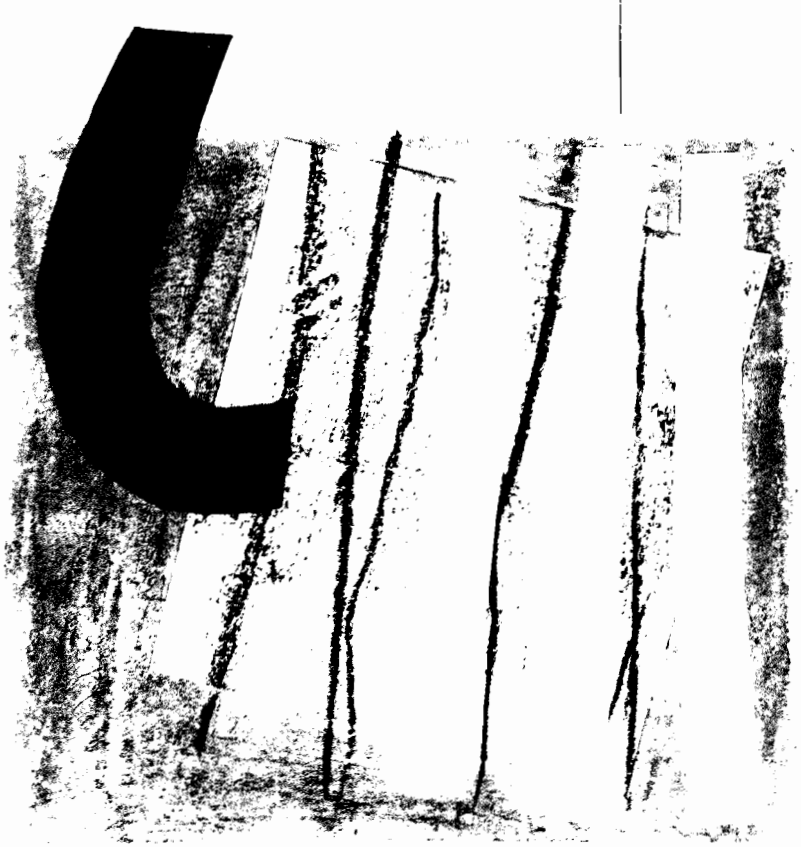
USING BOOTSTRAP TO DERIVE A PRIOR DISTRIBUTION

Pedro Delicado

94 - 07



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 94-07
Statistics and Econometrics Series 05
April 1994

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe, Madrid (Spain)
Fax (341) 624-9849

USING BOOTSTRAP TO DERIVE A PRIOR DISTRIBUTION

Pedro Delicado [1]

Abstract

Constructing a prior distribution when there is no available information is usually an interesting challenge. In this paper, a new method based on bootstrap and nonparametric density estimation ideas is proposed. Its ability to detect and partially correct misspecifications is illustrated with a simulation study.

Key words:
Bayesian analysis, prior distribution, bootstrap, density estimation.

[1] Universidad Carlos III de Madrid, Departamento de Estadística y Econometría. e-mail: pedrod@eco.uc3m.es. I am grateful to Professor Mark Steel for useful comments to a previous version of this paper.

1. INTRODUCTION

In a bayesian context, the researcher sums up his previous information about the parameter of interest in a prior distribution. This procedure is very useful whenever there is available information. Otherwise, we need a way to build a prior. Some methods have been proposed in the literature such as flat priors and Jeffrey's priors. In this paper we propose a different way to build the prior: it uses part of the sample to obtain information about the parameter of interest, e.g., β in Θ. This information is used to construct a density function over the parameter space using bootstrap methods and nonparametric density estimation. This density function becomes the prior distribution of β and it is combined with the remaining observations to complete the bayesian analysis in the usual way. The proposed priors are always proper priors. This is an interesting point in many contexts such as the Bayesian approach to model selection (see Berger and Pericchi, 1993).

In section 2 the proposed method is exposed in detail; a theoretical justification is given in section 3. The last section presents the results of an extensive simulation study and points out the capability to detect missespecifications and to correct them partially. Finally, the appendix gives the proofs of the results presented in section 3.

2. CONSTRUCTING THE PRIOR DENSITY

Let $X_1, \ldots, X_n$ be i.i.d. random variables with density function $P(X \mid \beta)$, β in Θ. Let $T_n : \mathrm{X}^n \rightarrow \Theta$ be an estimator of β, in the classical sense, with density $P(T_n \mid \beta)$. The sample $x_1, \ldots, x_n$ is observed and assumed to be generated with the specific parameter value $\beta = \beta_0$. Let $\hat{\beta}_n = T_n(x_1, \ldots, x_n)$ be the estimated value of β based in our sample. We try to define a density function in Θ which can be used as the prior density of β, $P(\beta)$.

The proposed method is the following:

Step 0 Choose $m | 0 \leq m \leq n$. Choose $x_{i_1}, \ldots, x_{i_m} \subseteq \{x_1, \ldots, x_n\}$, with $i_j \neq i_l$ if $j \neq l$.

(For instance, $i_j = j$).

Step 1 Take B bootstrap resamples of that subsample: $x_1^{*(b)}, \ldots, x_m^{*(b)}$, $b = 1, \ldots, B$, and obtain B bootstrap observations of $\hat{\beta}_m$: $\hat{\beta}_m^{*1}, \ldots, \hat{\beta}_m^{*B}$.

Step 2 Estimate the density $P(T_m \mid \beta_0)$ using any usual nonparametric estimator based in $\{\hat{\beta}_m^{*j}\}_{j=1}^{B}$. Let $P^*(\hat{\beta}_m \mid \beta_0)$ denote this estimation. (This

is only notation. In the next section we study what density is being estimated).

