Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n2p465

On the shortest pivotal confidence intervals: an entropy measure technique

By Al-Talib

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 02, October 2019, 465-475
DOI: 10.1285/i20705948v12n2p465

On the shortest pivotal confidence intervals: an entropy measure technique

Mohammad Al-Talib*

Yarmouk University, Department of Statistics
21163 Irbid, Jordan

For the important role of confidence intervals in statistical inference, we present in this article the shortest pivotal confidence interval using an entropy measure called the Resistor-Average distance, two examples are used to show the application of the proposed technique which include a simulation study.

keywords: Pivotal Quantity, Confidence Interval, Entropy Measure, Kullback-Leibler.

1 Introduction

In statistical inference, a Confidence Interval (CI) is an interval estimate of the unknown parameter of the population, where the degree of confidence refers to a given proportion of intervals which includes the true value of the parameter, it is usual to use 90% and 95% confidence levels in practice.
An accurate estimate of the unknown parameter is important to avoid uncertainty and any misleading inferences. In this sense, confidence intervals have the guarantee of capturing the true value of the unknown parameter as it gives a range of values which is likely (by likely, we mean with a high probability) to contain the population parameter. CI is thought of as a tool gives an impression of the precision of the estimate as it depends on the parameter and the error model, the certainty of the interval estimate is based on two important measures: (1) coverage probability and (2) interval size. For any interval estimate of the form $[L(X), U(X)]$ of any unknown parameter θ, the coverage probability is defined as the probability of covering the true parameter; i.e., $P_\theta\big(\theta \in [L(X), U(X)]\big)$. The infimum of the coverage probability is known as the confidence coefficient denoted by α, and the size of the interval estimate is its length, i.e., $\big(U(X) - L(X)\big)$.

*Corresponding author: m.altalib@yu.edu.jo

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

To get the most accurate interval estimate, we need to obtain the interval with the most coverage and the least size or equivalently the shortest length, many researchers addressed this issue, see; Nakagawa and Cuthill (2007),Casella and Berger (2002), Ferentinos and Karakostas (2006), Gao et al. (2012), among many others.
There are different methods for finding interval estimator, such as: inverting a test statistic, pivotal quantities, pivoting a cumulative distribution function and bayesian credible sets. Different construction methods were studied extensively in literature (for example; Nakagawa and Cuthill (2007), Casella and Berger (2002),Meeker et al. (2017)), a special concern of finding the shortest interval can be found in Nakagawa and Cuthill (2007), Guenther (1969), Hall (1988), Juola (1993) and many others.
As a method of finding confidence intervals, pivotal quantities is a very useful method and its f interest in this article. For general definitions and examples of pivotal quantities see Meeker et al. (2017), Casella and Berger (2002).

On the other hand, information theory provides a statistical basis for quantifying and storing information. In probability theory, the fundamental quantities of information theory (entropy, mutual information, and Kullback-Leibler distance) are defined as functionals of probability distributions.
Kullback-Leibler distance is the measure of the discrepancy between two distributions, it have many important statistical interpretations in different fields of statistics, especially in estimation and hypotheses testing.

The motivation of this article is to study the distance between the distributions of the CI bounds and to minimize this distance under the constriction of coverage probability to obtain the shortest confidence interval.

The outline of this article is as follows; In section 1 we presented an introduction on confidence intervals and definitions of entropy measures, we also stated our motivation. In section 2, we presented the bases of constructing the shortest pivotal confidence intervals, the classical method and proposed method from entropy point of view. Also, we introduced our main theorem. In section 3, we applied the proposed methodology as we studied two examples via a simulation study. Lastly, in section 4, we present our conclusions and remarks.

2 Shortest Pivotal Confidence Interval

Let X be a real valued random variable with a probability density function (*pdf*) $f(x;\theta), x \in R$, with unknown parameter $\theta \in \Theta$. A random quantity $Q(X;\theta)$ is a pivotal quantity if the distribution of $Q(X;\theta)$ is independent of all parameters. For a specific value $\alpha \geq 0$, the $(1-\alpha)$ confidence set for θ can be presented as

$$C(x) = \{\theta_0 : a \leq Q(x;\theta_0) \leq b\}, \tag{1}$$

with constants a and b satisfies,

$$P_\theta\big(a \leq Q(x;\theta) \leq b\big) \geq 1-\alpha, \tag{2}$$

which can be converted to (if θ is a real valued parameter) to

$$C(X) = \{L(x) \leq \theta \leq U(x)\}, \tag{3}$$

the $(1-\alpha)\%$ confidence interval for θ. The interval $C(x)$ is considered as the Shortest Pivotal Confidence Interval (SPCI) if $\big(U(x) - L(x)\big)$ is minimum.
Theorem 9.3.2 in Casella and Berger (2002) introduced some requirements that a confidence interval should satisfy in order to be considered as the shortest confidence interval. A disadvantage of the theorem that one should be careful when dealing with scale families as it may not be applicable, while on the other hand it works well with location cases.
The end points of the confidence interval presented in equation 3; $L(x)$ and $U(x)$ are functions of x. As our motivation states, it is tempting to study the distance between these two functions under the constraint that the coverage probability of the interval equals $1-\alpha$, and then to look for the bounds that gives the shortest distance.

