REPORTING UNCERTAINTY BY SPLINE FUNCTION APPROXIMATION OF LOG-LIKELIHOOD

by

Ahmet Sezer

Submitted to the Graduate Faculty of

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

FACULTY OF ART AND SCIENCES

This dissertation was presented

by

Ahmet Sezer

It was defended on

September, 2006

and approved by

Professor Leon J. Gleser

Professor Satish Iyengar

Professor Henry W. Block

Professor John Wilson

Dissertation Director: Professor Leon J. Gleser

REPORTING UNCERTAINTY BY SPLINE FUNCTION APPROXIMATION OF LOG-LIKELIHOOD

Ahmet Sezer, PhD

University of Pittsburgh, 2006

Reporting uncertainty is one of the most important tasks in any statistical paradigm. Likelihood functions from independent studies can be easily combined, and the combined likelihood function serves as a meaningful indication of the support the observed data give to the various parameter values. This fact has led us to suggest using the likelihood function as a summary of post-data uncertainty concerning the parameter.

However, a serious difficulty arises because likelihood functions may not be expressible in a compact, easily-understood mathematical form suitable for communication or publication. To overcome this difficulty, we propose to approximate log-likelihood functions by using piecewise polynomials governed by a minimal number of parameters. Our goal is to find the function of the parameter(s) that approximates the log-likelihood function with the minimum integrated (square) error over the parameter space. We achieve several things by approximating the log-likelihood; first, we significantly reduce the numerical difficulty associated with finding the maximum likelihood estimator. Second, in order to be able to combine the likelihoods that come from independent studies, it is important that the approximation of the log- likelihood should depend only upon a few parameters so that the results can be communicated compactly. By the simulation studies we compared natural cubic spline approximation with the conventional modified likelihood methods in terms of coverage probability and interval length of highest density region obtained from the likelihood and the mean squared error of the maximum likelihood estimator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Reporting uncertainty is one of the most important tasks in any statistical paradigm. In both the frequentest and Bayesian approaches, likelihoods play an important role in reporting uncertainty about parameters.

P-values and confidence intervals are commonly used to report uncertainty in statistical methodology. However, p-values and confidence intervals are often misinterpreted and misrepresented. Indeed, confidence intervals tend to hide the fact that some parameter values are more supportable by the data than others. Further, it has often been pointed out that one cannot directly combine confidence intervals from study to study in meta-analysis. In contrast, the likelihood functions from independent studies can be easily combined, and the combined likelihood function serves as a meaningful indication of the support the observed data give to the various parameter values. This fact has led many scientists to suggest using the likelihood function as a summary of post-data uncertainty concerning the parameter. Indeed, likelihood functions have several desired properties . They are objective, in that they depend only on the agreed-upon model and the data. They are also flexible, allowing us to combine information about competing models across studies.

However, a serious difficulty arises because likelihood functions may not be expressible in a compact, easily-understood mathematical form suitable for communication or publication. For example, likelihood functions in mixture models may only be computable for individual values of the parameters and otherwise cannot be given in "closed form". To overcome this difficulty, we propose to approximate log- likelihood functions by using piecewise polynomials governed by a minimal number of parameters. Our goal will be to find the function of the parameter(s) that approximates the log-likelihood function with the minimum integrated (square) error over the parameter space.

For fixed data $\mathbf{y}$, the likelihood is a function only of the parameter. We think of approximating the log-likelihood using an additive model of the form:

$$\log \prod_{i=1}^{n} f(y;\theta) = \sum_{j=1}^{J} \beta_j B_j(\theta) + \epsilon \tag{1.1}$$

We would like to achieve several things by approximating the log-likelihood; first, we significantly reduce the numerical difficulty associated with finding the maximum likelihood estimator. Second, in order to be able to combine the likelihoods that come from independent studies, it is important that the approximation of the log-likelihood should depend only upon a few parameters so that the results can be communicated compactly.

In large samples, central limit and saddle-point approximation theory suggest that spline functions having just two or three knots may be used to approximate the log- likelihood function for a single parameter, with the approximating polynomial being quadratic inside and linear outside of the knots. Natural cubic splines, which are piecewise cubic inside of the extreme knots and linear outside of these knots, are frequently used in non-parametric regression and data mining. These two facts suggest the use of natural cubic splines to approximate the log-likelihood functions.

A non-trivial problem is how to extend this theory to approximate the log-likelihood for two or more parameters. Of course, if one of the two parameters is viewed as a "nuisance parameter" of little interest, we can reduce to the one- parameter model after the elimination of that nuisance parameter by a variety of methods proposed in the literature. If we include both parameters in the likelihood to be approximated, the approximation becomes more complicated.

In Chapter 1, we will introduce methods that have been proposed for eliminating nuisance parameters from the likelihood: Integrated, conditional , marginal and profile likelihood. We compare the accuracy of these methods for small samples. In Chapter 2, piecewise polynomials (natural cubic spline and B-spline) are introduced to approximate log-likelihoods for one parameter. Chapter 3 discusses evaluation of natural cubic spline approximations and graphical presentation of those approximations as raindrop plots. In Chapter 4, we will show how to approximate the log-likelihood when there are two parameters in the model, and Chapter 5 will consist of approximation of posterior density functions and evaluation of those approximations. Finally, Chapter 6 will have a discussion of findings and possible future research.

2.0 LIKELIHOOD METHODS IN THE PRESENCE OF NUISANCE PARAMETERS

In parametric statistical inference, the goal is to make inference about the unknown value of a parameter θ. Suppose the observed data $y = (y_1, y_2, \ldots, y_n)$ can be modeled as independent observed values of a random variable Y distributed according to the density $f(y;\theta)$, $\theta \in \Omega$. The likelihood function of θ based on the data y is given by;

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta). \tag{2.1}$$

If θ_1 and θ_2 are two possible values of θ and such that $L(\theta_1) > L(\theta_2)$, then the probability of observing the data y is greater when the true parameter value is θ_1 than when the true parameter value is θ_2. The maximum likelihood estimator(MLE) is the parameter value for which the data are most likely. Intuitively, the MLE is a reasonable choice for an estimator. Under certain regularity conditions MLE have several important properties when the sample size is large:

1. Consistency: As the sample size increases, the MLE converges in probability to the true parameter value, e.g., $\hat{\theta} \to \theta$. (Wald,1949).

2. Invariance: If g(θ) is any function of the unknown parameters of the distribution, then the MLE of g(θ) is $g\left(\hat{\theta}\right)$.

3. Asymptotic normality and efficiency: As the sample size increases, the sampling distribution of the MLE converges to a normal distribution with mean θ and variance equal to $\frac{1}{nI(\theta)}$ where $I(\theta)$ is Fisher's information for one observation. A proof of the asymptotic normality of the maximum likelihood estimator is given in Cramer (1946). The result that MLE's have the minimum possible asymptotic variance has been studied by Kalianpur and Rao (1955) and Bahadur (1964). The choice between using the observed and expected information for estimating Fisher's information of the maximum likelihood estimate has been considered in Efron and Hinkley (1978).

Because the MLE has many desirable properties as mentioned above, it is a reasonable choice for a point estimator of the parameter θ. Finding the MLE is relatively straightforward in one parameter models; one can use calculus or simply plot the likelihood function. In general, maximum likelihood estimation involves several steps;

1) Construct the likelihood function.

2) Simplify the likelihood function and take its logarithm.

3) Take the partial derivative of the log-likelihood with respect to each parameter and set it to zero and solve to find the estimation of the parameter.

However, we may have models with more than one parameter where various components of the parameter vector have different levels of interest. Consider a model parameterized by a two-dimensional vector of parameters (θ, λ), where θ is the parameter of interest (called a structural parameter) and λ is a nuisance parameter. For models with only one parameter, θ, inference may be based on the likelihood function $\mathrm{L}(\theta)$. However, when a nuisance parameter is present, it is not possible to use the likelihood function to directly compare different values of θ. Indeed, the greater the dimension of the nuisance parameter, the greater its potential effect on the conclusions regarding the parameter of interest (Berger, Liseo and Wolpert, 1998).

In order to draw inferences regarding the parameter of interest in such a two-parameter model, we must deal with the nuisance parameter. General likelihood-based methods have been proposed for the elimination of a nuisance parameter to focus on the structural parameter only. Some of these methods are: profile likelihood, marginal likelihood, conditional likelihood and integrated likelihood. Before discussing the relationships among these methods, we will introduce each method briefly.

2.1 PROFILE LIKELIHOOD

The simplest likelihood approach to eliminating nuisance parameters is to replace them with their maximum likelihood estimates, leading to profile likelihood. Suppose that the model can be identified by parameters (θ, λ), where θ is a parameter of interest and λ is a nuisance parameter. To eliminate λ from the likelihood, it is replaced by the maximum likelihood estimator of λ, keeping θ fixed. The resulting likelihood function; $L_p(\theta) = \sup_\lambda L(\theta, \lambda)$ is called the profile likelihood function and $\log L_p(\theta)$ is called the profile log-likelihood function.

$L_p(\theta)$ can be used as if it were an ordinary likelihood to produce asymptotically correct inference about θ. It is the simplest method, but it does not take into account the uncertainty due to lack of knowledge of the nuisance parameter and can be misleading in both precision and location (Severini,1998b). In large samples, replacing λ by its maximum likelihood estimate has relatively minor effect on inferences regarding θ. However in small samples, replacing λ by the maximum likelihood estimator may have a large effect on inference, particularly when there are several nuisance parameters in the model.

2.2 INTEGRATED LIKELIHOOD

The integrated likelihood is obtained for each fixed value of the parameter of interest by integrating out the nuisance parameter with respect to a weighting function. Let $\pi(\lambda)$

denote a weighting function defined on Λ, the space of possible values λ. Then the integrated likelihood function for θ is given by;

$$L(\theta;\pi) = \int_{\Lambda} L(\theta;\lambda)\,\pi(\lambda)\,d\lambda \tag{2.2}$$

Inference about θ will be based on $L(\theta;\pi)$, or equivalently on the logarithm of the integrated log-likelihood function $logL(\theta;\pi)$. In this study, we will consider the use of a uniform weighting function $\pi(\lambda)$=1; the resulting integrated likelihood function will be denoted L($\theta;U$) to emphasize that the uniform weight function applies to λ. A comprehensive discussion of the use of integrated likelihood methods in the Bayesian approach is given by Berger, Liseo and Wolpert (1998). Integrated likelihood functions have the advantage that, unlike conditional or marginal likelihoods, they are generally available and, in principle, are relatively easy to determine, although sophisticated computational methods may be needed to evaluate the integrals that arise.

In the Bayesian literature, the noninformative prior plays an important role. When we apply the Bayesian approach, we may not have any prior information about the parameter. In this situation, the statistician tries to find a prior that provides as little information about the parameter of interest as possible. One of the most important noninformative priors is the Jeffreys' prior which is equal to the square root of the Fisher information $I(\theta)$(Jeffreys 1939).

Sweeting (1995) has shown that one-parameter methods which operate solely on the likelihood L(θ) can also be used with an integrated likelihood. Examples of such methods are: (i) using the mode $\hat{\theta}$ of the likelihood as the estimate of θ, and (ii) using,

$$C = \left\{\theta : -2log\left(L\left(\hat{\theta}\right)/L(\theta)\right) < X_p^2(1-\alpha)\right\} \tag{2.3}$$

as an approximate $100\,(1-\alpha)$ confidence set for θ, where $X_p^2(1-\alpha)$ is the $(1-\alpha)$th quantile of the chi-squared distribution with p degrees of freedom (for a scalar θ, p=1). By using these facts, we will be able to compare the accuracy of inference about θ based on the integrated likelihood with that based on other modified likelihoods (profile likelihood and conditional likelihood) by comparing the coverage probabilities of confidence intervals of the form (2.3) based on such modified likelihoods.

2.3 MARGINAL LIKELIHOOD

Another way to eliminate a nuisance parameter is to construct a likelihood function based on a statistic **T** having the property that the distribution of **T** depends only on θ. Then we may form a genuine likelihood function for θ based on the density function of **T**; such a likelihood function is called a marginal likelihood function, since it is based on the marginal distribution of **T**. Marginal likelihoods were considered in detail by Kalbfleisch and Sprott (1970). The main drawback of this approach is that we may not be using all of the available information about θ in the data.

Suppose that there exists a statistic $\mathbf{T}= \mathbf{T(y)}$ such that the density of the data $\mathbf{y}$ may be written

$$P\left(y;\theta,\lambda\right) = P\left(t;\theta\right)P\left(y|t;\theta,\lambda\right). \tag{2.4}$$

In (2.4), the marginal likelihood function based on **T**=t is given by

$$L\left(\theta;t\right) = P\left(t;\theta\right),$$

whereas the joint likelihood function for (θ, λ) when **Y**=y is given by P$(y;\theta,\lambda)$.

