Relative Likelihood Differences to Examine Asymptotic Convergence: A Bootstrap Simulation Approach

Milan Bimali* and Michael Brimacombe
Department of Biostatistics, University of Kansas Medical Center, Kansas City, 66160, KS, USA

Abstract

Maximum likelihood estimators (mle) and their large sample properties are extensively used in descriptive as well as inferential statistics. In the framework of the large sample distribution of the mle, it is important to know the relationship between the sample size and asymptotic convergence i.e. for what sample size does the mle behave satisfactorily "attaining asymptotic normality." Previous work has discussed the undesirable impact of using large sample approximations of the mles when such approximations do not hold. It has been argued that relative likelihood functions should be examined before making inferences based on the mle. Little has been explored regarding the appropriate sample size that would allow the mle achieve asymptotic normality from a relative likelihood perspective directly. Our work proposes a bootstrap/simulation based approach in examining the relationship between sample size and the asymptotic behaviors of the mle. We propose two measures of the convergence of observed relative likelihood function to the asymptotic relative likelihood function namely: differences in areas and dissimilarity in shape between the two relative likelihood functions. These two measures were applied to datasets from the literature as well as simulated datasets.

Keywords: Relative likelihood functions; Bootstrap; Sample size; Divergence; Exponential family; Convergence

Introduction

"Likelihood" is arguably the most pronounced terminology in the statistical realm and was defined and popularized by the eminent geneticist and statistician Fisher (1922) [1-4]. The likelihood function is a function of model parameter(s) based on a given set of data and a pre-defined probability density function (pdf). The likelihood function can be formally defined as follows:

If \( f(x | \theta) \) is the joint pdf (pmf) of the sample \( X = \{X_1, ..., X_n\} \) based on the random variables \( X, x \) iid, then the likelihood function of \( \theta \) is given by:

\[
L(\theta | X) = f(x | \theta) = c \prod_{i=1}^{n} f(x_i | \theta)
\]

where \( c \) is a constant with respect to \( \theta \).

A key point often reiterated in textbooks is that the likelihood function is a function of \( \theta \) and not to be viewed as a probability density itself [5]. However, the shape of the likelihood function relative to its mode is often of interest in estimating \( \theta \). Likelihood functions can be mathematically constructed for most statistical distributions; but maximum likelihood estimators may not always have closed form [6]. Nevertheless most of the distributions commonly used allow for the computation of maximum likelihood estimators either analytically, numerically or graphically. Several properties of maximum likelihood estimators such as asymptotic normality, invariance, and ease of computation have made maximum likelihood estimators popular [7]. In this paper we assume \( \theta \) is a scalar throughout.

The large sample distribution of the maximum likelihood estimators is often used for inferential purposes. If \( \hat{\theta} \) is the mle of \( \theta \), then \( \hat{\theta} \sim N\left(\theta, I^{-1}(\theta)\right) \) where \( I(\hat{\theta}) = E\left(\frac{\partial}{\partial \theta} \log f(X | \theta)\right) \) is the Fisher Information evaluated at \( \hat{\theta} \). In situations where the computation of the expectation for the Hessian of the log-likelihood is analytically tractable, \( J(\hat{\theta}) = \left\{ \frac{\partial}{\partial \theta} \log f(X | \theta) \right\}_{p \times p} \), the observed Fisher Information, has been used as an approximation in the computation of \( I(\hat{\theta}) \) [8].

A common question that often arises in statistics is in regard to sample size. In the framework of the large sample distribution of the mle, we are interested in knowing for what sample size the mle behaves satisfactorily, attaining the asymptotic normal distribution. Put a different way, does the existing sample size allow us to use the large sample properties of the mle with confidence? If not, what would be an ideal sample size?