2.1 Entropy Measures of Distance

Entropy is defined as a measure of uncertainty of a single random variable or equivalently the uncertaninty of its *pdf.* Kullback-Leibler distance was first defined by Kullback and Leibler (1951). It is also known as relative entropy, cross entropy, information distance, and information for discrimination. In recent literature, Kullback-Leibler distance had some attention to be used and applied in different research fields. Let $p_0(x), p_1(x)$ be two probability densities, the Kullback-Leibler distance denoted by $D(p_1||p_0)$ is defined to be

$$D(p_1||p_0) = \int p_1(x) \log \frac{p_1(x)}{p_0(x)}.$$

Kullback-Leibler distance have two fundamental properties:

- non-negativity; $D(p_1||p_0) \geq 0$, with equality if and only if $p_1 = p_0$.
- asymmetry; $D(p_1||p_0) \neq D(p_0||p_1)$.

Minimizing $D(p_1||p_0)$ is studied in many fields of statistics; information theory, in large deviations theory and maximum entropy (see; Csiszár (1967), Kullback (1997), Good (1963), Akaike (1998), Cover and Thomas (2012), Abbas et al. (2017) and many others).
Because the asymmetry property of Kullback-Leibler distance it is not a true distance between distributions as it does not satisfy the triangle inequality. Its often of interest to find the I-projection $p^* = \arg\min D(p_1||p_0)$; the "closest" distribution to p_1 of all the distributions in P_0, the family of the distribution p_0. Many work has been done in finding p^* in convex sets (Dykstra (1985), Bhattacharya and Dykstra (1997), and Bhattacharya and Al-Talib (2017)).

To overcome the disadvantage of Kullback-Leibler distance of being an asymmetric measure of distance, Johnson and Sinanovic (2001) calculated a symmetric distance that has many of the properties of Kullback-Leibler distance, called the Resistor-Average distance denoted by $R(p_0, p_1)$ and defined as

$$R(p_0, p_1) = \frac{D(p_0||p_1).D(p_1||p_0)}{D(p_0||p_1) + D(p_1||p_0)}, \tag{4}$$

it arises from geometric considerations similar to those used to derive the Chernoff distance.
The Kullback-Leibler distance is attractive to be used in the context of finding this distance as it is easy to evaluate, but lack of symmetry is a key disadvantage, also we cannot decide on which bound could be considered the true distribution. Hence, we should look for a measure that is symmetric and have the important properties of Kullback-Leibler distance in hypothesis testing.
The two Kullback-Leibler distances; $D(p_0||p_1)$ and $D(p_1||p_0)$ are asymmetric but their average

$$J(p_0, p_1) = \frac{D(p_0||p_1) + D(p_1||p_0)}{2}$$

is symmetric.
It can be proved following the definition of arithmetic and harmonic means that,

$$R(p_0, p_1) \le min\{D(p_0||p_1), D(p_1||p_0)\} \le J(p_0, p_1) \le max\{D(p_0||p_1), D(p_1||p_0)\}$$

its interesting to consider Resistor-Average distance as a minimum bound of the distances $D(p_0||p_1), D(p_1||p_0)$.

Recall that the coverage probability of $C(x) = \{\theta : L(x; a(\alpha_1)) \le \theta \le U(x; b(\alpha_2)) : \alpha = \alpha_1 + \alpha_2\}$ is $(1-\alpha)$, $C(x)$ is considered the shortest if $U(x; b(\alpha_2)) - L(x; a(\alpha_1))$ is the minimum.
The choice of $a(\alpha_1)$ and $b(\alpha_2)$ will be the decisive criterion in choosing the values that insures minimum length.
The following theorem proves that the confidence interval with the smallest Resistor-Average distance will be the shortest.

Theorem 1. Let $C(x)$ be a $(1-\alpha)100\%$ confidence interval with the smallest Resistor-Average distance, then $C(x)$ is the shortest confidence interval for θ.

Proof. Suppose that

$$C^*(x) = \{\theta : L^*(x; a^*(\alpha_1)) \le \theta \le U^*(x; b^*(\alpha_2)) : \alpha = \alpha_1 + \alpha_2\}$$

is the confidence interval with the smallest Resistor-Average distance, where,

$$R(L^*, U^*) = \frac{D(L^*||U^*) \cdot D(U^*||L^*)}{D(L^*||U^*) + D(U^*||L^*)}.$$