2.4 CONDITIONAL LIKELIHOOD

Another approach to eliminating nuisance parameters can be applied whenever there exists a statistic $\mathbf{S} = \mathbf{S}\ (\mathbf{y})$ such that the conditional distribution of the data $\mathbf{y}$ given $\mathbf{S}$=s depends only on θ. In this case, we may form a genuine likelihood function for θ based on the conditional density of $\mathbf{Y}$ given $\mathbf{S}$=s ; this is called a conditional likelihood function. Suppose that the data can be transformed to the vector (t, s) such that;

$$P\left(t, s : \theta, \lambda\right) = P\left(t|s; \theta\right) P\left(s; \theta, \lambda\right). \tag{2.5}$$

The statistic $\mathbf{S}$ is a sufficient statistic for θ in the model with λ held fixed. A likelihood function for θ may be based on P$(t|s;\theta)$ which does not depend on λ; the resulting conditional likelihood function for θ is a genuine likelihood function. The use of conditional likelihood inference in models with many nuisance parameters was discussed by Anderson (1970), Kalbfleisch and Sprott (1970) and van der Vaart (1988).

2.5 DISCUSSION AND COMPARISON

The simplest approach to eliminating nuisance parameters is to replace them with their maximum likelihood estimates, leading to the profile likelihood. Many examples of misleading behavior of the profile likelihood have been given, leading to various corrections of the profile likelihood. Among the proposed corrections are modified profile likelihood (Barndroff-Nielsen, 1983) and the conditional profile likelihood (Cox and Reid, 1987).

The profile likelihood and integrated likelihood are not genuine likelihood functions. That is, an integrated likelihood function or a profile likelihood function do not, in general, correspond to a likelihood function arising from an observed statistic. However, integrated likelihood is closely related to the profile likelihood function in the sense that the first order approximations of procedures based on the uniform integrated likelihood function are the same as the first-order approximations of procedures based on the profile likelihood function.

Severini (1998b) has shown that the profile likelihood function can be viewed as an estimate of the genuine likelihood function. Indeed the profile likelihood can be used as if it were an ordinary likelihood to produce asymptotically ($n \to \infty$) correct inferences about a structural parameter. Severini (1998b) has also shown that the modified profile likelihood can be derived as an approximation to either the conditional or marginal likelihood when either of the latter likelihoods exist.

In large samples, there are unlikely to be large differences between the results based on these modified likelihood methods. However for a given small sample size, the results may differ. Our goal is to compare the validity of each likelihood method in terms of the coverage probabilities of confidence regions derived from them, using a region of the form (2.3), when the sample size is small. In this study, we will apply profile, conditional and integrated likelihood methods to the two-parameter gamma distribution for the situation where the shape(θ) and scale(λ) parameters are regarded as the structural and nuisance parameters respectively. The probability density of the two-parameter Gamma distribution is given by;

$$f\left(y;\theta,\lambda\right) = \frac{1}{\Gamma(\theta)\lambda^{\theta}} y^{(\theta-1)} \exp^{(-y/\lambda)} \; y > 0,$$

for parameters $\theta, \lambda > 0$. We proceeded as follows: First, we eliminated the nuisance parameter (λ) from the two-parameter gamma distribution by the integrated, conditional and profile likelihood methods. Then we generated 10,000 replications of various sample sizes from the two-parameter gamma distribution with known parameter values θ=2 and λ=3. Generating the data from the density with known parameters allows us to determine coverage probabilities for confidence intervals of the form (2.3) obtained from each of the modified likelihood methods for different sample sizes. The results are shown in Table 1. Binomial standard deviations belonging to each coverage probability are shown in bold numbers.

Table 1: 95% coverage probabilities from the modified likelihoods of the gamma density

SAMPLE SIZE	INTEGRATED LIK.	CONDITIONAL LIK.	PROFILE LIK.
n=7	0.9476 **(0.00221)**	0.9422 **(0.00190)**	0.9223 **(0.00172)**
n=15	0.9478 **(0.00216)**	0.9442 **(0.00237)**	0.9285 **(0.00240)**
n=20	0.9495 **(0.00218)**	0.9461 **(0.00221)**	0.9326 **(0.00245)**
n=30	0.9518 **0.00212)**	0.9481 **(0.00215)**	0.9364 **(0.00225)**

As it is expected for small sample sizes(n=7,n=15, n=20), profile likelihood is not as accurate as integrated likelihood and conditional likelihood. However the integrated likelihood produces coverage probabilities very close to the nominal level (0.95) for all sample sizes.

In Chapter 3, we will justify use of the cubic spline approximation for one parameter models and will propose how to use raindrop plots to compare different likelihoods. Indeed raindrop plots (Barrowman and Myers,2003) are an effective way to communicate the relative plausibility of different values of parameters of interest and provide a compact and informative way of displaying groups of likelihoods.

Alternative to the modified likelihood methods, in Chapter 4 we will calculate coverage probabilities of the structural parameter based on projection pursuit regression method. We will compare the coverage probabilities of the modified likelihood methods with the marginal coverage probabilities of the projection pursuit regression method.

3.0 POLYNOMIAL SPLINES

Polynomials have played a central role in approximation theory and numerical analysis for many years. The main drawback of polynomials for approximation purposes is that the class is relatively inflexible. Polynomials work well on sufficiently small intervals, but when we go to larger intervals, severe oscillations often appear. In order to achieve a class of approximating functions with greater flexibility, we can divide up the interval of interest into smaller pieces. In this chapter we will justify use of the cubic spline approximation for one-parameter models.

Polynomial splines are piecewise polynomials of some degree **d**. The breakpoints marking a transition from one polynomial to the next are referred to as "knots". A piecewise polynomial function $f(y)$ is obtained by dividing the domain of Y into contiguous intervals and f can be separated by a polynomial in each interval. In the literature on approximation theory, the term "linear spline" is applied to a continuous piecewise linear function. Such a piecewise linear spline with two knots(t_1, t_2) will have the following basis functions:

$$h_1(Y) = I(Y < t_1), h_2(Y) = I(t_1 \leq Y < t_2), h_3(Y) = I(t_2 \leq Y)$$

Similarly, the term "cubic spline" is reserved for piecewise cubic functions having two continuous derivatives, allowing jumps in the third derivative at the knots. It is common in statistics to require a simple approximation for a smooth relationship between response and predictor variable. Such relationship may be known but complicated or unknown. The cubic spline functions are very popular in data mining to serve for this job.

Given a maximum polynomial degree **d** and a knot vector **t**, the collection of polynomial splines having s continuous derivatives form a linear space. For example the collection of linear splines with knot sequence $(t_1, ..., t_k)$ is spanned by the functions

$$1, y, (y - t_1)_+, \ldots, (y - t_k)_+ . \tag{3.1}$$

where $(.)_+ = \max(., 0\ )$. This set is called the truncated power basis of the space. Classical cubic splines have d=3 and s=2 so that the basis has elements

$$1, y, y^2, (y - t_1)^3_+, \ldots, (y - t_k)^3_+ \tag{3.2}$$

However the truncated power functions (3.1) and (3.2) are known to have rather poor numerical properties. In linear regression problems, for example, the condition number of the design matrix deteriorates rapidly as the number of knots increases (Hansen and Kooperberg,2002).

Extended linear models (ELMs) were defined as a theoretical tool to understand the properties of spline-based procedures in a large class of estimation problems (Hansen,1994; Huang 2001). This class contains all of the standard generalized linear models as well as density and conditional density estimation, hazard regression, censored regression and spectral density estimation.

Friedman (1991) introduced multivariate adaptive regression splines (MARS), which is a polynomial spline methodology for estimating regression functions. Multivariate adaptive regression splines (MARS) is a method for flexible modeling of high dimensional data. The MARS method has become very popular in data mining because it does not assume or require any particular type of relationship between predictor variables and response variable. The multivariate adaptive regression splines (MARS) model can be written as

$$f\left(y;\beta\right) = \sum_{j=1}^{J} \beta_j B_j\left(y\right) \tag{3.3}$$

for a given set of basis functions $B_1(y), ..., B_J(y)$. The unknown parameters $\beta_1, ..., \beta_J$

in MARS are estimated using least squares.

In functional ANOVA, spline basis elements and their tensor products are used to construct the main effects and interactions. Stone (1994) gave the first theoretical treatment of convergence of spline estimation with functional ANOVA decompositions.

Most of the early applications of splines were focused mainly on curve estimation. These tools also have proved effective for multivariate problems. In the context of density estimation, the log-spline procedure of Kooperberg and Stone (1991) shows excellent spatial adaptation, capturing the full height of spikes without overfitting smoother regions. Note that approximation of densities by log-spline is similar to approximation of log-likelihoods in the sense that the argument y of the density plays the same role as the value of the chosen parameter values θ in log- likelihood estimation.

Kooperberg and Stone (1991,1992) modeled the log-density as a natural cubic spline. Like the log-spline in density estimation, log-likelihood can be modeled as a natural cubic spline. Indeed in large samples, central limit and saddle-point approximation theory suggest that cubic splines may be used to approximate the log-likelihood functions.

Natural cubic splines are twice continuously differentiable, piecewise polynomials defined relative to a knot sequence $t = (t_1, ...t_k)$. Within each interval $[t_1, t_2], ...[t_{K-1}, t_K]$, natural cubic splines are cubic polynomials, but on $(L, t_1]$ and $[t_K, U)$ (beyond the first and last knots) they are forced to be linear. We assume that log-likelihood can be written in the form:

$$logL\left(y;\beta\right) = \sum_{j=1}^{J} \beta_j B_j\left(y_j\right) \tag{3.4}$$

where J is the number of basis functions. The basis of natural cubic spline with K knots is $B_1\left(y\right) = 1,\ B_2\left(y\right) = y,\ B_{k+2}\left(y\right) = d_k\left(y\right) - d_{K-1}\left(y\right),\ k = 1, ..., K-2$, where

$$d_k(y) = \frac{(y - t_k)_+^3 - (y - t_K)_+^3}{t_K - t_k}.$$

Another basis representation of the natural cubic splines is the B-spline basis (De Boor,1978). B-splines are constructed from polynomial pieces joined at certain values of y. Once the knots are given, it is easy to compute the B-splines recursively, for any desired degree of the polynomial.

An important question in spline modeling is to decide the number of knots in the model. The choice of knots has been a subject of much research; too many knots lead to overfitting of the data, too few knots lead to underfitting. Some authors have proposed automatic schemes for optimizing the number and the position of knots (Friedman and Silverman, 1989 ; Kooperberg and Stone 1991). Kooperberg and Stone (1991) found that it matters less where the knots are placed than how many knots are chosen. Fortunately, equally spaced fixed knots ensure that there are enough data within each region to get sensible fits. This choice also guards against outliers overly influencing the fitted curve. In this study we will determine the appropriate numbers of knots by the AIC method and will assume that knots are equally spaced.

The knot selection methodology in log-spline density estimation involves initial knot placement, stepwise knot addition, stepwise knot deletion and final model selection based on the Akaike information criteria (AIC). Log-spline density estimation is discussed by Stone (1990), who used a basis of the form $1, B_1(y,), ...B_J(y,)$ of natural cubic spline functions, where $j = k - 1$, $t = (t_1, ..., t_k)$. Let t be fixed and G denote the J-dimensional span of the functions $B_1, ...B_J$, so that any $g \in G$ is of the form;

$$g(y, \beta, t) = \beta_1 B_1(y, t) + ... + \beta_J B_J(y, t). \tag{3.5}$$

Then, density functions on (L, U) of the form

$$f(y:\beta,t) = \exp[g(y:\beta,t) - C(\beta,t)]$$

$$= \exp(\beta_1 B_1(y,t) + ... + \beta_j B_j(y,t) - C(\beta,t)), L < y < U$$

(called a log-Spline Model) are used to approximate a member of exponential family density functions of interest. Given a random sample $Y_1, ...Y_n$ of size n from a distribution on (L, U) having an unknown density function in an exponential family, the log-likelihood function corresponding to the log-spline model is given by

$$L(\beta,t) = \sum_i \log f(Y_i;\beta,t) = \sum_i \sum_j \beta_i B_j(Y_i,t) - nC(\beta,t)$$

Then, the maximum likelihood estimate of β is calculated by arg max $L(\beta, t)$. The MLE's of β and fixed choice of knots $\mathbf{t}$ can be found efficiently in reasonably-sized problems through simple Newton-Raphson iteration.

3.1 MODEL SELECTION CRITERION IN POLYNOMIAL SPLINE

To approximate the log-likelihood functions, first we will sample n values of the parameter vector uniformly from a chosen compact convex region in the parameter space. Next, the log-likelihood for each choice of parameter vector is calculated. Then the log-likelihood is taken to be the response and the elements of θ to be the predictors, and in a regression model we will fit by natural cubic splines. However before going through this process, we should determine the optimum number of knots that we include in the models. Clearly the number of knots will determine the complexity of the model.