Sprott et al (1969) discussed some of the undesirable impacts of using large sample approximation of the mle when such approximations do not seem to hold [9] and suggested examining likelihood functions before making inferences based on mle. They demonstrate via an example from Bartholomew [10] that drawing inferences from the mle without first examining the likelihood functions can be misleading. Figure 1 gives the plot of the observed relative likelihood (likelihood functions scaled by its mode) as obtained from Bartholomew’s data and relative normal likelihood based on the large sample theory of mle. The plot shows that for a pre-specified value of the relative likelihood, the range of \( \theta \) s can be in complete disagreement between the two likelihood functions. A relative likelihood of 10 % or higher, the ranges are roughly(20,110) and (7.81) for the relative and relative normal likelihood functions. A relative likelihood of 10 % or higher, the ranges are roughly(20,110) and (7.81) for the relative and relative normal likelihood functions [9], approximately a 17% drop in coverage. Sprott (1969) also demonstrated that transformation of the mle can help achieve the asymptotic normality with smaller sample sizes. However...
little has been explored regarding the appropriate sample size that would allow the mle to achieve asymptotic normality from a relative likelihood perspective directly. However little has been explored regarding the appropriate sample size that would allow the mle to achieve asymptotic normality from a relative likelihood perspective directly. This work proposes a bootstrap/simulation based approach to the above question via the behavior and properties of the relative likelihood function. This work proposes a bootstrap/simulation based approach to the above question via the behavior and properties of the relative likelihood function. In particular we measure the proximity of the observed likelihood function to the likelihood function based on large sample properties, both of which are scaled here by their modes to have a maximum at one. The convergence measures proposed by the authors are(i) difference in area under the two relative likelihood functions and (ii) dissimilarity in the shape of the two likelihood functions (dissimilarity index). We propose that, for a given sample size, if the difference in the area under the two relative likelihood functions and the dissimilarity index between them are both close to 0, the asymptotic approximation of mle is satisfactorily achieved. To study the properties of these measures and related likelihood convergence, we use the bootstrap to generate samples of varying size based on initial samples for examples in literature.

The paper is laid out as follows. Section 2 provides a review of the bootstrap method and the proposed measures of distance between distributions. In section 3, we provide the mathematical details of the two measures of convergence. In section 4 we provide examples by simulating data from exponential families of distribution and apply our method to data available in the literature and textbook.

Review of Bootstrap and Distances Between Distributions

Bootstrap

The Bootstrap is a re sampling technique introduced by Efron (1979) with a related long history [11] and has attracted much attention in the past three decades primarily due to its conceptual simplicity and the computational empowerment of statisticians due to advances in computer science [12]. The past three decades have witnessed much works dedicated to developing bootstrap methods [13-19]. Bootstrap at its core, is a re sampling technique that treats the data at hand as a “surrogate population” and allows for sampling with replacement with a goal of re-computing the statistic of interest many times. This allows us to examine its bootstrap distribution. Efron has demonstrated that the bootstrap method outperforms other methods such as jackknifing and cross-validation [12]. Despite the simplicity of the bootstrap algorithm, the large sample properties of bootstrap distributions are surprisingly elegant. Singh (1981), for example has demonstrated that the sampling distribution of $\hat{\theta}$, where $\theta$ is an estimate of $\theta$, is approximated well approximated by its bootstrap distribution [20]. Bickel and Freedman have also made substantial contributions in developing bootstrap theory [21-23]. The most common applications of the bootstrap in its basic form involve approximating the standard error of sample estimators, correcting the bias in the sample estimate, and in constructing confidence intervals. However in situations involving bootstrapping dependent data, modified bootstrap approaches such as moving-block bootstrap are recommended [24]. Romano (1992) has also discussed extensively applications of the bootstrap [18]. Here we use the bootstrap as an approach to simulating samples based on the observed data. The sampling properties of the bootstrap are not used directly as we observe convergence behavior on the likelihood scale.