Step 3 Use the density obtained in Step 2 as a prior density of β:

$$P(\beta) \propto P^*(\hat{\beta}_m \mid \beta_0),$$

or equivalently,

$$P_\beta(u) \propto P^*_{\hat{\beta}_m|\beta_0}(u) \text{ for all } u \text{ in } \Theta.$$

Step 4 Calculate the posterior distribution using the following elements:

sample: $X_{m+1}, \ldots, X_n$,
prior of β: the one obtained in Step 3,
likelihood: $\prod_{i=m+1}^{n} P(X_i \mid \beta)$.

Steps 2 and 3, denoted as the DIRECT method, can be considered a first approach to the problem. In *Section 3*, we will see that the following general procedure is, under certain assumptions, theoretically more appropriate.

Suppose that there exists a pivotal quantity Q, depending on the data only through the statistic T_n: $Q(T_n, \beta)$. Assume that the function $Q^b(a) = Q(a, b)$ has derivative (say $(Q^b)'$) and inverse function.

We propose the following method to estimate $P(\hat{\beta}_m \mid \beta)$, for all β in Θ:

Step 2' Estimate nonparametrically the density $P_{Q(T_m,\beta)|\beta}(u)$ (which does not depend on β) from observations $\{Q(\hat{\beta}^{*j}_m, \hat{\beta}_m)\}$. Calculate the value of this estimated density function when $u = Q^\beta(v)$ and multiply it by $|(Q^\beta)'(v)|$. Therefore, we have an estimation of $P_{T_m|\beta}(v)$: $P_{Q(T_m,\beta)|\beta}(Q^\beta(v))|(Q^\beta)'(v)|$. Finally, take $v = \hat{\beta}_m$ and obtain the estimation of

$$P_{T_m|\beta}(\hat{\beta}_m) \text{ for all } \beta \text{ in } \Theta.$$

Let us denote this function of β by $P^*(\hat{\beta}_m \mid \beta)$.

Step 3' Use the density obtained in Step 2' as the prior density of β:

$$P(\beta) \propto P^*(\hat{\beta}_m \mid \beta) \text{ for all } \beta \text{ in } \Theta,$$

that is,

$$P_\beta(u) \propto P_{T_m|u}(\hat{\beta}_m) \text{ for all } u \text{ in } \Theta.$$

This method will be denoted by PIVOTAL method. In Boss and Monahan(1986), the authors study the location parameter estimation. They use $Q(T_m, \beta) = T_m - \beta$ to obtain posterior parameter density functions. Our proposal is essentially similar: calculate a posterior density, although we will use it later on as a prior.

When $m = 0$ we are in the usual bayesian inference with a flat prior, and if $m = n$, we are in a pure bootstrap study. Therefore, the proposed methods can be seen as midpoints between those two extremes.

A different methodology to estimate likelihood functions using bootstrap and nonparametric density estimation is proposed in Davidson et al. (1992). They use nested bootstrap and they can to estimate the likelihood in wider contexts. This methodology could be introduced in the algorithms described here to substitute steps 2 or 2'.

3. THEORETICAL JUSTIFICATION

In bayesian analysis the following steps are equivalent:

(i) Constructing the posterior distribution of β from both the likelihood using the whole sample and a prior $P_0(\beta)$.

(ii) Dividing the sample into two subsamples; proceed as in (i) with one of the subsamples and use the obtained posterior as the prior in the analysis of the other one.

This equivalence is obvious because

$$\begin{aligned} P(\beta \mid X_1, \ldots, X_n) &\propto P(X_1, \ldots, X_n \mid \beta) P_0(\beta) \\ &\propto P(X_{m+1}, \ldots, X_n \mid X_1, \ldots, X_m, \beta) P(X_1, \ldots, X_m \mid \beta) P_0(\beta) \\ &\propto P(X_{m+1}, \ldots, X_n \mid X_1, \ldots, X_m, \beta) P(\beta \mid X_1, \ldots, X_m) \end{aligned}$$

We propose to replace $P(\beta \mid X_1, \ldots, X_m)$ by a nonparametrically estimated density based on a bootstrap sample of an estimator of β.

Observe that if $T_m : \mathrm{X}^m \to \Theta$ is a sufficient statistic for β, then

$$P(\beta \mid X_1, \ldots, X_m) = P(\beta \mid T_m) \propto P(T_m \mid \beta) P_0(\beta).$$

Taking a flat prior $P_0(\beta) \propto 1$, $P(\beta \mid X_1, \ldots, X_m) \propto P(T_m \mid \beta)$, that is,

$$P_{\beta \mid x_1, \ldots, x_m}(u) \propto P_{T_m \mid u}(\hat{\beta}_m) \text{ for all } u \text{ in } \Theta,$$

where $\hat{\beta}_m = T_m(x_1, \ldots, x_n)$.

Therefore, we are interested in the estimation of $P_{T_m|u}(\hat{\beta}_m)$. If nonparametric methods are used, likelihood especification for the first part of the sample is not needed. This fact may be an advantage over the usual bayesian analysis: if the model is missespecified (i.e., $P(X \mid \beta)$ is not the density of the data) the true posterior density of β, given the first observations, may be a better approximation to $P(T_m \mid \beta)$ than to $P(X_1, \ldots, X_m \mid \beta)$ (we omit the constants). This is because T_m may be a sufficient statistic for β in a wide range of models, including the true model. So, the nonparametric part of our proposal partially corrects model missespecifications.