In this chapter we describe and illustrate the methods that we used to select the number of knots in spline models. We assume that we have the target variable Y(log-likelihood), a vector of inputs x(parameter vector) and the prediction model $\hat{f}(x)$. The loss function for measuring errors between Y and $\hat{f}(x)$ is denoted by $L\left(Y, \hat{f}(x)\right)$ and is the squared error loss function.

$$L\left(Y, \hat{f}(x)\right) = \left(Y - \hat{f}(x)\right)^2 \tag{3.6}$$

The test error (Err) is the expected prediction error over an independent test sample.

$$Err = E[L\left(Y, \hat{f}(x)\right)]$$

We estimate the test error from the training error and training error as the average loss over the training sample:

$$\overline{err} = \frac{1}{N}\sum_{i=1}^{N} L\left(Y, \hat{f}(x_i)\right)$$

Training error decreases as model complexity increases, finally dropping to zero if we increase the model complexity enough. For this reason, training error is not a good estimate of the test error. Fortunately, there is an optimal model complexity that gives the minimum test error. Indeed Akaike and Bayesian information criterion trade between bias and variance to get optimum level of the test error.

3.1.1 Akaike Information Criterion

Akaike information criterion(AIC) is one of the most popular methods to determine the optimum number of variables in the ideal model. Assume that we have the linear model, $Y = f(x) + \epsilon$ where $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma_\epsilon^2$. We can derive an expression for the expected prediction error of a regression fit $\hat{f}(x)$ at an input point $X = x_0$ using squared error loss

$$Err\left(x_0\right) = E[(Y - \hat{f}\left(x_0\right))^2|X = x_0]$$

$$= \sigma_\epsilon^2 + [E\hat{f}\left(x_0\right) - f\left(x_0\right)]^2 + E[\hat{f}\left(x_0\right) - E\hat{f}\left(x_0\right)]^2$$

$$= \sigma_\epsilon^2 + Bias^2\left(\hat{f}\left(x_0\right)\right) + Var\left(\hat{f}\left(x_0\right)\right)$$

$$= IrreducibleError + Bias^2 + variance.$$

For a linear model fit $\hat{f}_p\left(x\right) = \hat{\beta}^T X$, where the parameter vector β with p components is fit by least squares,

$$Err\left(x_0\right) = E[(Y - \hat{f}\left(x_0\right))^2|X = x_0] = \sigma_\epsilon^2 + [\hat{f}\left(x_0\right) - E\hat{f}_p\left(x_0\right)]^2 + \|h\left(x_0\right)\|^2\sigma_\epsilon^2$$

$h\left(x_0\right)$ is the N-vector of linear weights that produce the fit $\hat{f}_p\left(x_0\right) = X_o^T\left(X^TX\right)^{-1}X^TY$, and hence

$$Var[\hat{f}_p\left(x_0\right)] = \|h\left(x_0\right)\|^2\sigma_\epsilon^2.$$

This variance changes with x_0 and its average(over the sample values x_i) is $\frac{p}{N}\sigma_\epsilon^2$ and the in-sample error is

$$\frac{1}{N}\sum_{i=1}^{N} Err\left(x_i\right) = \sigma_\epsilon^2 + \frac{1}{N}\sum_{i=1}^{N}[f\left(x_i\right) - E\hat{f}\left(x_i\right)]^2 + \frac{p}{N}\sigma_\epsilon^2. \qquad (3.7)$$

Here the model complexity is directly related to the number of parameters p. We define optimism as the expected difference between Err_{in} and the training error $e\bar{r}r$. Let Y^{new} indicate that we observe N new response values at each of the training points x_i, $i = 1, ..., N$.

Then

$$Err_{in} = \frac{1}{N}\sum_{i=1}^{N} E_y E_{Y^{new}} L\left(Y_i^{new}, \hat{f}(x_i)\right). \tag{3.8}$$

and

$$optimism = Err_{in} - E_y\left(\overline{err}\right). \tag{3.9}$$

For a squared error loss function, we can show that,

$$optimism = \frac{2}{N}\sum_{i=1}^{N} Cov\left(\hat{y}_i, y_i\right)$$

where Cov indicates covariance. Then we have the important relation

$$Err_{in} = E_y\left(\overline{err}\right) + \frac{2}{N}\sum_{i=1}^{N} Cov\left(\hat{y}_i, y_i\right)$$

For the additive error model $y = f(x) + \epsilon$,

$$Err_{in} = E_y\left(\overline{err}\right) + 2\frac{d}{N}\sigma_\epsilon^2$$

Simply, the AIC is an estimate of Err_{in} when a log-likelihood loss function is used.

$$AIC = -2loglik + 2\frac{d}{N}\sigma_\epsilon^2 \tag{3.10}$$

To use AIC for model selection in spline modeling, we choose the model giving smallest AIC over the set of models. Let $f_\alpha(x)$ be the set of models indexed by a tuning parameter α and denote by $\overline{err}(\alpha)$ and $d(\alpha)$ the training error and number of parameters for each model, then we can define

$$AIC(\alpha) = \overline{err}(\alpha) + 2\frac{d(\alpha)}{N}\hat{\sigma_\epsilon^2}$$

The function AIC(α) provides an estimate of the test error and we choose the tuning parameter $\hat{\alpha}$ that minimizes the test error. With the spline models the tuning parameter is typically the number of knots in the spline model. The number of knots controls the complexity of the spline model and we want to find the number of knots that minimizes the AIC. In the present study, before approximating the log-likelihoods by cubic splines, we determine the appropriate number of knots by AIC and then equally space the knots.

3.1.2 The Bayesian Information Criterion

The Bayesian information criterion (BIC) is applicable where the fitting is carried out by maximization of log-likelihood. The form of Bayesian information criterion (BIC) is

$$BIC = -2loglik + (logN)\, d \tag{3.11}$$

Under the Gaussian model, assuming the variance σ_ϵ^2 is known, we can write BIC as,

$$BIC = \frac{N}{\sigma_\epsilon^2}[\overline{err} + (logN)\, \frac{d}{N}]$$

BIC tends to penalize complex models more heavily than the Akaike information criterion, giving preference to simpler models in selection. Our simulation results indicate that to approximate the log-likelihood of the exponential distribution, both AIC and BIC agreed on 1 as the appropriate number of knots for the natural cubic spline model.

3.2 ASSESSING THE QUALITY OF THE APPROXIMATION

One of the most important tasks for us is to assess the quality of the approximation. In this chapter we use a one-parameter distribution (exponential distribution) to illustrate how to assess the approximation in terms of the following three criteria:

1) Coverage probability of the highest density region (HDR) of the approximated log-likelihood function.

2) Mean squared error (MSE) of the estimator obtained as the maximum over the parameter of the approximation of the log-likelihood.

3) Average interval length of the highest density region.

Before presenting our results, first we will show how to find the region from which we select parameter values, and then we will introduce the algorithms that we used to find the coverage probability of the (HDR) and the mean squared error (MSE) of the "maximum likelihood estimator" obtained from the approximated log-likelihood functions.

To find the region in the parameter space from which we sample parameter values, we have the following algorithm. Recall that in chapter 2, we have the equation (2.3):

$$C = \left\{\theta : -2log\left(L\left(\hat{\theta}\right)/L\left(\theta\right)\right) < X_p^2\left(1-\alpha\right)\right\}$$

1) First, from the data find the maximum likelihood estimator of the parameter.

2) Plug in the maximum likelihood value in (2.3) for $\widehat{\theta}$

3) Solve (2.3) to find the lower and upper limit of the interval of parameter of interest.

4) Finally, parameter values are sampled uniformly from the interval between the lower and upper limits found in step(3).

Assume that we want to approximate the log-likelihood of the exponential distribution:

$$f\left(y;\theta\right) = \frac{1}{\theta} exp \frac{-y}{\theta} \tag{3.12}$$

It is easy to show that the maximum likelihood estimator of θ is $1/\overline{y}$. We plug in $1/\overline{y}$ for $\widehat{\theta}$ in (2.3). Then we solve (2.3) to find the lower and upper limits of the interval. Finally, we sample the parameter values uniformly from this region.

3.2.1 Highest Density Region of the Approximated Log-likelihood Function

Statistical methods summarize a probability distribution by a region of the sample space covering a specified probability. One method of selecting such a region is to contain points of relatively high density. Hyndman (1996) proposed a simple method for computing a highest density region. If we have a distribution $f(y)$, we would like to find the region $R(y)$ that satisfies

$$(i) \int_{R(y)} f(y)\, dy = 1 - \alpha$$

$$(ii) Size R(y) \leq Size R^{'}(y)$$

for any region $R'(y)$ which satisfies $\int_{R'(y)} f(y) \geq 1 - \alpha$ such a region is called the highest density region (HDR) of the distribution $f(y)$.

It follows from the definition that the HDR has the smallest possible volume in the sample space of y. One of the most important advantages of using HDR is that the mode is contained in every HDR.

For normal distributions, the high density region (HDR) coincides with the usual probability region symmetric about the mean and this is also true for all unimodal and symmetric distributions. Another characteristic of HDR is that it can contain disjoint intervals when the underlying distribution is multimodal.

There have been several suggestions for constructing the HDR from a general bounded and continuous univariate density $f(y)$. Wright (1986) proposed an algorithm which includes numerical integration of $f(y)$. Hyndman (1996) developed a density quantile approach that computes the HDR. In this study we adapted Hyndman's idea to find the highest density region of the approximated log-likelihood function. We use the following algorithm to find the HDR of the approximated log-likelihoods;

1) Use (2.3) to sample the parameter values from the interval of parameter of interest.

2) Determine the appropriate number of knots for the log-likelihood function.

3) Approximate the log-likelihood function by a spline model by treating the likelihood values as response and parameter values as independent variable(s).

4) Find the fitted values of the approximated spline model for chosen θ values.

5) Treat those fitted values as the height and find the total area under the approximated function by summing those heights.

6) Divide each height by the total area to make the total area 1 under the spline function.

7) Next step is finding the biggest height(mode) and add the next biggest height and continue this until we reach the given confidence level.

8) Finally, claim the corresponding minimum and maximum parameter values as the lower and upper bound of the highest density region of approximated log-likelihood.

To evaluate the performance of our approximation in terms of coverage probability, we calculate intervals $[\hat{L}(y), \hat{U}(y)]$ from the highest density region of the approximated log-likelihood function obtained by the algorithm above. We generated 10,000 samples from the exponential distribution with different sample sizes and parameter values. After the approximation of log-likelihood by cubic spline function, we check whether the highest density region (HDR) includes the known true parameter value.

Table 2 shows 90% coverage probabilities from the HDR of natural cubic spline and B-spline approximations of the log-likelihood of the exponential distribution. The first value is the coverage probability calculated from natural cubic spline approximation and number in bold is the coverage probability from B-spline approximation. Both methods produce very similar results and those results are reasonably close to the asserted confidence level(α=90) of the interval. Standard deviation of coverage probabilities is around .003. The result of this simulation study indicates that both the natural cubic spline and B-spline approximation provide excellent coverage probabilities except when the sample size is 10.

Table 2: 90% coverage probabilities from the Highest Density Region(HDR) of approximated log-likelihood of the exponential distribution with θ=2

SAMPLE SIZE	$\theta = 2$	$\theta = 5$	$\theta = 10$
n=10	0.8752	0.8851	0.8677
	0.8773	**0.8844**	**0.8703**
n=20	0.8900	0.9053	0.9079
	0.8904	**0.9056**	**0.9075**
n=30	0.9061	0.8992	0.9075
	0.9061	**0.8991**	**0.9074**
n=50	0.8953	0.9027	0.9052
	0.8952	**0.9028**	**0.9052**

In addition when the sample size is 10, the coverage probabilities for the B-spline and the natural cubic spline slightly differ, but for other sample sizes they produce very close numbers. For the large sample sizes (n=30 and n=50), the agreement between the coverage probabilities is nearly perfect, agreeing to within +/- 0.001.

3.2.2 The Mean Squared Error Of The Approximated Log-likelihood Function

In sufficiently large samples, the log-likelihood function is known to be approximately a quadratic form in a neighborhood of the MLE of θ. In the tails (outside of a region around the MLE of θ), the likelihood function is approximately linear in either θ or in the natural parameter of an exponential family saddlepoint approximation. Saddlepoint approximations have been used successfully to approximate the tails of distributions; discussion of many such applications are given by Reid (1988), Goutis and Casella (1999), and Huzurbazar (1999).

By approximating the log-likelihood function using regression methodology, we have the advantage that an estimate of the mean squared error(MSE) of the structural parameter can be reported along with the approximation. The mean squared error (MSE) of an estimator W of a parameter θ is the function of θ defined by $E_\theta\left(W-\theta\right)^2$. MSE measures the average squared difference between the estimator W and the parameter θ.