4.2 Distance between distributions

Kullback-Leibler distance is a commonly used measure of the difference between two statistical distributions [25]. If $p(x)$ and $q(x)$ are two continuous distributions the KL distance between $p(x)$ and $q(x)$ is defined as follows:

$$KL(p, q) = \int p(x) \log \frac{p(x)}{q(x)} dx = \int p(x) \log p(x) dx - \int p(x) \log q(x) dx.$$  

Kullback-Leibler distance has been applied in areas such as functional linear models, Markovian processes, model selection, and classification analysis [26-29]. It should be noted that the Kullback-Leibler distance is not symmetric, $KL(p || q) \neq KL(q || p)$, but can be expressed in a symmetric form [30].

Bhattacharya distance is another popular measure of difference between two distributions [31]. If $p(x)$ and $q(x)$ are two continuous distributions the Bhattacharya distance between $p(x)$ and $q(x)$ is defined as follows:

$$B(p, q) = \int \sqrt{p(x)q(x)} dx.$$  

Bhattacharya distance has also found extensive applications in several fields [32-35]. Bhattacharya distance assumes the product of $p(x)$ and $q(x)$ to be non-negative.

In lieu of the above two distance measures, we could simply use $\int (f_1(\theta) - f_2(\theta))d\theta$ as a measure of proximity between the two functions $f_1(\theta)$ and $f_2(\theta)$. Geometrically this measure is the difference in the area under the two curves $f_1(\theta)$ and $f_2(\theta)$.

In this paper we make use of the bootstrap approach to resample from the actual sample (or simulate data from known distributions) to obtain a “bootstrap sample”. For each “bootstrap sample”, the observed

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Observed Relative and Asymptotic Relative Likelihood Functions for Bartholomew’s Data.}
\end{figure}
relative likelihood function and corresponding asymptotic (normal) relative likelihood function are constructed and the area under the two relative likelihood functions computed. As the size of “bootstrap sample” increases we measure the convergence of the observed relative likelihood function to the asymptotic relative likelihood function. The convergence is measured by the difference in area under the curves and a dot product based measure of curve similarity. We note that simulated data is not a real world.

**Method**

**Background**

Let \( X_1, \ldots, X_n \) be iid random variables from a specified distribution \( f(X \mid \theta) \) with observed values \( x_1, \ldots, x_n \). The observed relative likelihood function of \( \theta \) i.e. \( R(\theta) \) is defined as follows:

\[
R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}
\]

Since \( \theta \sim N(\theta', I^{-1}(\theta')) \), the asymptotic relative likelihood function of \( \theta \) can be defined as follows:

\[
R_n(\theta) = \frac{L_n(\theta)}{L_n(\hat{\theta})}
\]

For exponential families, the density function can be expressed in the following form:

\[
f(x \mid \theta) = \exp \left( c(\theta) \sum h(x_i) - n \alpha(\theta) + b(x) \right)
\]

and the related likelihood function expressed as:

\[
L(\theta | X) = \exp \left( c(\theta) \sum h(x_i) - n \alpha(\theta) + b(x) \right)
\]

If \( \hat{\theta} \) is the mle of \( \theta \), then the likelihood function evaluated at \( \hat{\theta} \) is:

\[
L(\hat{\theta} | X) = \exp \left( c(\hat{\theta}) \sum h(x_i) - n \alpha(\hat{\theta}) + b(x) \right)
\]

Thus the observed relative likelihood function \( R(\theta) \) is:

\[
R(\theta) = \frac{L(\theta | X)}{L(\hat{\theta} | X)}, \quad \exp \left( c(\theta) \sum h(x_i) - n \alpha(\theta) + b(x) \right) \frac{\exp \left[ \sum \left( c(\theta) - c(\hat{\theta}) \right) \right]}{\exp \left[ \sum \left( c(\hat{\theta}) - c(\hat{\theta}) \right) \right]}
\]