The sample has been randomly divided into two subsamples. There are several reasons for this randomness. It guarantees the independence between the two subsamples which is needed for the equivalence of statements (i) and (ii). Moreover, the first subsample extraction is essentially symmetrical since each possible subsample has the same probability to be selected. True symmetry is hard to get if we do not want to loose independence. A feasible way is to draw all possible first subsamples and take some average of the posteriors as the final posterior. This procedure is computationally very expensive if n is moderate and m is far from 0 and n. We could select the first subsample according to a sensible criterion. So, we would loose the independence between first and second subsamples. However, in the simulation study (see *Section 4*) we have tested one of these procedures: we select the subsample of size m having the same quantiles i/m, $i = 0, \ldots, m$ as the original sample. We are looking for the subsample wich is most similar to the whole sample. We will name these ways to select the first subsample RANDOM and NON-RANDOM extractions, respectively.

The procedures presented in *Section 2* need some assumptions to provide good approximations of $P_{T_m|u}(\hat{\beta}_m)$ as a function of u in Θ.

Let us first examine the method described in Steps 2 and 3. There we use $\{\hat{\beta}_m^{*j}\}_{j=1}^B$ to estimate a density. Then we estimate the following density:

$$P(T_m^* \mid \hat{\beta}_m) = P_{T_m^*|\hat{\beta}_m}(u) \text{ for all } u \text{ in } \Theta,$$

where T_m^* is the bootstrap version of T_m: T_m^* is the statistic T_m applied to $(X_1^*, \ldots, X_m^*)$, i.i.d. with distribution function F_m, the empirical distribution associated to the sample $(x_1, \ldots, x_m)$.

Next reasoning ignores two important problems: the nonparametric density estimation of $P_{T_m^*|\hat{\beta}_m}(u)$ and the bootstrap approximation $P_{T_m^*|\hat{\beta}_m}(u) \approx P_{T_m|\beta_0}(u)$. We suppose known the density $P_{T_m|\beta_0}(u)$, for all u in Θ, where β_0 is the fixed parameter value used to generate the sample. Thus, the approxi-

mation obtained from the DIRECT method is

$$P_{T_m|u}(\hat{\beta}_m) \approx P_{T_m|\beta_0}(u) \text{ for all } u \text{ in } \Theta,$$

where $\hat{\beta}_m = T_m(x_1, \ldots, x_m)$ is the statistic value in our sample (we might have the left side, and the method provides us the right side). Let us denote $P_{T_m|u}(v)$ by $f_u(v)$. We are assuming that

$$f_u(\hat{\beta}_m) = f_{\beta_0}(u) \text{ for all } u, \beta_0 \text{ in } \Theta.$$

Let us also assume that β_0 is near $\hat{\beta}_m$ (i.e., $\hat{\beta}_m \approx \beta_0$), then we must assume that

$$f_u(v) = f_v(u) \text{ for all } u, v \text{ in } \Theta.$$

The next proposition gives some properties of this family of densities f_u defined in Θ and indexed by elements of Θ. Observe the relation between the last assumption and the symmetry around a location parameter.

Proposition 1. *Let u be a location parameter ($f_u(v) = f_{u+k}(v+k)$). Then the following are equivalent:*

$(a) f_u(v) = f_v(u)$ for all u, v in Θ.

$(b) f_u$ is symmetric around u for all u in Θ.

If, moreover, we assume 0 is in Θ then (a) and (b) are equivalent to

$(c) f_0(u) = f_u(0)$ for all u in Θ.

As a summary, the next assumptions are needed in order to apply the DIRECT method proposed in Steps 2 and 3:

a. $T_m \mid \beta$ is a random variable which takes values in Θ and verifies $P_{T_m|\beta}(u) = P_{T_m|u}(\beta)$ for all u, β in Θ.

b. T_m is an estimator such that $P_{T_m|u}(\hat{\beta}_m) \approx P_{T_m|u}(\beta_0)$ for all u, β_0 in Θ, when $\hat{\beta}_m$ is obtained applying T_m to $x_1, \ldots, x_m$ i.i.d. with distribution $P(X \mid \beta_0)$.

c. The conditional density of the bootstrap estimator $T_m^* \mid \hat{\beta}_m$ is near the conditional density of $T_m \mid \beta_0$ (i.e., the bootstrap "works" in this case).

d. The nonparametric estimation of the bootstrap estimator density is near its true density.

e. The statistic T_m is sufficient for β.

Sufficient conditions for c can be found in Bickel and Freedman (1981). There are several references about asymptotic properties of nonparametric estimation in Silverman (1986). *Proposition 1* gives a sufficient condition for a: to have a location parameter and a symmetric density around it. Assumption b is more difficult to be verified.

We examine now our second proposal to estimate the function $P_{T_m|\beta}(\hat{\beta}_m)$, β in Θ. In Steps 2' and 3', observations $\{Q(\hat{\beta}_m^{*j}, \hat{\beta}_m =)\}_{j=1}^{B}$ are used to estimate the underlying density function: $P_{Q(T_m^*,\hat{\beta}_m)|\hat{\beta}_m}(u)$.

As before, we clear away the problems derived from the density estimation and the bootstrap approximation. We can consider for theoretical reasoning that the estimator of $P_{Q(T_m^*,\hat{\beta}_m)|\hat{\beta}_m}(u)$ agree with the density obtained if we substitute the bootstrap terms by the population terms: $P_{Q(T_m,\beta_0)|\beta_0}(u)$. Since Q is a pivotal quantity, this density is equal to $P_{Q(T_m,\beta)|\beta}$ for all β in Θ. This is just the density we are looking for in Step 2'.

Then. the following assumptions are needed to apply the general method:

a'. $Q(T_m, \beta)$ is a pivotal quantity.

b'. The bootstrap "works" in the following sense: $P_{Q(T_m^*,\hat{\beta}_m)|\hat{\beta}_m} \approx P_{Q(T_m,\beta_0)|\beta_0}$.

c'. We can obtain a good estimation of the density $P_{Q(T_m^*,\hat{\beta}_m)|\hat{\beta}_m}$ by nonparametric methods.

d'. The statistic T_m is sufficient for β.