MSE incorporates two components, one measuring the variability of the estimator(precision) and the other measuring its bias(accuracy).

$$E_\theta\left(W-\theta\right)^2 = VAR_\theta W + \left(E_\theta W-\theta\right)^2 = VAR_\theta W + \left(BIAS_\theta W\right)^2 \tag{3.13}$$

To find an estimator with good MSE properties, we need to find estimators that control both variance and bias. It is obvious that unbiased estimators do the best job of controlling bias and the MSE is equal to the variance when the estimator is unbiased. Since we approximate the log-likelihood function from the density distribution with known parameter, we can calculate the exact mean square error of the maximum likelihood estimator. We will compare the exact MSE, with the empirical MSE from the approximated log-likelihood. To find the estimated MSE from the approximated log-likelihood function, we have the following algorithm:

1) Use (2.3) to select parameter values from the interval of parameter of interest.

2) Calculate the log-likelihood values for chosen parameter values from step 1.

3) Treat the log-likelihood values as response and parameter values as independent variable and find the appropriate number of knots.

4) Approximate the log-likelihood by the cubic spline function with appropriate number of knot(s).

5) Find the fitted values corresponding to the chosen parameter values.

6) Find the biggest fitted value (mode) and claim the corresponding θ value as the "MLE" estimator of the unknown true parameter.

7) Repeat 1-6 to find the empirical MSE by using;

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\theta - \hat{\theta}_i\right)^2$$

where $\hat{\theta}_i$ is the estimation of the mle from the ith simulation, θ is known true parameter and N is the simulation size.

In our simulation, we generated 10,000 samples from an exponential distribution for each sample size and parameter value. Selecting such a large simulation size(10,000) will allow us to obtain small binomial standard deviations of the parameter estimates. Table 3 shows the empirical MSE of the maximum likelihood estimator (mle) from the cubic spline approximations. The first value is MSE of natural cubic spline approximation and the number in bold is MSE of B-spline approximation.

Table 4 presents the exact MSE of the mle from the exponential distribution. As it is expected, the MSE of the approximations results are getting closer to the expected MSE as the sample size increases.

Since we generate data from the exponential distribution with known parameter, we can calculate the exact MSE of the maximum likelihood estimator. It is easy to show that for the exponential distribution $1/\overline{y}$ is the mle of θ. Furthermore it can be shown that

$$E[1/\overline{y}] = \frac{n\theta}{n-1}$$

$$E[(1/\overline{y})^2] = \frac{n^2\theta^2}{(n-1)(n-2)}$$

From (3.14) we can write MSE of the mle;

$$MSE(1/\overline{y}) = E[(1/\overline{y})^2] - (E[1/\overline{y}])^2 + (E[1/\overline{y}] - \theta)^2$$

$$MSE(1/\overline{y}) = \frac{n^2\theta^2}{(n-1)(n-2)} - \frac{n^2\theta^2}{(n-1)^2} + (\frac{n\theta}{(n-1)^2} - \theta)^2$$

After some simplifications;

$$MSE(1/\overline{y}) = \frac{(n+2)\theta^2}{(n-1)(n-2)}$$

Now we can calculate the exact MSE of the maximum likelihood estimator of the exponential distribution. Table 4 gives the exact MSE's of the maximum likelihood estimator of the exponential distributions for different sample sizes and parameter values.

The result of the simulation study reveal that both the natural cubic spline approximation and cubic B-spline approximation provide an accurate point estimates of the known true parameter. Estimated MSE results are consistent with the coverage probability results in the sense that for large sample sizes difference between exact MSE and estimated MSE's are getting smaller and smaller.

Table 3: Empirical MSE's of approximated log-likelihoods of the exponential distribution

SAMPLE SIZE	$\theta = 2$	$\theta = 5$	$\theta = 10$
n=10	0.572	4.371	14.79
	0.577	**4.412**	**15.18**
n=20	0.234	1.63	6.98
	0.232	**1.667**	**6.93**
n=30	0.169	1.007	4.29
	0.174	**1.014**	**4.27**
n=50	0.093	0.537	2.152
	0.094	**0.538**	**2.153**

Table 4: Exact MSE's of maximum likelihood estimator of the exponential distribution

SAMPLE SIZE	$\theta = 2$	$\theta = 5$	$\theta = 10$
n=10	0.667	4.17	16.67
n=20	0.257	1.608	6.43
n=30	0.157	0.985	3.94
n=50	0.088	0.552	2.21

3.3 PRESENTATION OF LIKELIHOODS BY RAINDROP PLOT

Likelihood plays an important role in the frequentist and Bayesian statistical paradigms. In many areas it is important to display information about a parameter obtained from several independent studies, such as obtained in meta-analysis. However, instead of displaying a collection of likelihoods, it is more common to display point estimates with corresponding bars representing confidence intervals. This traditional display is sometimes misinterpreted by users as implying that all values within a confidence interval are equally supported by the data as possible values of the parameters.

This traditional display of point estimates with confidence intervals is sufficient if the log-likelihood or log density is close to being normal. But with small sample sizes and in nonlinear models departures from normality may occur, and the traditional display may be uninformative or even misleading. In contrast the likelihood ratio provides the relative plausibility of different values of θ and allows for asymmetry. For this reason a graphical display of the likelihood would be desirable.

To display likelihood ratios, Barrowman and Myers (2003) introduced raindrop plots. Let $\hat{\theta}$ be the maximum likelihood estimator and from (2.3) we have

$$C = \left\{\theta : -2log\left(L\left(\hat{\theta}\right)/L\left(\theta\right)\right) < X_p^2\left(1-\alpha\right)\right\} \tag{3.14}$$

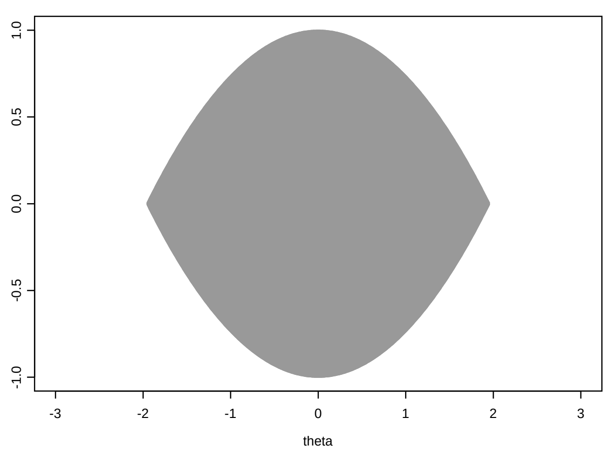
as an approximate $100\left(1-\alpha\right)$ confidence set for θ, where $X_p^2\left(1-\alpha\right)$ is the $(1-\alpha)$th quantile of the chi-squared distribution with p degrees of freedom (for a scalar θ, p=1). If we have one parameter in the model, $X_1^2\left(0.95\right)/2 \approx 1.92$ and a 95 percent interval is equivalent to

$$C = \left\{\theta : log\left(L\left(\hat{\theta}\right)/L\left(\theta\right)\right) \geq -1.92\right\}. \tag{3.15}$$

To produce a raindrop shape, the log-likelihood ratio is graphed over the range where it is greater than or equal to -1.92. It is then reflected at the horizontal line at zero and the resulting region is shaded. The reflection and shading simplify the perceptual task for the viewer who wishes to compare several likelihoods and detect asymmetries and other differences in shape.

In Figure 1 we show a raindrop plot obtained from normally distributed random variables. The variation in the height of the raindrop shows the relative plausibility of parameter values within this range. The area of the raindrop thus relates to the shape and scale of a distribution.

Figure 1: Raindrop plot of standard normal distribution



The raindrop plot is based on reflecting the log-likelihood. However, alternative approaches are possible using, for example, the likelihood or log densities (Miettinen ,1985). One advantage of using the log-likelihood is that the log scale allows a tighter display vertically, so that many raindrops can be shown on a single page. Another advantage occurs when there are multimodes in the likelihoods. In this case raindrops may consist of separate pieces which represent confidence sets.

Barrowman and Myers (2003) also showed how to use raindrop plots to display distributions. Instead of the log-likelihood, they used the log density, with a cutoff corresponding to a probability of 0.95. They call this an high density region-raindrop(HDR) plot. HDR-raindrop plots provide information on where distributions have the bulk of their mass and also skewness and kurtosis. Although an HDR-raindrop plot does not show the extreme tails of a distribution, the shape of the sides of an HDR-raindrop are suggestive of the tail behavior.

In this study we modify the HDR-raindrop plot for the approximation of the log-likelihoods. Our modified raindrop plot will reflect the region from the highest density region of the approximated log-likelihood. Edge of the modified raindrop plot will give the upper and lower bound of the highest density region of the approximation. Interpretation of those graphs will be same as the classical raindrop plots. Figure 2,3,4 show several HDR-raindrop plot of approximated log-likelihood of exponential distribution with different parameter values and sample sizes.

In figure 2, we have the 90% HDR-raindrop plot of natural cubic spline approximation of log-likelihood of the exponential distribution with θ=2 and n=30. This graph provides some useful information such as the mode is around 2.2 and lower and upper bound of the 90% confidence region is approximately 1.65 and 2.8. Furthermore we can conclude that the approximated log-likelihood function is fairly symmetric. We can give a similar interpretation for the other HDR-raindrop plots.

The raindrop plot is particularly useful for Bayesian analysis because posterior distributions are rarely normal (Barrowman and Myers, 2003). In chapter 5 we will define the highest posterior density raindrop plot to display the approximated log-posterior densities.

Figure 2: 90% HDR raindrop plot of the approximated log-likelihood of the exponential distribution with $\theta = 2$, n=30

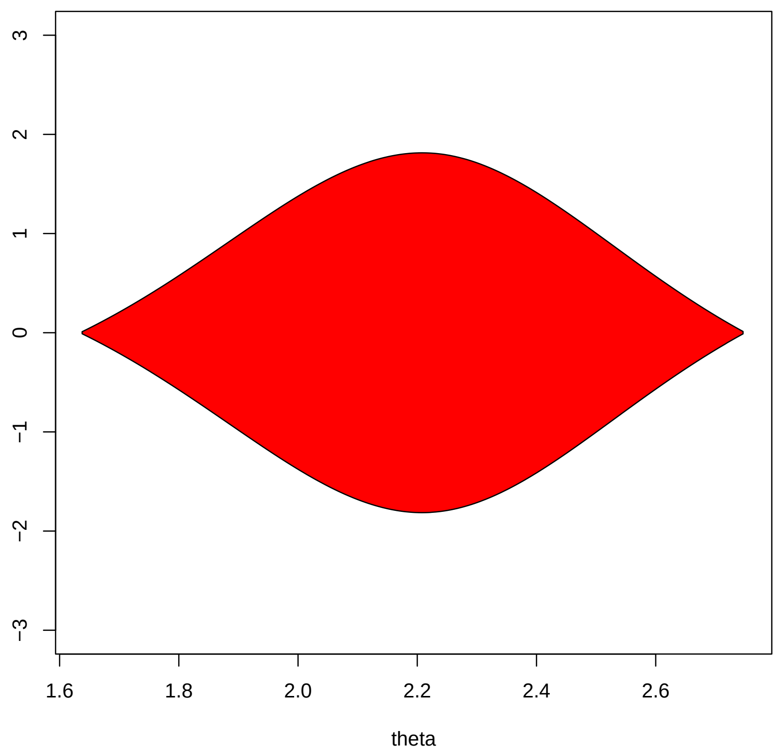


Figure 3: 70% HDR-raindrop plot of approximated log-likelihood of the exponential distribution with $\theta = 2$, n=30

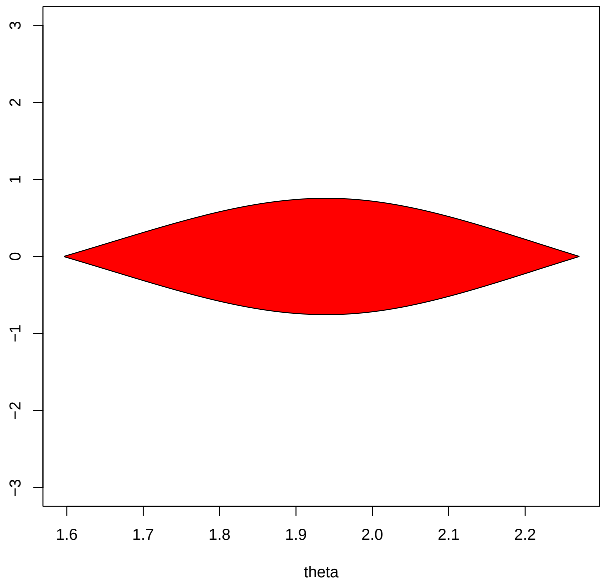
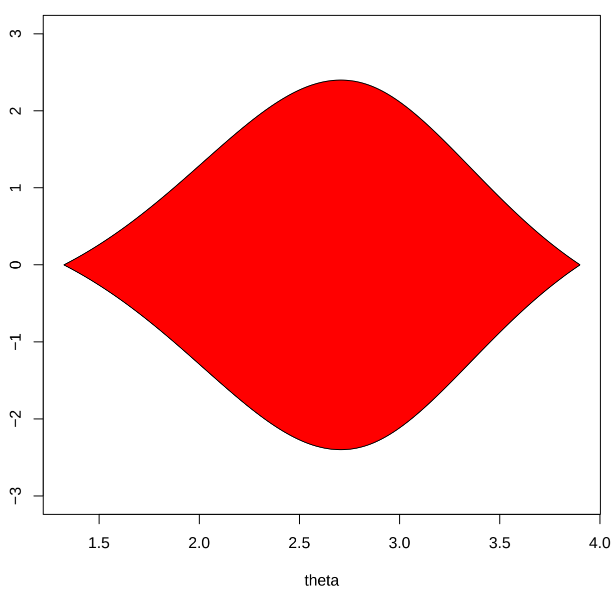


Figure 4: 95% HDR-raindrop plot of approximated log-likelihood of the exponential distribution with $\theta = 3$, n=10



4.0 APPROXIMATION OF LOG-LIKELIHOOD BY PROJECTION PURSUIT REGRESSION FOR MORE THAN ONE VARIABLE

In Chapter 3, we showed how to approximate the log-likelihood by spline functions when there is only one parameter. Specifically we approximated the log-likelihood of the exponential distribution by the natural cubic spline and B-spline. And we assessed the quality of approximation by calculating the coverage probabilities depend on the highest density region and mean squared error of the structural parameter.