The asymptotic relative likelihood function for \( \theta \) assumes the following form:

\[
R_n(\theta) = \exp \left( \frac{1}{2} \frac{\hat{\theta} - \theta}{\text{std.err}(\theta)} \right)^2 = \exp \left( \frac{1}{2} \hat{I}(\theta) \right)
\]

where

\[
\text{std.err}(\theta) = \sqrt{\hat{I}(\theta)}
\]

and

\[
\hat{I}(\theta) = -E \left( \frac{\partial^2}{\partial \theta^2} \log(L(X | \theta)) \right) = -\sum h(x_i) + n \alpha(\theta) \text{ evaluated at } \theta.
\]

In situations where computation of expectations is not analytically tractable, \( \hat{I}(\theta) \) can be estimated by \( J(\theta) = -\exp \left( \frac{\partial^2}{\partial \theta^2} \log f(X | \theta) \right) \) evaluated at \( \hat{\theta} \).

Here both \( R(\theta) \) and \( R_n(\theta) \) are positive since both are exponential functions.

**Measure of distance between** \( R(\theta) \) and \( R_n(\theta) \).

If \( R(\theta) \) and \( R_n(\theta) \) are defined over the interval \( \theta, \theta' \), the difference in area under the two likelihood curves will serve as the measure of discrepancy between \( R(\theta) \) and \( R_n(\theta) \) and can be computed as follows:

\[
\Delta R = \int \frac{R(\theta)}{\hat{\theta}} d\theta - \int \frac{R_n(\theta)}{\hat{\theta}} d\theta
\]

(1)

If the expression does not have a closed form solution, numerical methods such as Simpson’s rule [36] can be applied:

\[
\Delta R = \sum_{i=1}^n \left( R(\theta_i) - R_n(\theta_i) \right) \delta
\]

(2)

Where \( n \) is the number of intervals.

For similar curves we would expect \( \Delta R \) to be very small. A tolerance level may set: \( |\Delta R_n - \Delta R| < \epsilon ; n_i < n_2 \) Typically \( \epsilon = 0.01 \) will be acceptable.

**Property of \( \Delta R \):**

1. On a log scale, \( R(\theta | X) \) can be approximated by \( R_n(\theta | X) \) up to a second term [37].

Proof:

The general expression for Taylor expansion of a function \( f(x) \) around ‘a’ is as follows:

\[
f(x) = \sum_{k=0}^{\infty} \frac{(x-a)^k}{k!} f^{(k)}(a)
\]

Using Taylor expansion on \( \log(\hat{R}(\theta | X)) \) around \( \hat{\theta} \) we have:

\[
\log(\hat{R}(\theta | X)) = \log(\hat{R}(\theta | X)) + \log(\hat{R}(\theta | X)) + \text{higher order powers of } k
\]

Now:

\[
\frac{\partial}{\partial \theta} \log(\hat{R}(\theta | X)) = \frac{1}{\hat{R}(\theta | X)} \frac{\partial}{\partial \theta} \log(L(\theta | X)) = - \frac{1}{\hat{R}(\theta | X)} \frac{\partial}{\partial \theta} L(\theta | X) = -J(\theta)
\]

This is the derivative of the score function evaluated at mle. \( \frac{\partial}{\partial \theta} \log(L(\theta | X)) = -J(\theta) \)

Thus \( \log(\hat{R}(\theta | X)) \) can be approximated as:

\[
\log(\hat{R}(\theta | X)) = -\frac{1}{2} J(\theta)
\]

The \( ! \) in the higher order terms of the Taylor expansion shrinks it to 0.

For exponential families:

\[
\log(\hat{R}(\theta | X)) = -\frac{1}{2} \left( \sum_h \left( c(\theta) - c(\hat{\theta}) \right) \right)^2 - \frac{1}{2} \left( c(\theta) - c(\hat{\theta}) \right)
\]
Let \( L_i(\theta) \) and \( L_j(\theta) \) be two different functions of \( \theta \) with the same domain \( U \). Graphically \( L_i(\theta) \) and \( L_j(\theta) \) can be visualized as two curves constructed on the same support. The two curves need not necessarily have closed functional form. Here we propose a simple and computationally efficient algorithm that uses the dot product to measure the similarity of the two curves in terms of their curvature.