To apply DIRECT method in the location parameter case we need assume symmetry around β (by *Proposition 1*, hypothesis a is equivalent to symmetry). The PIVOTAL method does not need symmetric distributions. In this sense we can say that the general method is theoretically more appropriate than the first proposal in the location problem.

To finish this section, we will see in a particular case the relationship between the two density estimators proposed. Let $\hat{f}_D$ and $\hat{f}_P$ the estimated densities by DIRECT and PIVOTAL procedures, respectively.

Proposition 2. *Let β a location parameter. For kernel estimators of the density, we have*

$$\hat{f}_D(u) = \hat{f}_P(2\hat{\beta}_m - u) \quad \textit{for all } u \textit{ in } \Theta.$$

$$\hat{f}_P(u) = \hat{f}_D(2\hat{\beta}_m - u) \quad \textit{for all } u \textit{ in } \Theta.$$

Moreover, if one estimator is symmetric around $\hat{\beta}_m$ then both estimators are the same.

4. SIMULATION STUDY

In the present simulation study we evaluate the two proposed methods. The involved density estimations have been constructed using kernel estimators. We have used *CURVDAT* routines (see STATCOM, 1990). The bandwith selection came from plug-in method. The kernels orders were 2 and 6. The results were very similar using both orders, so we will only refer to the first one. Numerical integration was evaluated by Simpson method.

We work with a location parameter β. The T_n statistics used through this section are sufficient statistics in each case. We assume a certain likelihood for the data: $X \sim N(\beta, \sigma), \sigma = 1$, or $X - \beta + 1/\lambda \sim Exp(\lambda), \lambda = 1$. In the normal case we take out data from normal distributions with the same mean and standard deviation $\sigma = .8(.05)1.2$. The values $\sigma = .6, .7, 1.3, 1.4$ were also examined. In the shifted exponential case, we draw data with the same mean as in the nominal model and $\lambda = .96(.01)1.04$. We adjust the simulated cases to the required hypotheses as much as possible.

The range of models is different in the normal and the exponential cases because in the second one the probability is very concentrated in the right neighborhood of the point $\beta - 1/\lambda$, so slight changes in λ lead to significant variations in the probability mass distribution.

Two sample sizes are used: $n = 40$ and $n = 100$. The first subsample size m is taken in the following way: with $n = 40$, $m = 0, 10, 20, 30, 40$, with $n = 100$, $m = 0, 20, 40, 60, 80, 100$. The number of bootstrap replications of the first subsample is $B = 400$ when $n = 40$, and $B = 1000$ when $n = 100$. Finally, 200 replications of each case have been made. The values we will show are the mean values for all replications.

Our interest is concentrated in the L_1 distance between two posterior densities of β: the first one is obtained under the supposed likelihood using DIRECT and/or PIVOTAL methods, and the second one is obtained using a flat prior and the true likelihood. We hope this two posterior densities are close if $\sigma = 1$ or $\lambda = 1$ and, in any other case, their distances decrease as m increases.

In the *normal* case we use the sample mean as statistic T_n. It is a sufficient statistic. Moreover, the central limit theorem guarantees that the bootstrap works in this case. In *Table 1* we can see the results for $\sigma = 1$ (i.e., the nominal model is the true model). We build the prior distribution of location β by the two methods given in *Section 2*: DIRECT and PIVOTAL. No significant differences are found between them. This is true also for all the considered values of σ. Then, from now on we only show the outcomes for the PIVOTAL method in the normal case.

Subsampling:		DIRECT method		PIVOTAL method	
		RANDOM	NON-RANDOM	RANDOM	NON-RANDOM
	$m = 10$	.13614336	.06800597	.14108823	.06178522
$n = 40$	20	.11760730	.07819342	.12468031	.07903041
	30	.11991236	.09990099	.11858867	.09867791
	40	.11581332	.11274376	.11637762	.11227307
	$m = 20$	.09569074	.04334227	.09397230	.03731273
$n = 100$	40	.09182197	.04977523	.08714055	.04951791
	60	.08614807	.06298065	.08339045	.05985488
	80	.08380556	.07173784	.08055644	.07310174
	100	.08096730	.08213384	.07820505	.07991870

Table 1: *Standard normal distribution: L_1 distances between posterior built with a flat prior and the posteriors obtained by the proposed methods. For $m = 0$ this distance is always 0.*

Always with $\sigma = 1$, we examine before the RANDOM way to choose the first subsample. The proposed ways to build the prior lead to posteriors that are not very far from the true posterior in L_1 sense. Moreover the results are quite uniform in m (approximately .12 if $n = 40$ and .09 if $n = 100$).

The NON-RANDOM way of drawing the first subsample gives better results. For $n = 40$ and $m = 10$ (resp., $n = 100$ and $m = 20$) the L_1 distances are reduced to .6 (resp., .04). The L_1 distances increase with m and they are even smaller than the obtained with RANDOM selection.

Let us leave the true model ($\sigma = 1$). In *Figures 1, 2* and *3* we can see a summary of outcomes for $n = 100$ and $\sigma = .8(.05)1.2$. The cases for $\sigma = .6..7, 1.3, 1.4$ have also been carried out. The results for $n = 40$ are essentially similar, but the distance between true and supposed σ should be larger with $n = 100$ to observe the advantages of a specific value of m. For instance, if $\sigma = .9$ $m = n$ is better than $m = 0$ for $n = 100$. If $n = 40$ this is false and $\sigma = .85$ is needed to observe $m = n$ beat $m = 0$.