In chapter 2, we used the modified likelihood methods to eliminate the nuisance parameter from the model. Alternative to the modified likelihood methods, in this chapter we will approximate the log-likelihood by projection pursuit regression (PPR) method without eliminating the nuisance parameter. Our goal is to calculate the marginal coverage probabilities of the structural parameter from PPR and compare those results with the coverage probability results from the modified likelihoods. As it is in chapter 3, the marginal coverage probabilities of the structural parameter will depend on highest density region of the approximated log-likelihood.

Projection pursuit regression (PPR) was introduced by Friedman and Stuetzle (1981) and refined by Friedman (1984). Projection pursuit regression (PPR) models the regression surface as a sum of general smooth functions of linear combinations of the predictor variables. The PPR method has attracted much attention as a method for estimating smooth functions of several variables from noisy scattered data.

Recall that we have an input vector x (parameter vector) with p components and a target

y(Log-likelihood). Let w_m, $m = 1, 2, ..., M$ be a unit p-vectors of unknown parameters. The projection pursuit regression model has the form

$$f(y) = \sum_{m=1}^{M} g_m \left(w_m^T X \right) \tag{4.1}$$

The function $g_m \left(w_m^T X \right)$ is called a ridge function in R^p. It varies only in the direction defined by vector w_m. The idea of PPR is to approximate the mean regression function by a sum of such ridge functions. Diaconis and Shahshahani (1984) provide an approximation theory justification as the number of ridge functions goes to infinity.

To fit PPR model we minimize the error function

$$\sum_{i=1}^{N} [y_i - \sum_{m=1}^{M} g_m \left(w_m^T x_i \right)]^2 \tag{4.2}$$

For given g, we want to minimize (4.2) over w. We impose the natural cubic spline for g and estimate w by Gauss-Newton method. This method can be simply derived as follows. Let w_{old} be the current estimator for w. Then we can write

$$g\left(w^T x_i\right) \approx g\left(w_{old}^T x_i\right) + g'\left(w_{old}^T x_i\right)\left(w - w_{old}^T x_i\right) \tag{4.3}$$

to give

$$\sum_{i=1}^{N} [y_i - \sum_{m=1}^{M} g\left(w^T x_i\right)]^2 \approx \sum_{i=1}^{N} g'\left(w_{old}^T x_i\right)^2 [(w_{old}^T x_i + \frac{y_i - g\left(w_{old}^t x_i\right)}{g'\left(w_{old}^T x_i\right)} - w^T x_i]^2 \tag{4.4}$$

To minimize the right-hand side, we carry out a least square regression with target $w_{old}^T x_i + \frac{y_i - g\left(w_{old}^t x_i\right)}{g'\left(w_{old}^T x_i\right)}$ on the input x_i, with weights $g'\left(w_{old}^T x_i\right)^2$ and no intercept term. This produces the updated coefficient vector w_{new}. Estimation of w continues until it converges. In high dimensional problems the number of terms M is usually estimated as part of the step-wise strategy. The model building stops when the next term does not improve the fit of the model. Cross-validation and aikake information criterion can be used to determine M.

To assess the natural cubic spline approximation by the projection pursuit regression method in terms of coverage probabilities, we generated data from the density distributions (gamma, pareto and log-normal densities) with known parameter values. We follow the similar algorithm to find the marginal coverage probabilities as we had for the one parameter models;

1) Find the intervals for the each parameter by using (2.3).

2) Sample parameter values for the each parameter from these interval using the uniform distribution.

3) Apply the projection pursuit method to find the regression function (impose natural cubic spline for the smoothing function).

4) Find the highest density region for the each parameter as explained in 3.2.1.

5) Finally, for each parameter calculate the marginal coverage probabilities from the highest density regions.

We also calculate joint confidence interval for the nuisance parameter and structural parameter. For the maximum likelihood estimator $\left(\hat{\theta}, \hat{\lambda}\right)$, the boundary of an elliptical $100(1-\alpha)\,\%$ approximate confidence region for(θ, λ) is given by

$$\left(\theta-\hat{\theta}, \lambda-\hat{\lambda}\right)[V_L\left(\hat{\theta}, \hat{\lambda}\right)]^{-1}\left(\theta-\hat{\theta}, \lambda-\hat{\lambda}\right)^T \tag{4.5}$$

where $V_L(\theta, \lambda)$ is the covariance matrix of $\left(\hat{\theta}, \hat{\lambda}\right)$ given(with regularity conditions) by the inverse of

$$E\left(\frac{\partial lnL}{\partial \theta_i}\frac{\partial lnL}{\partial \lambda_j}\right)$$

To get joint coverage probabilities of θ and λ, we choose the highest density region from (4.5) with given confidence level. For the simulation, we generated data from the gamma density for different θ and λ values. In total we have 9 cases, since we use 2,3,4 for θ and 4,7,10 for λ. Tables 5 provides the marginal coverage probabilities of θ and λ from the projection pursuit regression method and also joint coverage probabilities of θ and λ from (4.5).

Table 6 provides the marginal coverage probabilities of θ from the modified likelihood methods (profile, conditional and integrated likelihood methods) and marginal coverage probabilities of θ from the projection pursuit method.

Our main interest is to compare the marginal coverage probabilities, mean squared error (MSE) of structural parameter and the average interval length from the projection pursuit regression method with the modified likelihood methods(conditional, integrated and profile likelihood method). We will discuss our findings at the last chapter.

In table 7, we present mean squared error(MSE) of the θ from modified likelihoods and from the projection pursuit regression method. And Table 8 presents average interval length for the modified likelihood methods and average interval length for the PPR method.

Table 5: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression method(PPR) of approximated log-likelihood of the gamma density with θ=2, λ=4

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9427	0.9392	0.7449
n=15	0.9462	0.9448	0.7511
n=20	0.9488	0.9482	0.7562
n=30	0.9510	0.9536	0.7598

Table 6: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=2, λ=4

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9417	0.9378	0.9192	0.9427
n=15	0.9445	0.9444	0.9217	0.9462
n=20	0.9472	0.9476	0.9264	0.9488
n=30	0.9503	0.9514	0.9386	0.9510

Table 7: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	11.24	11.83	13.29	10.93
n=15	4.82	5.24	6.35	4.62
n=20	0.73	0.75	1.49	0.76
n=30	0.42	0.43	0.99	0.44

Table 8: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=2, λ=4

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	45.79	56.45	50.46	46.77
n=15	19.76	26.68	22.2	19.64
n=20	2.96	6.17	3.12	3.11
n=30	1.66	3.96	1.75	1.76

Table 9: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=2, λ=7

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9418	0.9327	0.7517
n=15	0.9412	0.9363	0.7563
n=20	0.9462	0.9381	0.7648
n=30	0.9512	0.9462	0.7643

Table 10: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=2, λ=7

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9387	0.9394	0.9182	0.9418
n=15	0.9411	0.9402	0.9237	0.9412
n=20	0.9465	0.9468	0.9268	0.9462
n=30	0.9497	0.9527	0.9375	0.9512

Table 11: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	15.17	15.8	16.26	14.24
n=15	2.83	2.92	3.16	2.75
n=20	0.75	0.75	0.88	0.76
n=30	0.59	0.60	0.64	0.63

Table 12: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=2, λ=7

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	62.81	67.18	64.91	63.92
n=15	12.65	14.12	13.14	12.27
n=20	3.42	3.86	2.92	3.61
n=30	2.41	2.74	2.51	2.71

Table 13: 95% marginal and joint coverage probabilities θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=2, λ=10

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9378	0.9411	0.7619
n=15	0.9392	0.9468	0.7624
n=20	0.9472	0.9491	0.7633
n=30	0.9528	0.9517	0.7666

Table 14: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=2, λ=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9382	0.9362	0.9127	0.9378
n=15	0.9431	0.9399	0.9138	0.9392
n=20	0.9461	0.9466	0.9243	0.9472
n=30	0.9501	0.9523	0.9346	0.9528

Table 15: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	26.38	27.49	29.15	26.83
n=15	6.23	6.74	7.25	6.6
n=20	2.19	2.39	2.99	2.25
n=30	0.90	0.93	1.01	0.99

Table 16: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=2, λ=10

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	63.16	71.36	65.64	62.13
n=15	14.44	19.71	15.32	14.75
n=20	10.12	11.94	9.22	9.46
n=30	3.68	4.13	3.66	3.69

Table 17: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=4

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9363	0.9486	0.7411
n=15	0.9372	0.9471	0.7432
n=20	0.9396	0.9518	0.7508
n=30	0.9472	0.9529	0.7533

Table 18: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=4

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9387	0.9337	0.9182	0.9363
n=15	0.9392	0.9367	0.9199	0.9372
n=20	0.9467	0.9394	0.9273	0.9396
n=30	0.9489	0.9459	0.9396	0.9472

Table 19: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	21.49	22.89	24.69	22.19
n=15	7.72	8.39	8.92	8.22
n=20	1.79	1.85	2.24	1.84
n=30	0.90	1.13	1.4	0.97

Table 20: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=3, λ=4

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	88.17	97.16	90.51	87.26
n=15	31.84	36.82	34.52	33.84
n=20	7.24	8.44	7.82	7.52
n=30	4.62	5.41	5.12	5.33

Table 21: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=7

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9456	0.9362	0.7457
n=15	0.9491	0.9359	0.7452
n=20	0.9518	0.9407	0.7492
n=30	0.9532	0.9455	0.7567

Table 22: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=7

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9455	0.9387	0.9217	0.9456
n=15	0.9468	0.9386	0.9238	0.9491
n=20	0.9482	0.9467	0.9307	0.9518
n=30	0.9496	0.9517	0.9409	0.9532

Table 23: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	28.12	29.1	31.31	27.27
n=15	7.44	7.38	8.56	7.23
n=20	1.89	1.91	2.45	1.93
n=30	0.98	1.21	1.42	0.99

Table 24: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=3, λ=7

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	115.48	127.18	119.48	111.98
n=15	30.36	35.83	30.52	29.92
n=20	7.63	9.93	7.92	8.32
n=30	3.91	5.68	4.84	4.62

Table 25: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=10

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9299	0.9414	0.7518
n=15	0.9312	0.9439	0.7542
n=20	0.9372	0.9504	0.7618
n=30	0.9425	0.9568	0.7714

Table 26: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=3, λ=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9387	0.9294	0.9238	0.9299
n=15	0.9438	0.9317	0.9287	0.9352
n=20	0.9467	0.9360	0.9314	0.9372
n=30	0.9502	0.9409	0.9367	0.9425

Table 27: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	26.85	28.35	30.87	27.94
n=15	13.64	13.99	14.49	14.22
n=20	1.79	1.58	1.85	1.64
n=30	1.09	1.26	1.39	1.14

Table 28: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=3, λ=10

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	109.56	123.17	115.62	108.82
n=15	55.64	58.92	57.82	57.92
n=20	7.32	7.93	6.94	7.49
n=30	4.48	6.31	5.36	4.58

Table 29: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=4

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9412	0.9447	0.7682
n=15	0.9448	0.9437	0.7677
n=20	0.9437	0.9467	0.7712
n=30	0.9521	0.9538	0.7734

Table 30: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=4

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9411	0.9433	0.9181	0.9412
n=15	0.9442	0.9489	0.9199	0.9448
n=20	0.9482	0.9497	0.9256	0.9437
n=30	0.9512	0.9527	0.9372	0.9521