The idea is to divide the support of the two curves into sufficiently small segments so that each of them can be approximated by a line segment (Figure 2). Each of these segments is equivalent to a vector in two dimensions and hence we can compute the dot product for the two vectors in each of these segments. If in general the two vectors are similar, the dot product between the two vectors is equal to the product of their individual L2 norms over each segment (Figure 2). Each of these segments is equivalent to a vector.

Two vectors in each of these segments. If in general the two vectors are similar, the dot product between the two vectors is equal to the product of their individual L2 norms over each segment. If in general the two vectors are similar, the dot product between the two vectors is equal to the product of their individual L2 norms over each segment.

For a given sample \( X_i, \ldots, X_n \), compute \( R(\theta) \) and \( R_k(\theta) \).

Choose an error level for \( \Delta R \) and \( D \).

If \( \Delta R \) and \( D \) are not sufficiently close to 0, bootstrap from the original sample and compute \( \Delta R \) and \( D \) again for the bootstrap sample.

Repeat step 3 until satisfactory convergence in achieved i.e. \( D \) and \( \Delta R \) are less than chosen tolerance level.

The next section contains several simulated examples to demonstrate the application of the above method.

**Results**

In this section, we examine the convergence of likelihood functions for some of the common distributions, using simulated data as well as for data obtained from the literature. Expression for \( R(\theta) \) and \( R_k(\theta) \) for some common distributions are tabulated in Table 1. \( R(\theta) \) and \( R_k(\theta) \) are the observed and large sample normal likelihood functions scaled by their modes.

**Simulation studies**

The convergence of the observed relative likelihood function to

\[
\sum_{i=1}^{n} d_i = n.
\]  

A Dissimilarity Index

Equation (3) can be used to express disagreement between the two curves (a dissimilarity index). If \( D \) is the dissimilarity index between the two curves then,

\[
D = \frac{n - \sum d_i}{n}.
\]

Note that: \( 0 \leq D \leq 1 \).

The proposed simulation based approach can be summarized in the following steps.

For a given sample \( X_i, \ldots, X_n \), compute \( R(\theta) \) and \( R_k(\theta) \).

Choose an error level for \( \Delta R \) and \( D \).

Compute \( \Delta R \) and \( D \) for the given sample.

If \( \Delta R \) and \( D \) are not sufficiently close to 0, bootstrap from the original sample and compute \( \Delta R \) and \( D \) again for the bootstrap sample.

Repeat step 3 until satisfactory convergence in achieved i.e. \( D \) and \( \Delta R \) are less than chosen tolerance level.

The next section contains several simulated examples to demonstrate the application of the above method.

**Table 1:** Observed and relative functions for some distributions in exponential families.

| Distribution | \( R(\theta) \) | \( R_k(\theta) \) | Note |
|--------------|----------------|----------------|------|
| Poisson      | \( e^{x+\beta} \frac{\lambda}{\beta!} \) | \( \exp(\frac{\eta}{\beta} (x - \hat{\beta})) \) | \( \hat{\lambda} \neq \lambda \) |
| Binomial     | \( \binom{n}{x} \frac{p^x (1-p)^{n-x}}{\hat{p}^x (\hat{1-p})^{n-x}} \) | \( \exp\left( \frac{\bar{p}}{\beta} (x - \hat{\beta}) \right) \) | \( \hat{\lambda} = \lambda \) |
| Exponential  | \( \frac{\beta^x}{\Gamma(x+1)} \) | \( \frac{\beta^x}{\Gamma(x+1)} \) | \( \hat{\beta} = \beta \) |
| Weibull      | \( \left( \frac{x}{\beta} \right)^{\gamma} \) | \( \left( \frac{x}{\beta} \right)^{\gamma} \) | \( \gamma \) : a constant |

Table 1: Observed and relative functions for some distributions in exponential families.
The values of difference in area (red line) and dissimilarity index (green line) for data of different sample sizes generated under the Poisson distribution.