For the RANDOMLY selected subsample (see *Figure 1*) the most important conclusions of the experiment are the following: pure bootstrap procedure ($m = n$) is uniformly better than mixtures ($0 < m < n$); the L_1 distance to the true posterior is constant in σ for pure bootstrap; for $n = 100$ when $|\sigma - 1| \leq .1$ pure bootstrap ($m = n$) is better than flat prior ($m = 0$); the non-extreme cases ($0 < m < n$) are also better than $m = 0$ for some σ. For instance, for $n = 100$ the value $m = 60$ is better for $\sigma \leq .85$ or $\sigma \geq 1.2$, and

$m = 20$ is better for $\sigma \leq .6$ (this last result is not included in *Figure 1*).

The results improve if we use a NON-RANDOMLY selected first subsample (see *Figure 2*). The non-extreme cases ($0 < n < m$) are better than pure bootstrap ($m = n$) for values of σ near 1, the true value. The smaller m is, the better results for these values of σ are. The opposite happens when σ is far from 1: big m are better. We can say that NON-RANDOM selection is better than RANDOM for all m (of course, $0 < m < n$) and for all σ.

In *Figures 2.b* and *2.c* we can see L_1 distance as a function of m. Each curve corresponds to a value of σ. For $\sigma < 1$ we have increasing functions of m (σ near 1) and decreasing functions of m (σ far from 1). If $\sigma > 1$ the functions have a minimum for some value of m. This optimum value of m is bigger if σ is farther from 1. Then if the data are farther from the supposed likelihood we need a bigger first subsample size to improve the results. The *Figures 1.b* and *1.c* show analogous functions in the RANDOM first subsample case. There, all functions have their minimum in $m = n$ (pure bootstrap) or $m = 0$.

The capability to detect deviation from the nominal model is other useful utility of the methodology proposed here. We can calculate the distance between the generated posteriors and the obtained posteriors if we take the nominal likelihood, the whole sample and a flat prior. These distances will be different for different true models and it is hopped that the more different the models are the bigger the distances are.

Only known elements are involved in these calculations. It is possible to reproduce them from the observed data and from simulated data under the nominal model. Two functions of m are then obtained and their comparison may help to decide if the proposed model is appropriate to the data. The really different profiles of L_1 distance as a function of m for several values of parameter σ are shown in *Figure 3*. We can see that both RANDOM and NON-RANDOM selection detect well the cases with a σ far from 1. The values of $\sigma > 1$ are better detected.

In the shifted exponential case we use the estimator $T_n = \min_i X_i + 1/\lambda_0$ (λ_0 is the supposed value for λ and will be 1 in our study). The nonparametric bootstrap does not reproduce the density of T_n. It is known that this is an example of bootstrap failure (see Efron, 1993, page 81). This is the reason why a parametric bootstrap was used. The value of β was estimated assuming $\lambda = \lambda_0$ and then several subsamples of size m were drawn under the model $X_i \sim Exp(\lambda_0) - 1/\lambda_0 + \hat{\beta}_m$. A bootstrap value $\hat{\beta}^*_m$ was obtained for each new subsample.

Only PIVOTAL procedure is considered because the hard asymmetry of this distribution provokes that the likelihood function and the bootstrap prior built by DIRECT method have not supports intersection.

The results are less satisfactory than in the normal case. For $\lambda > 1$, the L_1 distances between the true posteriors (with flat prior) and the one obtained using bootstrap are almost constant for $0 \leq m \leq n$. Slight improvements for $m > 0$ are found if $\lambda < 1$. The advantage for $m > 0$ is clearer in the NON-RANDOM selection case. There, $m = 10$ (resp., $m = 20$) is recommended for $n = 40$ (resp., $n = 100$). In *Figure 4* are shown the L_1 distance values for $n = 100$ and NON-RANDOM selection. The detection of missespecifications performs also badly in the exponential case.

The main reason for this performance in the exponential case is the next one: the statistic $T_n = \min_i X_i + 1/\lambda_0$, is not appropriated for $\lambda \neq \lambda_0$. The differences between the densities of $\min_i X_i + 1/\lambda_0$ and $\min_i X_i + 1/\lambda$ are very important between $1/\lambda_0$ and $1/\lambda$ and the probability is highly concentrated there. But this is the best statistic we can use if we assume the right model is for $\lambda = \lambda_0$ and we do not know in which sense the data are far from it.

We have shown two examples of performance of the proposed techniques. In the first one (*normal distributions*) the results are highly satisfactory. The *exponential* case is not so satisfactory because only slight improvements are achieved using bootstrap. In both cases the supremacy of NON-RANDOM selection for the first subsample had been stated.

5. APPENDIX

Proof of Proposition 1:

$(b) \Longrightarrow (a)$:

$$f_v(u) = f_v(v + u - v) = f_v(v - u + v) = f_u(v)$$

The second equality is true because the symmetry and the third one because we have a location parameter.

$(a) \Longrightarrow (b)$:

$$f_u(u - v) = f_{u-v}(u) = f_u(u + v)$$

The first equality follows due to assumption (a) and the second one because we have a location parameter.

If 0 is in Θ:

$(a) \Longrightarrow (c)$: obvious.