Table 31: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	45.69	46.25	48.14	44.16
n=15	14.83	15.69	16.24	14.11
n=20	3.58	3.50	4.39	3.61
n=30	1.67	1.68	2.23	1.71

Table 32: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=5, λ=4

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	185.17	192.13	187.36	179.7
n=15	61.31	66.9	63.72	59.5
n=20	15.31	18.5	15.07	15.5
n=30	6.84	8.96	7.34	6.9

Table 33: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=7

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9387	0.9317	0.7591
n=15	0.9428	0.9342	0.7583
n=20	0.9501	0.9397	0.7592
n=30	0.9534	0.9436	0.7618

Table 34: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=7

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9402	0.9327	0.9217	0.9387
n=15	0.9433	0.9391	0.9225	0.9428
n=20	0.9471	0.9443	0.9233	0.9501
n=30	0.9501	0.9488	0.9366	0.9534

Table 35: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	61.79	62.14	64.27	62.51
n=15	16.29	17.83	19.39	17.22
n=20	3.88	3.93	4.57	3.64
n=30	1.88	1.83	2.84	1.94

Table 36: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=5, λ=7

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	250.12	260.18	252.24	253.17
n=15	67.37	79.36	73.28	70.84
n=20	16.52	19.24	16.68	15.44
n=30	7.48	11.28	7.24	7.72

Table 37: 95% marginal and joint coverage probabilities of θ and λ from the projection pursuit regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=10

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR λ	JOINT COV. PROB.
n=7	0.9472	0.9462	0.7712
n=15	0.9491	0.9486	0.7723
n=20	0.9482	0.9497	0.7738
n=30	0.9523	0.9533	0.7756

Table 38: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma density with θ=5, λ=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9391	0.9458	0.9237	0.9472
n=15	0.9422	0.9447	0.9242	0.9491
n=20	0.9463	0.9456	0.9255	0.9482
n=30	0.9492	0.9491	0.9387	0.9523

Table 39: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	44.79	46.81	48.17	43.82
n=15	18.26	19.86	21.56	17.13
n=20	4.21	4.56	5.28	3.47
n=30	1.79	1.98	2.37	1.82

Table 40: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the gamma with θ=5, λ=10

SAMPLE SIZE	INTEG. LIK.	PROFILE LIK.	COND. LIK.	PROJ.PUR.R.
n=7	182.18	194.16	190.56	178.36
n=15	75.12	85.36	78.2	73.26
n=20	17.84	22.36	22.36	16.82
n=30	6.53	7.16	9.48	7.32

Example-2: We also generated data from Pareto distribution to compare how the procedure works on this heavy tailed distribution. The pareto probability distribution is a two parameter continuous probability distribution with density given by

$$f\left(y\right)=\frac{\theta K^{\theta}}{y^{\theta+1}},y\geq K \tag{4.6}$$

where the two parameters are $\theta > 1$ and $K > 0$. The pareto distribution is usually used to model the variation of income across a population. For the simulation we used θ=3,5, and K=10,20. We generated 10,000 observation from pareto distribution for each pair of θ and K values.

The likelihood function for the pareto density is;

$$L\left(\theta,K\right)=\prod_{i=1}^{n}\frac{\theta K^{\theta}}{y^{\theta+1}}$$

$$L\left(\theta,K\right)=\theta^{n}K^{n\theta}\prod_{i=1}^{n}\frac{1}{y_i^{\theta+1}}$$

Then the log-likelihood is;

$$l\left(\theta,K\right)=nlog\left(\theta\right)+n\theta log\left(K\right)-\left(\theta+1\right)\sum_{i=1}^{n}log\left(y_i\right)$$

since $l\left(\theta,K\right)$ is monotonically increasing with K, and $y\geq K$, maximum likelihood of K is $\widehat{K}=min\left(y_i\right)$.

To find the maximum likelihood estimator for the θ, we take the derivative of log-likelihood with respect to θ and set equal to zero.

$$\frac{\partial l}{\partial\theta}=\frac{n}{\theta}+nLog\left(K\right)-\sum_{i=1}^{n}log\left(y_i\right)=0$$

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n}\left(logy_i - log\widehat{K}\right)}$$

Table 41 provides the marginal coverage probabilities of θ and K from the projection pursuit method and also joint coverage probabilities of θ and K from (4.5)

Table 42 provides the coverage probabilities of θ(structural parameter) from the modified likelihood methods(profile, conditional and integrated likelihood methods) and marginal coverage probabilities of θ from the projection pursuit method.

In table 43 we present the mean squared error(MSE) of the θ from the modified likelihoods and from projection pursuit regression method. And Table 44 presents average interval length for the modified likelihood methods and average interval length for the PPR method.

Table 41: 95% marginal and joint coverage probabilities of θ and K from the projection pursuit regression(PPR) of approximated log-likelihood of the Pareto density with θ=3, K=10

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR K	JOINT COV. PROB.
n=7	0.9381	0.9288	0.7817
n=15	0.9397	0.9373	0.7814
n=20	0.9451	0.9422	0.7824
n=30	0.9504	0.9489	0.7848

Table 42: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=3, K=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK	PROFILE LIK.	PROJ. PUR.
n=7	0.9414	0.9369	0.9212	0.9381
n=15	0.9422	0.9361	0.9222	0.9397
n=20	0.9471	0.9404	0.9241	0.9451
n=30	0.9513	0.9454	0.9358	0.9504

Table 43: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.69	0.72	1.28	0.71
n=15	0.37	0.42	0.45	0.35
n=20	0.08	0.08	0.18	0.07
n=30	0.032	0.04	0.06	0.037

Table 44: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=3, K=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	2.92	3.01	5.32	2.91
n=15	1.52	1.74	1.89	1.42
n=20	0.37	0.38	0.69	0.35
n=30	0.13	0.16	0.26	0.14

Table 45: 95% marginal and joint coverage probabilities of θ and K from the projection pursuit regression(PPR) of approximated log-likelihood of the Pareto density with θ=3, K=20

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR K	JOINT COV. PROB.
n=7	0.9442	0.9312	0.7717
n=15	0.9448	0.9351	0.7748
n=20	0.9467	0.9393	0.7761
n=30	0.9490	0.9447	0.7784

Table 46: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=3, K=20

SAMPLE SIZE	INTEG. LIK.	COND. LIK	PROFILE LIK.	PROJ. PUR.
n=7	0.9410	0.9356	0.9282	0.9442
n=15	0.9437	0.9381	0.9297	0.9448
n=20	0.9478	0.9442	0.9337	0.9467
n=30	0.9506	0.9459	0.9372	0.9490

Table 47: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	2.41	2.58	3.17	2.36
n=15	1.33	1.47	2.58	1.39
n=20	0.25	0.26	0.52	0.25
n=30	0.09	0.12	0.15	0.13

Table 48: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=3, K=20

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	9.75	10.53	12.89	9.65
n=15	5.32	5.99	8.75	5.67
n=20	1.16	1.38	2.56	1.48
n=30	0.39	0.48	0.58	0.50

Table 49: 95% marginal and joint coverage probabilities of θ and K from the projection pursuit regression(PPR) of approximated log-likelihood of the Pareto density with θ=5, K=10

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR K	JOINT COV. PROB.
n=7	0.9414	0.9337	0.7817
n=15	0.9453	0.9372	0.7819
n=20	0.9482	0.9414	0.7881
n=30	0.9511	0.9462	0.7918

Table 50: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=5, K=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK	PROFILE LIK.	PROJ. PUR.
n=7	0.9422	0.9382	0.9182	0.9414
n=15	0.9469	0.9391	0.9293	0.9453
n=20	0.9487	0.9418	0.9337	0.9482
n=30	0.9514	0.9449	0.9402	0.9511

Table 51: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.22	0.25	0.29	0.24
n=15	0.09	0.110	0.17	0.08
n=20	0.04	0.037	0.079	0.038
n=30	0.014	0.015	0.024	0.012

Table 52: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=5, K=10

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	0.87	0.98	1.22	1.09
n=15	0.36	0.45	0.59	0.39
n=20	0.14	0.16	0.31	0.14
n=30	0.06	0.062	0.09	0.07

Table 53: 95% marginal and joint coverage probabilities of θ and K from the projection pursuit regression(PPR) of approximated log-likelihood of the Pareto density with θ=5, K=20

SAMPLE SIZE	COV.PROB. FOR θ	COV. PROB.FOR K	JOINT COV. PROB.
n=7	0.9414	0.9391	0.7881
n=15	0.9436	0.9441	0.7897
n=20	0.9486	0.9468	0.7917
n=30	0.9517	0.9511	0.7942

Table 54: 95% Marginal coverage probabilities of θ from the modified likelihood methods and marginal coverage probability of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=5, K=20

SAMPLE SIZE	INTEG. LIK.	COND. LIK	PROFILE LIK.	PROJ. PUR.
n=7	0.9461	0.9442	0.9227	0.9414
n=15	0.9468	0.9462	0.9233	0.9436
n=20	0.9501	0.9474	0.9261	0.9486
n=30	0.9507	0.9508	0.9382	0.9517

Table 55: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.86	0.84	1.931	0.83
n=15	0.32	0.31	0.37	0.31
n=20	0.26	0.26	0.43	0.25
n=30	0.042	0.05	0.07	0.06

Table 56: Average interval length of θ from the modified likelihood methods and average interval length of θ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the pareto density with θ=5, K=20

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	3.43	3.37	6.22	3.28
n=15	1.29	1.25	1.73	1.33
n=20	1.05	1.20	1.64	1.22
n=30	0.17	0.22	0.36	0.25

Example-3: We have another application using the log-normal density, which does not belong to the exponential family. Let $y_1, ..., y_n$ be a random sample from the log-normal distribution with density

$$f\left(y, \mu, \sigma^2\right) = \frac{1}{y\sigma\sqrt{2\pi}} \exp -\frac{1}{2\sigma^2}\left(logy - \mu\right)^2, \sigma^2 > 0$$

then, It can be shown that

$$f_L\left(y; \mu, \sigma\right) = \frac{1}{y} f_N\left(logy; \mu, \sigma\right)$$

where f_L is the density function of the log-normal and f_N is the density of the normal distribution. Therefore we can write the log-likelihood of log-normal distribution as;

$$Log_L\left(\mu, \sigma | y_1, ..., y_n\right) = -\sum_{i=1}^{n} log\left(y_i\right) + Log_N\left(\mu, \sigma | logy_1, ..., logy_n\right)$$

$$= constant + Log_N\left(\mu, \sigma | logy_1, ..., logy_n\right)$$

where Log_L is the log-likelihood of log-normal density and Log_N is the log-likelihood of normal density. Since the first term is constant with respect to μ and σ, L_N and L_L reach their maximum with same μ and σ, hence maximum likelihood estimates for μ and σ are;

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} log\left(y_i\right)}{n}$$

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}\left(logy_i - \widehat{\mu}\right)^2}{n}$$

Tables, 57 provides the marginal coverage probabilities of μ and σ^2 from projection pursuit method and also joint coverage probabilities of μ and σ^2 from (4.5)

Tables 58 provides the coverage probabilities of μ from the modified likelihood methods(Profile, conditional and integrated likelihood methods) and coverage probabilities of μ from the projection pursuit method.

In table 59 we present the mean squared error(MSE) of the μ from the modified likelihoods and from projection pursuit regression method. And Table 60 presents average interval length for the modified likelihood methods and average interval length for the PPR method.