**Figure 3:** Poisson Distribution-Change in values of Difference in Area and Dissimilarity Index.

Observed relative likelihood function (blue curve) and asymptotic relative likelihood function (red curve) show increasing overlap as sample size increases.

**Figure 4:** Poisson Distribution-Observes and asymptotic relative likelihood functions.

Asymptotic relative likelihood function was examined using simulated dataset. For different families of exponential distributions, data were simulated for a given sample size. For the given data, the two convergence measures namely $\Delta R$ and $D$ were computed. This process was repeated for different sample sizes and the values of $\Delta R$ and $D$ thus obtained were recorded. The examples of some of the distributions from exponential families and the required sample sizes that makes the large sample approximation of the mle reasonable are presented in Tables 1 and 2 and in Figures 3-6. Additional examples for more distributions from exponential families are provided in supplementary

| n  | $\Delta R$:Difference in Area | $D$:Dissimilarity Index |
|----|-------------------------------|-------------------------|
| 1  | 0.0691                        | 0.01968                 |
| 3  | 0.01318                       | 0.01183                 |
| 5  | 0.00604                       | 0.00767                 |
| 7  | 0.00359                       | 0.005752                |
| 10 | 0.00201                       | 0.00374                 |
| 15 | 0.00114                       | 0.00303                 |

Table 2: Poisson Distribution: Values of difference in area and dissimilarity index for data of different sample sizes simulated from Poisson distribution.
The values of difference in area (red line) and dissimilarity index (green line) for data of different sample sizes generated under weibull distribution. The Pearson’s and Spearman’s rank correlation between the two measures of convergence are also reported.

**Figure 5:** Weibull Distribution-Change in values of Difference in Area and Dissimilarity Index.

Observed relative likelihood function (blue curve) and asymptotic relative likelihood function (red curve) show increasing overlapping as sample size increases.

**Figure 6:** Weibull Distribution-Observed and asymptotic relativelikelihood functions.

| n  | ΔR: Difference in Area | D: Dissimilarity Index |
|----|------------------------|------------------------|
| 10 | 0.35706                | 0.34744                |
| 15 | 0.20736                | 0.23947                |
| 30 | 0.08872                | 0.16687                |
| 50 | 0.04498                | 0.10846                |
| 75 | 0.02406                | 0.07855                |
| 100| 0.0158                 | 0.0615                 |

**Table 3:** Weibull Distribution Values of difference in area and dissimilarity index for data of different sample sizes simulated from Weibull distribution.

| Number of days until death | Mean | Variance |
|-----------------------------|------|----------|
| control                     | 5,6,7,8,8,8,9,12 | 7.778    | 3.944 |
| drug                        | 7,8,8,8,9,12,1,3,14,17 | 10.5 | 10.944 |

**Table 4:** Data from Gibbon’s et al. Values represent number of days each mice survived when exposed to control or drug.

Example 1: Poisson distribution ($\lambda = 10$)
Example 2: Weibull Distribution \( (\gamma = 2, \beta = 6) \)

Examples Using Data from Literature

a) Data from Gibbons et al’s book “Nonparametric Statistical Inference” [38].

b) Data from Williams et al. (1995).

The following data was obtained from Williams et al. [39,40]. The data is the weight (in grams) of dry seed in the stomach of each spinifex.
The values of difference in area (red line) and dissimilarity index (green line) for bootstrapped data (Williams et al.) of different sample sizes.

**Figure 9:** Change in values of Difference in Area and Dissimilarity Index.

Observed relative likelihood function (blue curve) and asymptotic relative likelihood function (red curve) show increasing overlapping as sample size increases.

**Figure 10:** Williams et al Data Bootstrapped-Observed and asymptotic relative likelihood functions.

pigeon captured in desert. The data is as follows:

0.457,3.751,2.967,2.509,1.384,1.454,0.818,0.335,1.436,1.603,1.309,0.201,0.530,2.144,0.834.

