigma^2} \right) \exp\left(-\frac{n(\mu - \bar{y})^2}{2\sigma^2} \right), \quad y_{(n)} \le \mu,\ \sigma > 0 \tag{25}$$

Now, the marginal posterior distribution for σ, is

$$\pi(\sigma \mid \mathbf{y}) \propto \sigma^{-n} \exp\left(-\frac{(n-1)S^2}{2\sigma^2} \right) \left(1 - \Phi\left(\frac{\left(y_{(n)} - \bar{y}\right)\sqrt{n}}{\sigma} \right) \right), \sigma > 0 \tag{26}$$

where $\Phi(\bullet)$ is the standard normal distribution function and $y_{(n)}$ is the maximum value of $\mathbf{y} = (y_1,..., y_n)$ in the sample.

To calculate the marginal posterior of the parameter μ, we make the transformation $t = (\mu - \bar{y})\sqrt{n} / S$. Thus,

$$\pi(t \mid \mathbf{y}) \propto \left(1 + \frac{t^2}{n-1} \right)^{-n/2}, \quad t \ge \frac{\left(y_{(n)} - \bar{y}\right)\sqrt{n}}{S} \tag{27}$$

that is, the marginal posterior distribution for μ is a truncated t-Student density distribution with *n-1* degrees of freedom, where t is greater than t_0 and $t_0 = \left(y_{(n)} - \bar{y}\right)\sqrt{n}\Big/s$ is an ancillary statistic.

7. Conclusions.

The original idea of Jeffreys' rule has been generalized to nonregular models. This rule proposes a prior distribution that gives more density of probability to those values of the parameter space where local information is higher. The only difference with the regular models is about the local behavior of information measures. In non-regular models we apply a first order approximation while second order is applied in the regular ones. Therefore, both regular and non-regular models can be integrated in a single method as in (7).

Our proposal of generating one-dimensional noninformative prior distributions in nonregular models is quite satisfactory, since the calculation is simple, they are parameterisation-invariants and the Bayesian intervals have a good frequentist behaviour. If the necessary conditions to obtain the Ghosal's prior hold, then our proposal has the same functional form as the matching prior of Ghosal (1999). If the necessary conditions don't hold, then we have the advantage of being able to obtain the prior distribution applying (6).

In the multi-parameter case, we have only considered prior distributions when all parameters are of interest. If the supposition of (12) is not true, we can obtain a parameterisation to apply this definition.

For each one model, in sections 5 and 6, we have obtained their prior distributions. In these models, the Bayesian method offers a satisfactory solution to the inference problem and also avoids the problem the maximum likelihood estimator has with nonregular models.

References

Aigner, D. and Chu, S. (1968). On estimating the Industry Production Function, *American Economic Review*, 58: 826-839.

Akahira, M. and Takeuchi, K. (1991). A Definition of Information Amount Applicable to Non-Regular Cases, *Journal of Computing and Information*, 2: 71-92.

Akahira, M. and Takeuchi, K. (1995). *Non-regular Statistical Estimation*, New York: Springer-Verlag.

Apóstol, T.M. (1960). *Análisis Matemático,* Madrid: Reverté.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999), *Conditional Specification of Statistical Models,* Springer Series in Statistics, New YorK: Springer-Verlag.

Azzalini, A. (1996). *Statistical Inference Based on the Likelihood*, London: Chapman and Hall.

Basulto, J. (1997). Funciones a Priori Imparciales Unidimensionales, *Estadística Española*, 142: 99-128.

Bernardo, J.M. and Smith, A.F.M. (1994), *Bayesian Theory,* Chichester: John Wiley and Sons.

Coelli, T.J., Rao, P., O'Donnell, C.J. and Battese, G. (2005), *An Introduction to Efficiency and Productivity Analysis, 2nd edition,* New York: Springer.

Ghosal, S. (1999), Probability Matching Priors for Non-Regular Cases, *Biometrika*, 86: 956-964.

Jeffreys, H. (1946), An invariant Form for the Prior Probability in Estimation Problems, *Proceedings of the Royal Society of London, Series A*, 186: 453-461.

Jeffreys, H. (1961), *Theory of Probability, 3rd. edition*, London: Oxford University Press.

Lancaster, T. (1997), Exact Structural Inference in Optimal Job-Search Models, *Journal of Business and Economic Statistics*, 15: 165-179.

Müller, U.U. and Wefelmeyer, W. (2010), Estimation in Nonparametric Resgression with Non-Regular Errors, *Communications in Statictics: Theory and Methods*, 39: 1619-1629.

Nicolau, A. (1993), Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters, *Journal of the Royal Statistical Society, Series B*, 55: 377-390.

Ortega, F. J. y Basulto, J. (2003), Distribuciones a priori unidimensionales en Modelos No Regulares: Medidas de Información, *Estadística Española*, 154: 363-383.

Pitman, E.J. (1979), *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.

Sareen, S. (2003), Reference Bayesian inference in nonregular models, *Journal of Econometrics*, 113: 265-288.

Schmidt, P. (1976), On the Statistical Estimation of Parametric Frontier Production Functions, *Review of Economics and Statistics*, 58: 238-239.

Tchumtchoua, S. and Dey, D.K. (2007), Bayesian Estimation of Stochastic Frontier Models with Multivariate Skew t Error Terms, *Communications in Statictics: Theory and Methods*, 35: 907-916.

Welch, B.L. and Peers, H.W. (1963), On Formulae for Confidence Points Based on Integral of Weithed Likelihoods, *Journal of the Royal Statistical Society, Series B,* 25: 318-329.