Table 57: 95% marginal and joint coverage probabilities of μ and σ^2 from the projection pursuit regression(PPR) of approximated log-likelihood of the log-normal density with μ=2, σ^2=1

SAMPLE SIZE	COV.PROB. FOR μ	COV. PROB.FOR σ^2	JOINT COV. PROB.
n=7	0.9397	0.9402	0.7852
n=15	0.9418	0.9433	0.7883
n=20	0.9482	0.9467	0.7891
n=30	0.9532	0.9526	0.7912

Table 58: 95% Marginal coverage probabilities of μ from the modified likelihood methods and coverage probabilities of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal with μ=2, σ^2=1

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9433	0.9414	0.9289	0.9397
n=15	0.9428	0.9433	0.9291	0.9418
n=20	0.9468	0.9482	0.9317	0.9482
n=30	0.9502	0.9529	0.9427	0.9532

Table 59: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.17	0.17	0.23	0.18
n=15	0.12	0.12	0.18	0.12
n=20	0.08	0.07	0.14	0.078
n=30	0.05	0.051	0.06	0.052

Table 60: Average interval length of μ from the modified likelihood methods and average interval length of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal density with μ=2, σ^2=1

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	0.68	0.67	0.91	0.71
n=15	0.44	0.45	0.76	0.47
n=20	0.32	0.28	0.53	0.31
n=30	0.20	0.21	0.27	0.22

Table 61: 95% marginal and joint coverage probabilities of μ and σ^2 from the projection pursuit regression(PPR) of approximated log-likelihood of the log-normal density with μ=2, σ^2=5

SAMPLE SIZE	COV.PROB. FOR μ	COV. PROB.FOR σ^2	JOINT COV. PROB.
n=7	0.9489	0.9379	0.7748
n=15	0.9507	0.9391	0.7893
n=20	0.9519	0.9427	0.7841
n=30	0.9513	0.9449	0.7862

Table 62: 95% Marginal coverage probabilities of μ from the modified likelihood methods and marginal coverage probabilities of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal with μ=2, σ^2=5

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9448	0.9462	0.9271	0.9489
n=15	0.9421	0.9455	0.9318	0.9507
n=20	0.9452	0.9488	0.9314	0.9519
n=30	0.9492	0.9518	0.9412	0.9513

Table 63: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	4.56	4.78	5.87	4.23
n=15	3.48	3.81	4.38	3.62
n=20	1.47	1.61	2.3	1.58
n=30	1.36	1.42	1.53	1.24

Table 64: Average interval length of μ from the modified likelihood methods and average interval length of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal density with μ=2, σ^2=5

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	18.22	19.25	23.46	16.90
n=15	13.90	15.22	17.50	14.46
n=20	5.86	6.42	8.0	6.30
n=30	5.42	5.66	6.24	4.94

Table 65: 95% marginal and joint coverage probabilities of μ and σ^2 from the projection pursuit regression(PPR) of approximated log-likelihood of the log-normal density with μ=5, σ^2=1

SAMPLE SIZE	COV.PROB. FOR μ	COV. PROB.FOR σ^2	JOINT COV. PROB.
n=7	0.9386	0.9467	0.7812
n=15	0.9396	0.9482	0.7814
n=20	0.9412	0.9492	0.7834
n=30	0.9447	0.9526	0.7842

Table 66: 95% Marginal coverage probabilities of μ from the modified likelihood methods and marginal coverage probabilities of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal with μ=5, σ^2=1

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9425	0.9397	0.9278	0.9386
n=15	0.9419	0.9411	0.9266	0.9396
n=20	0.9433	0.9423	0.9308	0.9412
n=30	0.9478	0.9432	0.9438	0.9447

Table 67: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.17	0.148	0.26	0.15
n=15	0.14	0.153	0.19	0.124
n=20	0.05	0.06	0.09	0.55
n=30	0.039	0.05	0.06	0.04

Table 68: Average interval length of μ from the modified likelihood methods and average interval length of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal density with μ=5, σ^2=1

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	0.67	0.63	0.90	0.61
n=15	0.54	0.58	0.67	0.52
n=20	0.20	0.21	0.36	0.23
n=30	0.12	0.14	0.31	0.15

Table 69: 95% marginal and joint coverage probabilities from the projection pursuit regression of log-likelihood of the log-normal density with μ=5, σ^2=5

SAMPLE SIZE	COV.PROB. FOR μ	COV. PROB.FOR σ^2	JOINT COV. PROB.
n=7	0.9461	0.9387	0.7767
n=15	0.9488	0.9351	0.7792
n=20	0.9493	0.9422	0.7781
n=30	0.9517	0.9437	0.7812

Table 70: 95% Marginal coverage probabilities from modified likelihoods and marginal coverage probabilities of μ from the PPR method of log-likelihood of log-normal with μ=5, σ^2=5

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	0.9414	0.9412	0.9308	0.9461
n=15	0.9428	0.9428	0.9341	0.9488
n=20	0.9478	0.9451	0.9381	0.9493
n=30	0.9512	0.9527	0.9451	0.9517

Table 71: Mean squared error(MSE)of the structural parameter(θ) from the modified likelihood methods and from the PPR method.

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROFILE LIK.	PROJ.PUR.R.
n=7	4.45	4.31	5.81	4.1
n=15	2.91	3.12	4.62	2.82
n=20	1.29	1.34	3.24	1.37
n=30	0.89	0.95	1.18	0.93

Table 72: Average interval length of μ from the modified likelihood methods and average interval length of μ from the Projection Pursuit Regression(PPR) method of approximated log-likelihood of the log-normal density with μ=5, σ^2=5

SAMPLE SIZE	INTEG. LIK.	COND. LIK.	PROF. LIK.	PROJ.PUR.R.
n=7	17.8	17.24	23.24	16.4
n=15	11.64	12.48	18.48	11.28
n=20	5.14	5.35	10.77	5.49
n=30	3.56	3.80	4.26	3.70

5.0 APPROXIMATION OF LOG-POSTERIOR DENSITY

In the classical approach the parameter, θ, is known to be unknown, but a fixed quantity. A random sample $y_1, ..., y_n$ is drawn from the population and knowledge about the value of θ is obtained based on the observed values in the sample. However in the Bayesian approach, θ is considered to be a quantity whose variation can be described by a probability distribution. This probability distribution is called the prior distribution. The prior distribution is subjective and based on the the experimenter's belief. A sample is drawn from the population and the prior distribution is updated with this sample information. The updated distribution is called the posterior distribution. The posterior density describes what is known about θ given the data Y.

If we denote the prior distribution by $p(\theta)$ and the likelihood $L(\theta|Y)$, the posterior density for θ is given as

$$p(\theta|Y) = cp(\theta)\, L(\theta|Y)\,, \tag{5.1}$$

where

$$c^{-1} = \int_{\Theta} p(\theta)\, L(\theta|Y)\, d\theta,$$

when θ is continuous. When θ is discrete, we have $c^{-1} = \sum_{\Theta} p(\theta)\, L(\theta|Y)$.

However, often there is no closed form for $\int_{\Theta} p(\theta)\, L(\theta|Y)\, d\theta$, so that posterior distribution should be approximated. We will use natural cubic splines to approximate the log-posterior densities as we did before for log-likelihoods. We have two different approaches to approximate the log-posterior densities. From (5.1) we have that the shape of the posterior density is determined by the product of the likelihood and the prior distribution. We can rewrite (5.1) as

$$logp\left(\theta|Y\right) = logp\left(\theta\right) + logL\left(\theta|Y\right)$$

In the first approach (method 1), we approximate the log-prior and the log-likelihood at the same time. In other words we consider the sum of log-prior and log-likelihood value as the combined response and we approximate the log-posterior density by the natural cubic spline.

In the second approach (method 2), we approximate the log-likelihood first and then we add the value of the log-prior to this approximation. To assess the approximation of the log-posterior, we will calculate the coverage probabilities which will depend on the highest density region of approximated log-posterior densities.

The goal of obtaining a smallest confidence set with a specified coverage probability can also be obtained using Bayesian criteria. Indeed Highest Posterior density (HPD) credible region attempts to capture the smallest region of the parameter space which contains most of the mass of the posterior distribution. The HPD region is motivated as the region where the probability density of every point inside the region is at least as large as that of any point outside the region.

More formally, if we have a posterior distribution $p\left(\theta|Y\right)$, we would like to find the region $R\left(Y\right)$ that satisfies

$$(i) \int_{R(y)} p\left(\theta|y\right) dy = 1 - \alpha$$

$$(ii) SizeR\left(y\right) \leq SizeR^{'}\left(y\right)$$

for any region $R'(y)$ satisfying $\int_{R'(y)} p(\theta|Y) \geq 1 - \alpha$.

We have the following algorithm to approximate the log-posterior density and calculating the coverage probabilities from the highest density region of approximated log-posterior density. This procedure is similar to the one that we had in chapter 3 for log-likelihoods.

1) Approximate the log-posterior density by natural cubic spline with two different methods. In the first approach treat the addition of log-likelihood and log-prior value as the combined response value and the chosen parameter values as the predictors and then approximate by the natural cubic spline. For the second approach first approximate the log-likelihood by the natural cubic spline and then add the value of the log-prior value to this approximation.

2) Find the fitted values of the approximated log-posterior density for chosen θ values.

3) Treat those fitted values as the height and find the total area under the approximated function by summing those heights.

4) Divide each height (fitted value) by the total area to make the total area 1 under the spline function.

5) Next step is finding the biggest height (mode) and add the next biggest height (fitted value) and continue this until we reach the given confidence level.

6) Claim the corresponding smallest and biggest parameter values as the lower and upper bound of the highest density region of approximated log-posterior function for given confidence level.

7) Finally, calculate the coverage probability by using the upper and lower bound of the highest posterior density of approximated log-posterior density.

For the simulation, we generated random samples from an exponential distribution with known parameter (θ) for which the likelihood is:

$$f\left(y|\theta\right) = \theta^{-n} \exp^{-\sum_i y_i/\theta}$$

If we re-parameterize the likelihood in terms of $\Lambda = \Theta^{-1}$

$$f\left(y|\lambda\right) = \lambda^{n} exp^{-\lambda \sum_i y_i},$$

and the prior distribution for $\Lambda = \theta^{-1}$ is $Gamma\left(\alpha, \beta\right)$

$$g\left(\lambda|\alpha, \beta\right) = \frac{1}{\beta^{\alpha}\Gamma\left(\alpha\right)}\lambda^{\alpha-1} exp{-\lambda|\beta},$$

then the posterior distribution for λ given Y is $Gamma\left(\alpha^{*}, \beta^{*}\right)$:

$$g\left(\lambda|y, \alpha, \beta\right) = g\left(\lambda|\alpha^{*}, \beta^{*}\right) = \frac{1}{\left(\beta^{*}\right)^{\alpha^{*}}\Gamma\left(\alpha^{*}\right)}\lambda^{\alpha^{*}-1} exp^{-\lambda|\beta}$$

where

$$\alpha^{*} = \alpha + n$$

and

$$\beta^{*} = \frac{1}{\beta^{-1} + \sum y_i}$$

Table 73 presents 90% coverage probability results, from the highest posterior density region of approximated log-posterior density. Here the second method clearly performs better than the first method in terms of the coverage probabilities. We also presented highest posterior density-raindrop plot of approximated log-posterior densities for different confidence levels and sample sizes.

Table 73: 90% coverage probabilities from the approximated highest posterior density region

SAMPLE SIZE	Method-1	Method-2
n=10	0.8611	0.8759
n=20	0.8723	0.8814
n=30	0.8801	0.9027

Figures 5-8 provide some useful information about the posterior densities that we approximated. For example in, figure 5, we can see that the mode is around 1.9 and the upper and lower bounds of the 95% highest posterior credible region are 1.4 and 2.3. From the shape of the raindrop plot, we can also conclude that the posterior density is fairly symmetric. We have similar interpretations for the other raindrop plots.