The plot of relative and relative normal likelihood functions together with the values of $\Delta A$ and $D$ is in figure 8:

While the difference in area is small enough, the value of dissimilarity index seems fairly high. It was seen that that with larger samples (bootstrap) the dissimilarity index and difference in area both decrease (Table 5, Figure 9 and 10).

c) Data from Breslow (1984)

The data set is taken from a paper by Breslow who proposed an iterative algorithm for fitting over-dispersed Poisson log-linear model. The dataset provides the number of revertant colonies of TA98 Salmonella observed on each of plates processed at 6 dose levels of quinoline [39].
The two convergence measures (Table 6 and Figure 11) suggest that the data at each dose level is large enough for the mle to satisfy asymptotic normality.

Discussion

Our work discusses the issue of appropriateness of sample size required for asymptotic normality of mles to hold true. We essentially proposed two different diagnostic measures for this purpose viz. $\Delta R$ – difference in the area under the relative observed likelihood and relative asymptotic likelihood curves and $D$ – dissimilarity index which measures the shape of the curves. The simulated results show that different distributions have different threshold of $\Delta R$ and $D$. It gives an informal measure of convergence in real world. For example if we believe that the data at hand follows $\text{Poi}(\lambda = 10)$ distribution we could compute $\Delta R$ and $D$ and compare it with the tabulated values in Table 2. If the $\Delta R_{\text{computed}}$ and $D_{\text{computed}}$ are close to the tabulated values for the given sample size, assumption of asymptotic normality of mles is reasonable.

The two measures of convergence were also applied to data from the literature and bootstrap techniques were used in assessing the convergence of relative likelihood functions as the sample size increased. As seen from the simulated examples as well as the example from literature, a “sample size of 30” can be far more than what is actually needed and the sample size requirements for satisfactory asymptotic convergence differs for different distributions. For example with Poisson $(\lambda = 10)$ distribution, it was seen that samples of sizes less than 10 show convincing convergence. Our future work is directed at generalizing these diagnostic measures to distributions taking into account parameters within more than one dimension.