$(c) \Longrightarrow (a)$:

$$f_v(u) = f_{v-u}(0) = f_0(v - u) = f_u(v)$$

The first and last equalities are true because u is a location parameter, and the central equality is obtained using the assumption (c). ■

Proof of Proposition 2:

The kernel estimator of $P(T_m^* \mid \hat{\beta}_m)$ given by the DIRECT method is

$$\hat{f}_D(u) = \frac{1}{Bh_B} \sum_{j=1}^{B} K\left(\frac{u - \hat{\beta}_m^{*j}}{h_B}\right).$$

The density estimator provided by the PIVOTAL method is

$$\frac{1}{Bh_B} \sum_{j=1}^{B} K\left(\frac{v - (\hat{\beta}_m^{*j} - \hat{\beta}_m)}{h_B}\right),$$

what is calculated in the points with the pattern $v = x - u$, with $x = \hat{\beta}_m$, giving the following function of u:

$$\hat{f}_P(u) = \frac{1}{Bh_B} \sum_{j=1}^{B} K\left(\frac{2\hat{\beta}_m - u - \hat{\beta}_m^{*j}}{h_B}\right).$$

Therefore,

$$\hat{f}_D(u) = \hat{f}_P(2\hat{\beta}_m - u) \text{ for all } u \text{ in } \Theta,$$
$$\hat{f}_P(u) = \hat{f}_D(2\hat{\beta}_m - u) \text{ for all } u \text{ in } \Theta.$$

Moreover, if we assume $\hat{f}_D(u)$ is symmetric respect to $\hat{\beta}_m$, then

$$\hat{f}_D(u) = \hat{f}_D(\hat{\beta}_m + (u - \hat{\beta}_m)) = \hat{f}_D(\hat{\beta}_m - (u - \hat{\beta}_m)) =$$

$$= \hat{f}_D(2\hat{\beta}_m - u) = \hat{f}_P(u) \text{ for all } u \text{ in } \Theta,$$

and similarly if $\hat{f}_P(u)$ is symmetric respect to $\hat{\beta}_m$. ■

6. BIBLIOGRAPHY

Berger, J. O. and Pericchi, L. R. (1993). The intrinsic Bayes factor for model selection and prediction. Technical report 93-43C, Department of Statistics, Purdue University.

Bickel, P. J. and Freedman, D. A. (1981). "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196-1217.

Boss, D. D. and Monahan, J. F. (1986). "Bootstrap Methods using prior information," *Biometrika*, 73, 77-83.

Davidson,A.C., Hinkley,D.V., Worton,B.J. (1992). "Bootstrap likelihoods," *Biometrika* 79, 113-130.

Efron (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

STATCOM (1990). Nonparametric smoothing, differentiation and density estimation. Curvdat 90.1 User's manual. Institut für Statistik und Computing. Walter Köhler. Am Mühlrain 24B. D-6903 Neckargemünd

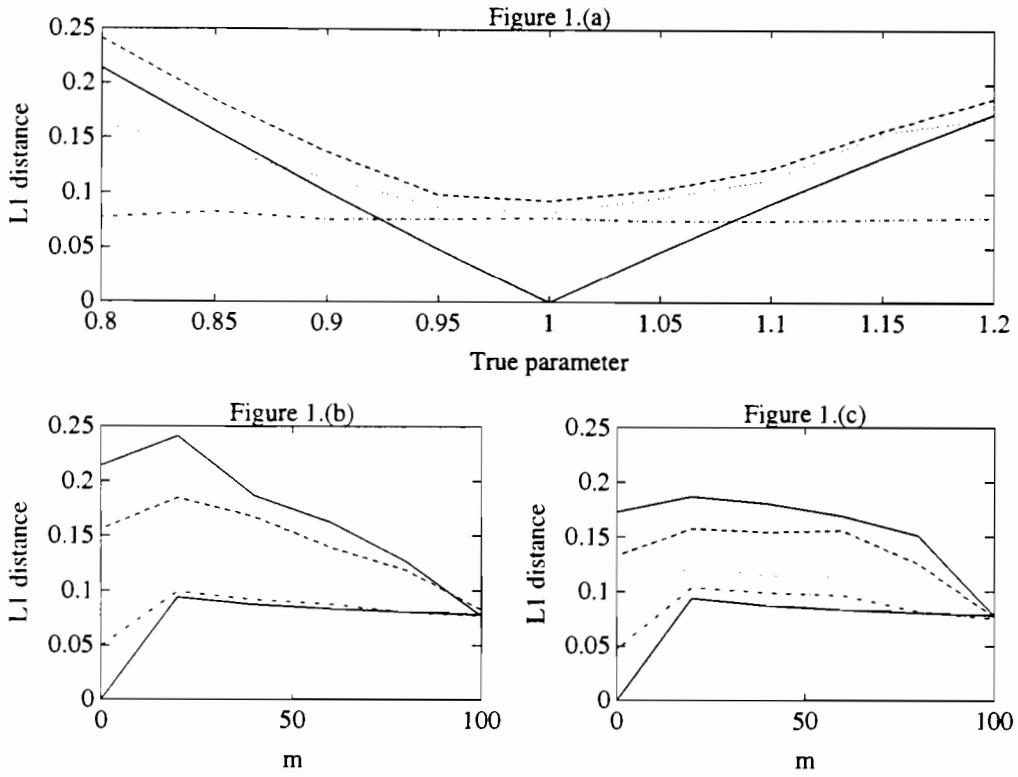


Figure 1: *Normal case*, $n = 100$, PIVOTAL *method*, RANDOM *selection.*

(a) L_1 *distance between built posteriors (assuming* $\sigma = 1$ *and using different sizes* m *for the first subsample) and true posterior (using true* σ *and flat prior) versus true values of* σ.
$m = 0$: ———; $m = 20$: – – –; $m = 60$: · · · · ·; $m = 100$: – · – · –.
(b) L_1 *distances versus size* m*, for different values of* $\sigma \leq 1$.
$\sigma = 1.0$: ——— *(starting at 0)*; $\sigma = .95$: – · – · –; $\sigma = .90$: · · · · ·; $\sigma = .85$: – – –; $\sigma = .80$: ——— *(starting at .214).*
(c) L_1 *distances versus size* m*, for different values of* $\sigma \geq 1$.
$\sigma = 1.0$: ——— *(starting at 0)*; $\sigma = 1.05$: – · – · –; $\sigma = 1.10$: · · · · ·; $\sigma = 1.15$: – – –; $\sigma = 1.20$: ——— *(starting at .173).*