Figure 5: 95% Highest Posterior Density-raindrop plot from method1

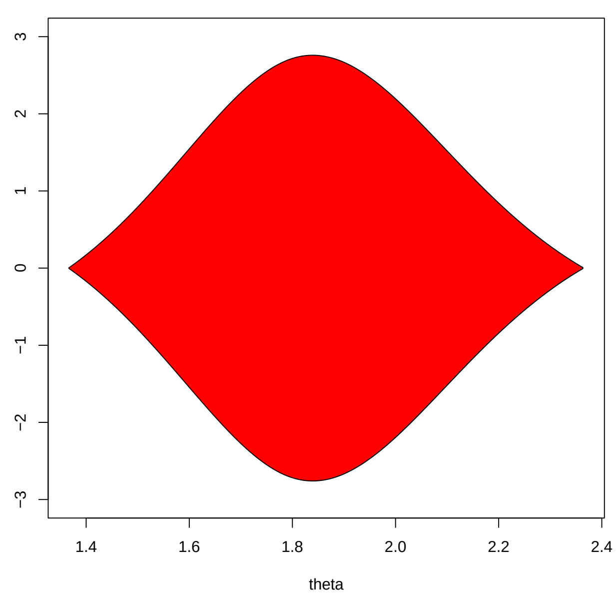


Figure 6: 70% HPD-raindrop plot from method1

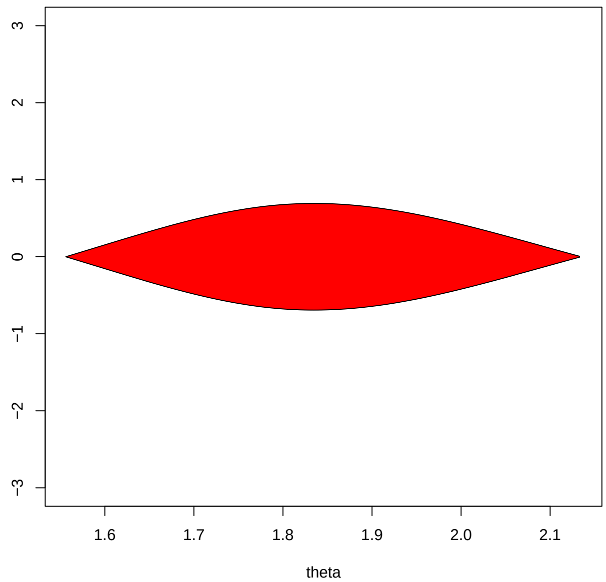


Figure 7: 95% HPD-raindrop plot from method2

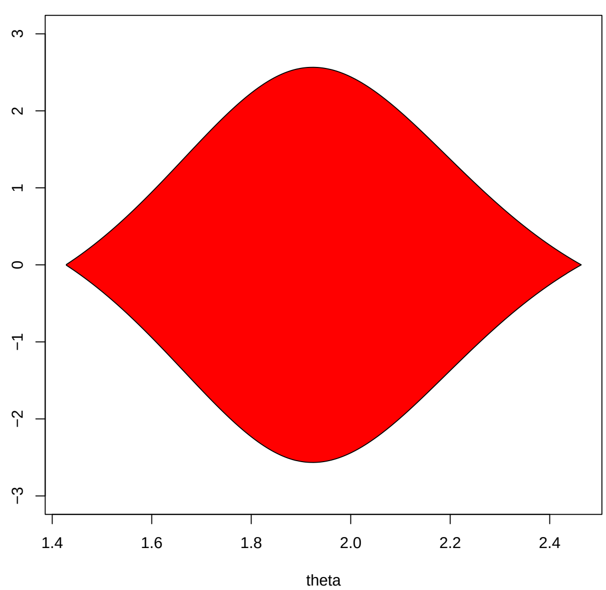
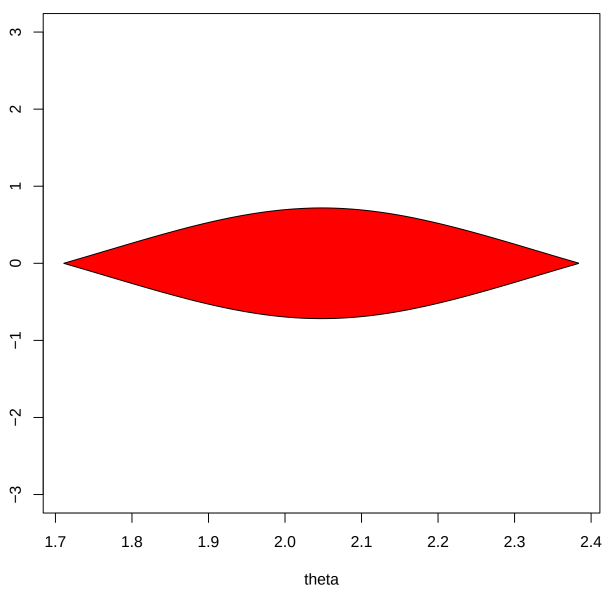


Figure 8: 70% HPD-raindrop plot from method2



6.0 DISCUSSION AND FUTURE WORK

To assess the quality of the approximation by the cubic spline function, we have determined three criteria:

1) Coverage probability of the highest density region of the approximated log-likelihood function.

2) Mean squared error(MSE) of the estimator obtained as the maximum over the parameter of the approximation of the log-likelihood

3) Average interval length of the highest density region from the approximated log-likelihood.

We applied natural cubic spline approximations on log-likelihood of the gamma, Pareto and log-normal densities and we presented the results in chapter 4. Now we briefly summarized our findings for each density function.

For the two parameter gamma distribution(θ is structural parameter), overall the integrated likelihood method provides the closest coverage probabilities to the nominal level among the methods that we consider. Marginal coverage probabilities from the conditional likelihood and marginal coverage probabilities from the projection pursuit regression method produce close coverage probability numbers, but they are not as close to the nominal level as the integrated likelihood. The profile likelihood seems to be the weakest one in terms of the marginal coverage probabilities.

However, it is not enough to evaluate the performance of these methods by just analyzing their coverage probabilities. Besides the coverage probabilities we calculated the mean squared error of the parameter and the average length of the highest density interval for each method. Clearly, the method with close coverage probabilities to the nominal level and low MSE and short average interval length will be the ideal one.

Furthermore, for the approximation of log-likelihood of the gamma density we note that the projection pursuit regression provides the smallest MSE and shortest average interval length for small samples(n=7). For larger sample sizes, integrated likelihood, conditional likelihood and projection pursuit regression perform close to each other in terms of coverage probabilities, MSE and the average interval length.

For the approximation to the log-likelihood of pareto density, the integrated likelihood method provides the closest coverage probabilities to the nominal level. Unlike the case of the gamma distribution, the projection pursuit regression method tended to perform better than conditional likelihood in terms of coverage probabilities. As it was in the gamma density, the PPR method performs slightly better than the integrated likelihood and conditional likelihood in terms of MSE and average interval length for the small samples. Results for the profile likelihood are not competitive with the other methods especially for the small samples, but get better as the sample size increases.

The results for the log-normal density are consistent with the results from the gamma density and pareto density. The Integrated likelihood method outperforms the other methods in terms of coverage probabilities. As it was in the gamma density, the overall coverage probability of μ from the conditional likelihood and the marginal coverage probabilities from the projection pursuit regression methods are close to each other. Projection pursuit regression continues to perform slightly better than the modified likelihood methods in terms of MSE and average interval length for small samples.

Our simulation results indicate that projection pursuit regression performs well as frequently as the integrated likelihood, and the conditional likelihood, and performs better than the profile likelihood. The PPR method seems to be the best choice for small samples to obtain an accurate point estimator of the parameter.

Although the integrated likelihood method outperforms the projection pursuit regression (PPR) with respect to coverage probabilities, PPR has some important advantages over the modified likelihood methods in that it is always available and easy to determine the projection pursuit function by most statistical packages. On the other hand, the integrated likelihood might be very difficult computationally, especially for the complex models.

For our future work, we will continue to evaluate the performance of raindrop plots and confidence intervals derived from the different methods of approximating log-likelihoods and dealing with nuisance parameters in the context mixture of exponential family. Dealing with mixture distributions will be challenging especially in terms of computations.

In Chapter 1, we calculated coverage probabilities to judge the accuracy of the integrated likelihood, profile likelihood and conditional likelihood methods. We may include some other likelihood methods such as modified profile likelihood, marginal likelihood and signed likelihood method in such comparisons.

In chapter 3, we showed how to get the raindrop plots for the one parameter distributions, and extended this idea to show the raindrop plots on approximated log-posterior densities. We believe that constructing raindrop plots of approximated log-likelihoods with more than one variable can be another challenging application.

Since the natural cubic splines and cubic B-splines provide very similar results we used natural cubic splines for our applications. However, we may try some other semi-parametric methods such as smoothing splines, and compare efficiency of those methods with the ones used in the present study.

In chapter 4, we compared the modified likelihood methods with the projection pursuit method on log-likelihood of gamma density, Pareto density and log-normal density. We can extend our analysis to other two parameter densities, such as the Weibull distribution.

From the Bayesian point of view, we investigated how well the cubic spline approximations lead to accurate approximations of highest posterior credible regions. In our simulation we selected a simple likelihood function (exponential distribution) and prior distribution(gamma density). We believe that application of more complicated likelihood functions(with several parameters) will make cubic spline approximation more attractive than the modified likelihood methods.

BIBLIOGRAPHY

[1] Anderson, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Society B* **32**, 283-301.

[2] Bahadur, R.R. (1964). On Fisher's bound for asymptotic variances.*Annal of Mathematical Statistics* **35**, 1545-52.

[3] Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrica*, **70**, 343-365.

[4] Barndorff-Nielsen, O. (1988). Parametric Statistical Models and Likelihood.*Lecture Notes in Statist.* **50**. Springer, New York.

[5] Barrowman, N.J. and Myers, R.A (2003). Raindrop plots: A new way to display collections of likelihoods and distributions.*The American Statistician*, **57**, No.4 268-274.

[6] Basu, D. (1978). On the elimination of nuisance parameters.*Journal of the American Statistical Association*, **72**, 355-66.

[7] Berger, J.O., Liseo, B. and Wolpert, R.L.(1999). Integrated likelihood methods for eliminating nuisance parameters.*Statistical Science*. Vol. **14**, No. 1, 1-28.

[8] Buse, A. and Lim, L. (1977). Cubic splines as a special case of restricted least squares.*Journal of the American Statistical Association*, Vol **72** ,No.357, 64-68.

[9] Cramer, H. (1946).*Mathematical methods of statistic. Princeton*: Princeton University Press.

[10] Chung, H, Loken, E. and Schaffer, L. (2004). Diffuculties in drawing inferences with finite-mixture models: A simple example with a simple solution.*The American Statistician*, Vol. **58**, No.2.

[11] Cooper, H. and Hedges, L.V. (1994) *The handbook of research synthesis*. New York: Russel Sage Foundation.

[12] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49** 1-39.

[13] Efron, B and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected fisher information (with discussion). *Biometrika* **65**, 457-487.

[14] Eubank, R.L. (1994) A simple smoothing spline. The American Statistician, vol, **48**, No.2, 103-106.

[15] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics.*Phi. Trans. Roy. Soc. A***222**, 309-68.

[16] Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion).*Technometrics* **31** 3-39.

[17] Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion).*Ann.Statist.***19** 1-141

[18] Gleser, L.J. (1998). Assessing uncertainty in measurement.*Statistical Science*, Vol. **13**, No.3, 277-290.

[19] Goutis, C. and Casella, G. (1999). Explaining the saddlepoint approximation.*The American Statistician* Vol. 53, No.3 216-224.

[20] Hansen, M. (1994). Extended Linear models, multivariate splines and ANOVA. Ph.D. dissertation, Univ. California, Berkeley.

[21] Hansen, M, Kooperberg, C and Sardy, S. Triogram models.*Journal of the American Statistical Association*, Vol.**93**, No. 441,101-119.

[22] Hansen H.M and Kooperberg C. (2002). Spline adaptation in extended linear models. *Statistical Science*, Vol **17**, No.1, 2-51.

[23] Hodges, J.S. Assessing the accuracy of normal approximation. *Journal of the American Statistical Association*, Vol **82**, No.397,p.149.

[24] Huang, J.Z. (2001). Concave extended linear modeling: A theoretical synthesis. *Statistical synthesis. Statist.Sinica* **11** 173-197

[25] Huzurbazar, S. (1999). Practical Saddlepoint Approximation. *American Statistician*, **53**, 216-224.

[26] Hyndman, R.J. (1996). Computing and graphing highest density regions. *The American Statistician*. **50**, 120-126.

[27] Kalbfleisch, J.D. and Sprott, D.A (1970). Application of likelihood methods to models involving large numbers of parameters.*Journal of the Royal Statistical Society B* **32**, 175-208.

[28] Kalianpur, G and Rao, C. R. (1955). On Fisher's lower bound to the asymptotic variance of a consistent estimate. *Sankhya* **15**, 331-42.

[29] Kooperberg, C. and Stone, C.J. (1991). A study of logspline density estimation. *Comput. Statist. Data Anal.***12** 327-347.

[30] Kooperberg, C. and Stone, C.J. (1992). Logspline density estimation for censored data. *J. Comput. Graph. Statist.* **1** 301-328.

[31] McLachlan, G.J. and Peel, D. (2000) *Finite mixture models*. New york: Jhon Wiley Sons. Miettinen, O.S. (1985), Theoretical Epidemology, New York: Wiley.

[32] Reid, N. (1988). Saddlepoint Methods and statistical Inference(with discussion). *statistical science* **3**, 213-238.

[33] Rosen, O., Jiang, W., and Tanner, M.A. (2000). Mixtures of marginal models. *Biometrica* **87**, 391-404.

[34] Severini, T.A. (1998b). Likelihood functions for the elimination of nuisance parameters. *Biometrica* **85**, 403-411.

[35] Severini, T.A. (2000). *Likelihood methods in statistics*. New York: Oxford.

[36] Severini, T. A. (1998a) An approximation to the modified profile likelihood function. *Biometrica* **85**, 403-11.

[37] Stone, C, J. (1990) Large-sample inference for Log-spline models. *The Annals of Statistics*, Vol. **18**, No.2, 717-741.

[38] Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118-184.

[39] Stone, C.J, Hansen, M.H, Kooperberg C. and Young K.T (1997). Polynomial splines and their tensor products in extended linear modeling. *The Annals of Statistics*, Vol.**25**, No. 4 1371-1425.

[40] Sweeting, T. (1995). A framework for Bayesian and likelihood approximations in statistics. *Biometrica* **82** 25-36.

[41] Tanner, M.A. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions*, 3rd edition. New York: Springer.

[42] Van der Vaart , A. W. (1998) Estimating a real parameter in a class of semiparametric models. *Annals of Statistics* **16**, 1450-74.

[43] Wald, A. (1949). Note on the consistency of the maximum likelihood estimator. *Annals of Mathematical Statistics* **20**, 595-601.

[44] Wold, S. (1974). Spline functions in data analysis. *Technometrics*. Vol **16**, No. 1 1-11.