Wiper, M.P., Girón, F.J. and Pewsey, A. (2008), Objective Bayesian Inference for the Half-Normal and Half-t Distributions, *Communications in Statictics: Theory and Methods,* 37: 3165-3185.

Appendix: proof of proposition 2.

We assume that the sets $S(\theta)$ are increasing in θ; that this, $a_1^{'}(\theta) < 0$ and $a_2^{'}(\theta) > 0$; the other case is handled by the reparameterisation $\theta \to -\theta$ (it is permitted that one of endpoint is free of θ or may be plus or minus infinity). In this situation, the function $\log f(x|\theta)$ is decreasing in θ, and consequently, we must to prove that $\pi(\theta) \propto -E\left[\partial \log f(x|\theta)/\partial\theta\right]$. In other case, the proof is similar, but the result will be $\pi(\theta) \propto E\left[\partial \log f(x|\theta)/\partial\theta\right]$, because of the function $\log f(x|\theta)$ will be increasing in θ.

For simplicity in the notation, we will use h instead of $\Delta\theta$.

A) For $h > 0$, we need to prove that $\lim\limits_{h\to 0^+} \dfrac{J(\theta,\theta+h)}{h} = -4E\left[\dfrac{\partial \log f(x|\theta)}{\partial\theta}\right]$.

We define the functions $g_\theta(x,h) = \sqrt{f(x|\theta) f(x|\theta+h)}$ and $G_\theta(h) = \int_{a_1(\theta)}^{a_2(\theta)} g_\theta(x,h)dx$.

Since $G_\theta(h)$ is continuous in $[0,\varepsilon]$, we have $\lim\limits_{h\to 0} G_\theta(h) = G_\theta(0) = 1$, and consequently,

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h} = \lim_{h\to 0}\frac{-8\log G_\theta(h)}{h} = -8\lim_{h\to 0}\frac{\partial \log G_\theta(h)}{\partial h} =$$

$$= -8\lim_{h\to 0}\frac{\dfrac{\partial G_\theta(h)}{\partial h}}{G_\theta(h)} = -8\lim_{h\to 0}\frac{\int_{a_1(\theta)}^{a_2(\theta)}\dfrac{\partial g_\theta(x,h)}{\partial h}}{G_\theta(h)} =$$

$$= -8\lim_{h\to 0}\frac{\int_{a_1(\theta)}^{a_2(\theta)}\sqrt{f(x|\theta)}\,\frac{\partial f(x|\theta+h)/\partial h}{2\sqrt{f(x|\theta+h)}}\,dx}{G_\theta(h)} = -4\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x|\theta)}{\partial\theta}dx = -4E\left[\frac{\partial\log f(x|\theta)}{\partial\theta}\right]$$

.

B) For h>0, we need to prove that $\lim_{h\to 0}\frac{J(\theta,\theta-h)}{h} \;=\; -4E\left[\frac{\partial\log f(x|\theta)}{\partial\theta}\right]$.

We define the functions $q_\theta(x,h)=\sqrt{f(x|\theta)f(x|\theta-h)}$ and

$Q_\theta(h)=\int_{a_1(\theta-h)}^{a_2(\theta-h)}q_\theta(x,h)dx$. Since $Q_\theta(h)$ is continuous in $[0,\varepsilon]$, we have

$\lim_{h\to 0}Q_\theta(h)=Q_\theta(0)=1$, and hence

$$\lim_{h\to 0}\frac{J(\theta,\theta-h)}{h}=\lim_{h\to 0}\frac{8\log Q_\theta(h)}{h}=8\lim_{h\to 0}\frac{\partial\log Q_\theta(h)}{\partial h}=8\lim_{h\to 0}\frac{\frac{\partial Q_\theta(h)}{\partial h}}{Q_\theta(h)}=8\left.\frac{\partial Q_\theta(h)}{\partial h}\right|_{h=0}.$$

In this case, since the integration limits of $Q_\theta(h)$ depend on h, we apply the formula of Leibniz (Apóstol, 1960) to calculate $\partial Q_\theta(h)/\partial h$. We have

$$\frac{\partial Q_\theta(h)}{\partial h}=\int_{a_1(\theta-h)}^{a_2(\theta-h)}\sqrt{f(x|\theta)}\frac{\partial\sqrt{f(x|\theta-h)}}{\partial h}dx+a_2'(\theta-h)\sqrt{f(a_2(\theta-h)|\theta)}\sqrt{f(a_2(\theta-h)|\theta-h)}-$$
$$-a_1'(\theta-h)\sqrt{f(a_1(\theta-h)|\theta)}\sqrt{f(a_1(\theta-h)|\theta-h)}$$

thus, we obtain:

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h}=8\left\{\int_{a_1(\theta)}^{a_2(\theta)}\sqrt{f(x|\theta)}\frac{\partial\sqrt{f(x|\theta)}}{\partial\theta}dx+a_2'(\theta)f(a_2(\theta)|\theta)-a_1'(\theta)f(a_1(\theta)|\theta)\right\},$$

and applying again the formula of Leibniz, we obtain:

$$a_2'(\theta)f(a_2(\theta)|\theta)-a_1'(\theta)f(a_1(\theta)|\theta)=-\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x|\theta)}{\partial\theta}dx,$$

and hence

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h}=4\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x|\theta)}{\partial\theta}dx-8\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x|\theta)}{\partial\theta}dx=-4E\left[\frac{\partial\log f(x|\theta)}{\partial\theta}\right].$$

This proves the proposition.

■