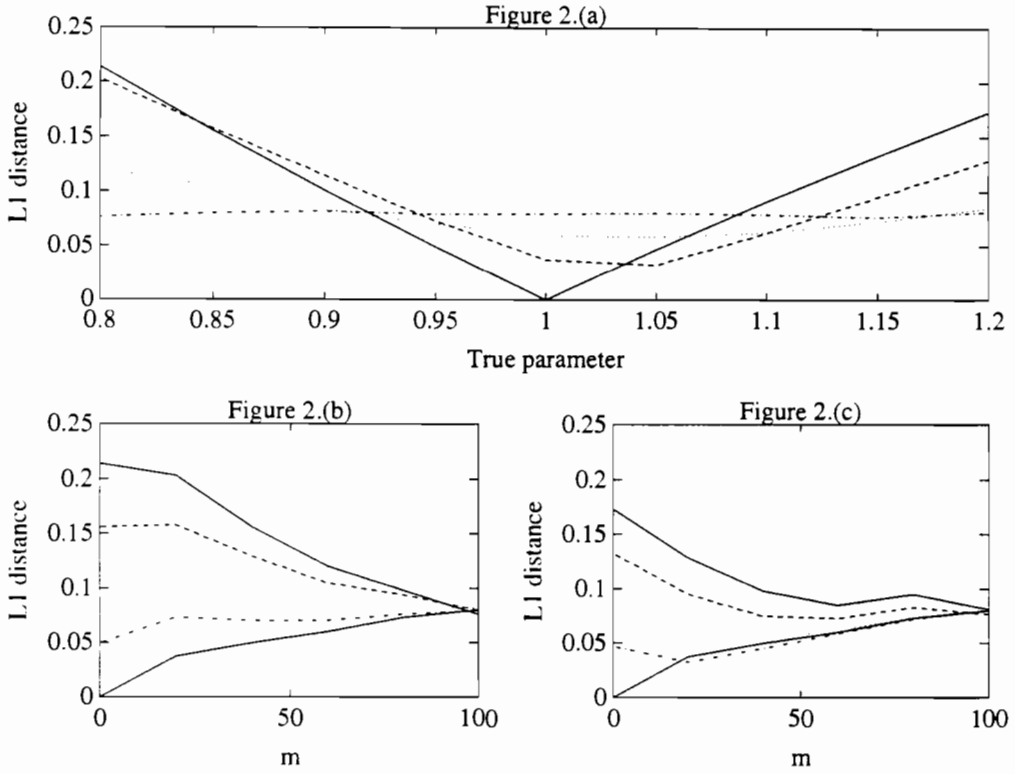


Figure 2: *Normal case*, $n = 100$, PIVOTAL *method*, NON-RANDOM *selection*.

(a) L_1 *distance between built posteriors (assuming* $\sigma = 1$ *and using different sizes* m *for the first subsample) and true posterior (using true* σ *and flat prior) versus true values of* σ.
$m = 0$: ———; $m = 20$: − − −; $m = 60$: · · · · ·; $m = 100$: − · − · −.
(b) L_1 *distances versus size* m, *for different values of* $\sigma \leq 1$.
$\sigma = 1.0$: ——— *(starting at 0)*; $\sigma = .95$: − · − · −; $\sigma = .90$: · · · · ·; $\sigma = .85$: − − −; $\sigma = .80$: ——— *(starting at .214)*.
(c) L_1 *distances versus size* m, *for different values of* $\sigma \geq 1$.
$\sigma = 1.0$: ——— *(starting at 0)*; $\sigma = 1.05$: − · − · −; $\sigma = 1.10$: · · · · ·; $\sigma = 1.15$: − − −; $\sigma = 1.20$: ——— *(starting at .173)*.