References

1. Fisher RA (1932) Inverse probability and the use of likelihood. Proceedings of the Cambridge Philosophical Society 28: 257-261.
2. Fisher RA (1934) Probability likelihood and quantity of information in the logic of uncertain inference. Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character 146: 0001-0008.
3. Fisher RA (1934) Two new properties of mathematical likelihood. Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character 144: 0285-0307.
4. Fisher RA (1941) The likelihood solution of a problem in compounded probabilities. Annals of Eugenics 11: 306-307.
5. Casella G, Berger Roger L (2001) Statistical Inference. Second ed., Belmont, California: Duxbury Press.
6. Altman M, Gill J, McDonald M (2003) Convergence Problems in Logistic Regression, in Numerical Issues in Statistical Computing for the Social Scientist.
7. Efron B (1982) Maximum Likelihood and Decision Theory. The Annals of Statistics 10: 340-356.
8. Efron B, Hinkley D (1978) Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. Biometrika 65: 457-482.
9. Sprott DA, Kalbflei JD (1969) Examples of Likelihoods and Comparison with Point Estimates and Large Sample Approximations. Journal of the American Statistical Association 64: 468-484.
10. Bartholomew DJ (1957) A problem in life testing. Journal of Americal Statistical Association 52: 350-355.
11. Hall P (1994) A short history of the bootstrap, in Acoustics, Speech, and Signal Processing, IEEE International Conference, Australia.
12. Efron B (1977) Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife. Annals of Statistics 7: 1-26.
13. Hall P (1986) On the Bootstrap and Confidence-Intervals. Annals of Statistics 14: 1431-1452.
14. Hall P (1988) Rate of Convergence in Bootstrap Approximations. Annals of Probability 16: 1665-1684.
15. Hall P (1988) On Symmetric Bootstrap Confidence-Intervals. Journal of the Royal Statistical Society Series B-Methodological 50: 35-45.
16. Hall P (1989) On Efficient Bootstrap Simulation. Biometrika 76: 613-617.
17. Hall P, Wilson SR (1992) Bootstrap Hypothesis-Testing - Reply. Biometrics 48: 970-970.
18. Leger C, Politis DN, Romano JP (1992) Bootstrap Technology and Applications. Technometrics 34: 378-398.
19. Singh K (1986) Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion. Annals of Statistics 14: 1328-1330.
20. Singh K (1981) On the Asymptotic Accuracy of Efrons Bootstrap. Annals of Statistics 9: 1187-1195.
21. Bickel PJ, Freedman DA (1981) Some Asymptotic Theory for the Bootstrap. Annals of Statistics 9: 1196-1217.
22. Bickel PJ, Freedman DA (1984) Asymptotic Normality and the Bootstrap in Stratified Sampling. Annals of Statistics 12: 470-482.
23. Freedman DA (1986) Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - Discussion. Annals of Statistics 14: 1305-1308.
24. Carlstein E (1998) Matched-block bootstrap for dependent data. Bernoulli 4: 305-328.
25. Kulback S, Leibler RA (1951) On Information and Sufficiency. Annals of Mathematical Statistics 22: 79-86.
26. Gersch W (1980) A Kulback-Lebler Nearest Neighbor Rule Classification of Eegs - the Eeg Population Screening Problem, an Anesthesia Level Eeg Classification Application. Computers and Biomedical Research 13: 283-296.
27. Kn'kov EA (2007) Application of the Kulback-Lebler measure for estimating the instants of a change in the statistical properties of a binary Markovian process. Journal of Communications Technology and Electronics 52: 1350-1354.
28. Kubokawa T, Tsukuma H (2007) Estimation in a linear regression model under the Kulback-Lebler loss and its application to model selection. Journal of Statistical Planning and Inference 137: 2467-2508.
29. Rodrigues J (1992) The Kulback-Lebler Approximation of the Marginal Posterior Density - an Application to the Linear Functional-Model. Communications in Statistics-Theory and Methods 21: 2861-2868.
30. Johnson DH, Sinanovic S (2001) Symmetrizing the kullback-leibler distance. IEEE Transactions on Information Theory 1: 1-10.
31. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distribution. Calcutta Mathematical Society 35: 99-110.
32. Chaudhuri G, Borwankar JD, Rao PRK (1991) Bhattacharyya Distance Based Linear Discriminant Function for Stationary Time-Series. Communications in Statistics-Theory and Methods 20: 2195-2205.
33. Chen K (2014) Vessel attachment nodule segmentation using integrated active contour model based on fuzzy speed function and shape-intensity joint Bhattacharyya distance. Signal Processing 103: 273-284.
34. Jain AK (1976) Estimate of Bhattacharyya Distance. Ieee Transactions on Systems Man and Cybernetics 6: 763-766.
35. Schwepp FC (1967) On Bhattacharyya Distance and Divergence between Gaussian Processes. Information and Control 11: p. 373-395.
36. Atkinson KE (1989) An Introduction to Numerical Analysis. (2ndedn) John Wiley and Sons.
37. Sprott DA (1973) Normal Likelihoods and Their Relation to Large Sample Theory of Estimation. Biometrika, 60: 457-465.
38. Gibbons JD, Chakraborti S (2011) Nonparametric Statistical Inference,4th ed. IN: Balakrishnan, SWR,CRC Press, Taylor and Francis Group.
39. Breslow NE (1984) Extra-Poisson variation in log-linear models. Applied Statistics 33: 38-44.
40. Williams JB, Bradshaw D, Schmidt L (1995) Field metabolism and water requirements of spinifex pigeons (Geophaps plumifera) in Western Australia. Australian Journal of Zoology 43: 1-15.