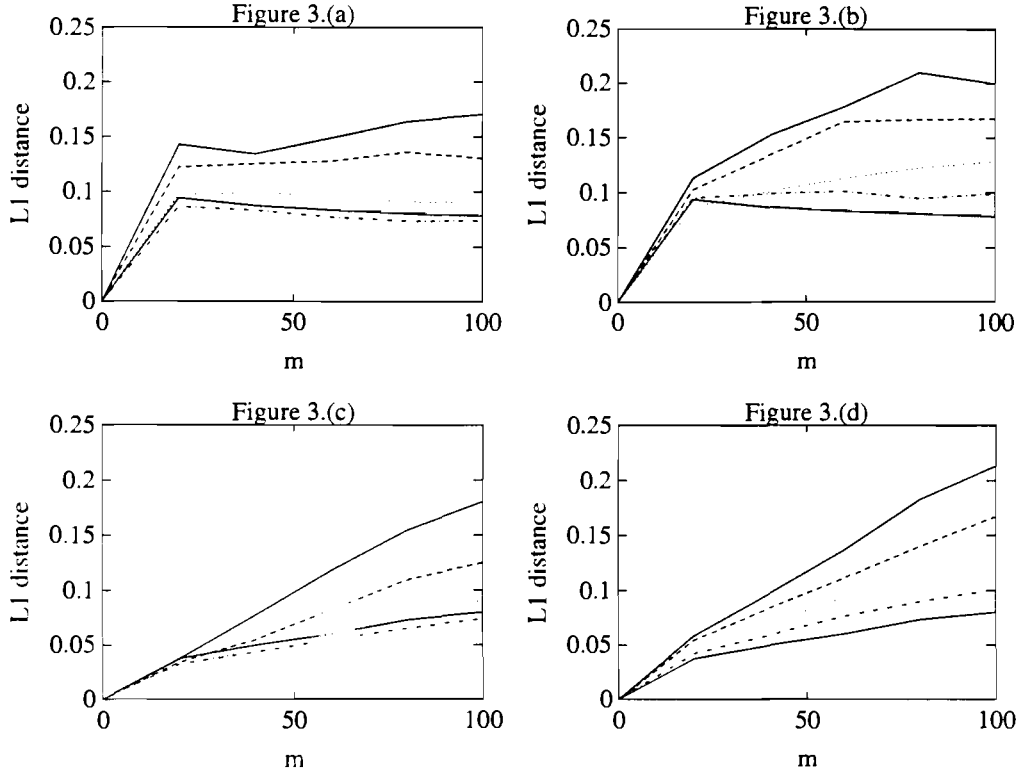


Figure 3: *Normal case,* $n = 100$, PIVOTAL *method.*

(a) RANDOM *selection and different values of* $\sigma \leq 1$*:* L_1 *distance between built posteriors (assuming* $\sigma = 1$ *and using different sizes* m *for the first subsample) and false posterior (using* $\sigma = 1$*, the whole sample and flat prior) versus size* m*.*
$\sigma = 1.0$*:* ——— *(finishing at .078);* $\sigma = .95$*:* $- \cdot - \cdot -$*;* $\sigma = .90$*:* $\cdot \cdot \cdot \cdot \cdot$*;*
$\sigma = .85$*:* $- \; - \; -$*;* $\sigma = .80$*:* ——— *(finishing at .170).*
(b) RANDOM *selection and different values of* $\sigma \geq 1$*:* L_1 *distance between built posteriors (assuming* $\sigma = 1$ *and using different sizes* m *for the first subsample) and false posterior (using* $\sigma = 1$*, the whole sample and flat prior) versus size* m*.*
$\sigma = 1.0$*:* ——— *(finishing at .078);* $\sigma = 1.05$*:* $- \cdot - \cdot -$*;* $\sigma = 1.10$*:* $\cdot \cdot \cdot \cdot \cdot$*;* $\sigma = 1.15$*:* $- \; - \; -$*;* $\sigma = 1.20$*:* ——— *(finishing at .203).*
(c) and (d) are equivalent to (a) and (b) for NON-RANDOM *selection.*

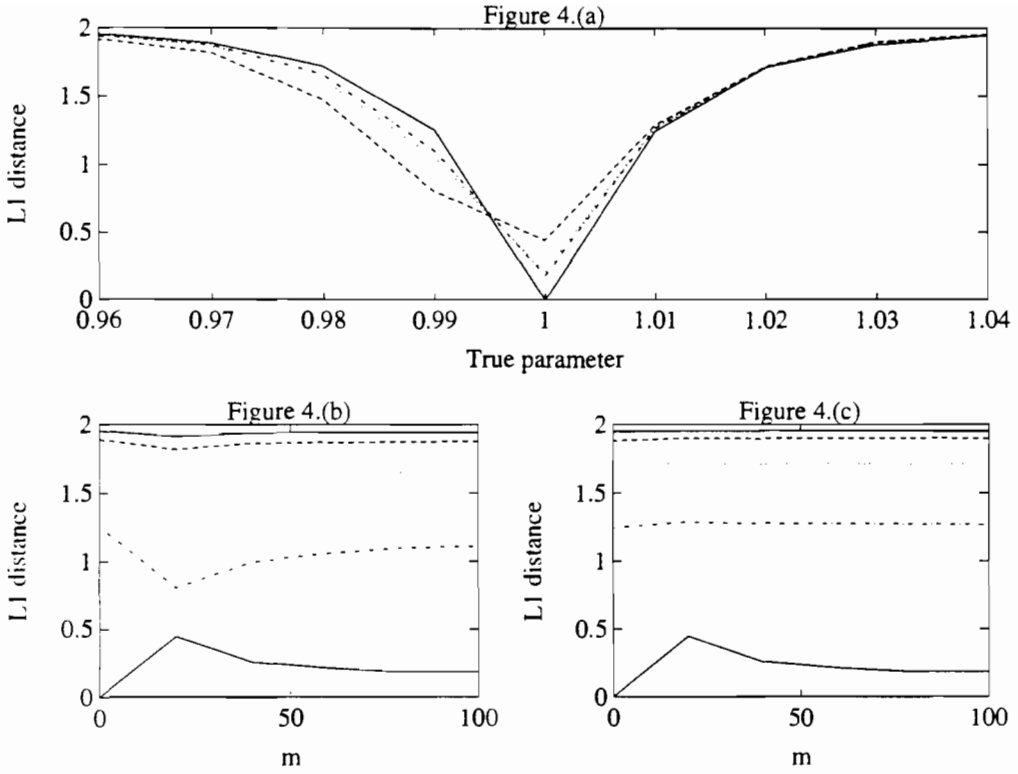


Figure 4: *Exponential case,* $n = 100$, PIVOTAL *method,* NON-RANDOM *selection.*

(a) L_1 *distance between built posteriors (assuming* $\lambda = 1$ *and using different sizes* m *for the first subsample) and true posterior (using true* λ *and flat prior) versus true values of* λ*.*
$m = 0$: ———; $m = 20$: $-\ -\ -$; $m = 60$: $\cdots\cdots$; $m = 100$: $-\cdot-\cdot-$.
(b) L_1 *distances versus size* m*, for different values of* $\lambda \leq 1$*.*
$\lambda = 1.0$: ——— *(starting at 0);* $\lambda = .99$: $-\cdot-\cdot-$; $\lambda = .98$: $\cdots\cdots$; $\lambda = .97$: $-\ -\ -$; $\lambda = .96$: ——— *(starting at 1.95).*
(c) L_1 *distances versus size* m*, for different values of* $\lambda \geq 1$*.*
$\lambda = 1.00$: ——— *(starting at 0);* $\lambda = 1.01$: $-\cdot-\cdot-$; $\lambda = 1.02$: $\cdots\cdots$; $\lambda = 1.03$: $-\ -\ -$; $\lambda = 1.04$: ——— *(starting at 1.95).*