Assume that,

$$C'(x) = \{\theta : L'(x; a'(\alpha_1)) \le \theta \le U'(x; b'(\alpha_2)) : \alpha = \alpha_1 + \alpha_2\},$$

be any interval such that $U'(x; b'(\alpha_2)) - L'(x; b'(\alpha_1))$ is less than $U^*(x; b^*(\alpha_2)) - L^*(x; b^*(\alpha_1)$. Now, let

$$\int_{a^*}^{b^*} L^* \log \frac{L^*}{U^*} = K_1^* \text{ and } \int_{a^*}^{b^*} U^* \log \frac{U^*}{L^*} = K_2^*,$$

then, $R(L^*, U^*) = \frac{K_1^* \cdot K_2^*}{K_1^* + K_2^*}$
If $a' < a^*$, then,

$$\begin{aligned}
\int_{a'}^{b'} L' \log \frac{L'}{U'} &= \int_{a'}^{b'} g(x)dx && ;\text{ say } K_1' \\
\int_{a'}^{b'} g(x)dx &\leq g(b')(b' - a') && ;\{g(x) \leq g(b')\} \\
&\leq g(a^*)(b' - a') && ;\{g(b') \leq g(a^*)\} \\
&< g(a^*)(b^* - a^*) && ;\{b' - a' < b^* - a^*, g(a^*) > 0\} \\
&\leq \int_{a^*}^{b^*} g(x)dx && ;\{g(x) \geq g(a^*)\} \\
\text{hence,} \qquad K_1' &\leq K_1^*.
\end{aligned} \tag{5}$$

In similar fashion, $K_2' \leq K_2^*$. And hence, $R(L', U') \leq R(L^*, U^*)$.
So, one can find a measure with smaller Resistor-Average distance, but does it have $(1 - \alpha)$ coverage probability? the proof of Theorem 9.3.2 of Casella and Berger (2002) assures that if the function is unimodal, and if there is a shorter interval,then it has a coverage probability less than or equal to $(1 - \alpha)$. Hence, The theorem is proved.

In the special case when probability distribution of $L(x)$ $\big($or $U(x)\big)$ is symmetric, then the Kullback-Leiber distance is also symmetric. Johnson and Sinanovic (2001) stated, in such case $R\big(L(x), U(x)\big)$ equals to $\frac{D\left(L(X)||U(x)\right)}{2}$. This statement draw the following proposition.

Proposition
If the probability distributions of the lower and the upper bounds of a $(1 - \alpha)$ confidence interval are symmetric, then the smallest Resistor-average distance is found when splitting α in half.

3 Simulation and Examples

Example 1: The shortest pivotal confidence interval for gamma distribution.

Let X be a random variable with a $Gamma(k; \beta)$ pdf, the quantity $Y = \frac{X}{\beta}$ is a pivotl quantity with Y $Gamma(k; 1)$. It is straight forward to state the confidence interval for β as,

$$\left\{\beta : \frac{x}{b} \leq \beta \leq \frac{x}{a}\right\}, \tag{6}$$

with the constraint $P(a \leq Y \leq b) = 1 - \alpha$.
The lenght of such interval is proportional to $\frac{1}{a} - \frac{1}{b}$ not to $b - a$, hence Theorem 9.3.2 of Casella and Berger (2002) is not applicable here. The shortest confidence interval is found by solving the following constrained minimization problem,

$$\begin{aligned} \textit{minimize, with respect to } a &: \frac{1}{a} - \frac{1}{b(a)} \\ \textit{subject to } &: \int_a^{b(a)} f(y)dy = 1 - \alpha. \end{aligned}$$

For $\alpha = 0.05$, the solution of the above constrained minimization problem for different k values are presented in Table 1.

Table 1: values of a and b for SPCI_{CM}

k	a	b
2	0.358	10.448
3	0.818	11.316
4	1.360	12.410
5	1.954	13.588

On the other hand, Theorem 1 will assist in finding the shortest pivotal confidence interval using the Resistor-Average distance.
Notice that the lower bound in equation 6; $L(x)$ is a random variable that has a Gamma distribution,

$$L(x) = \frac{x}{b} \sim Gamma(k_L = k, \beta_L = \frac{\beta}{b}),$$

similarly,

$$U(x) = \frac{x}{a} \sim Gamma(k_U = k, \beta_U = \frac{\beta}{a}).$$

The Kullback-Leibler distance from $L(x)$ to $U(x)$ is

$$\begin{aligned} D\big(U(x)||L(x)\big) &= \int U(x) \log\left(\frac{U(x)}{L(x)}\right) \\ &= (k_U - k_L)\psi(k_U) - \log\Gamma(k_U) + \log\Gamma(k_L) + k_L(\log\beta_L - \log\beta_U) + k_U\frac{\beta_U - \beta_L}{\beta_L}, \end{aligned}$$

(Penny, 2001)

$$\begin{aligned} &= (k - k)\psi(k) - \log\Gamma(k) + \log\Gamma(k) + k(\log\beta_L - \log\beta_U) + k\frac{\beta_U - \beta_L}{\beta_L} \\ &= k(\log\beta_L - \log\beta_U) + k\frac{\beta_U - \beta_L}{\beta_L} \\ &= k\Big[\log\left(\frac{\beta_L}{\beta_U}\right) + \frac{\beta_U}{\beta_L} - 1\Big] \end{aligned}$$

$$= k\Big[\log\Big(\frac{2\beta/b}{2\beta/a}\Big) + \frac{2\beta/a}{2\beta/b} - 1\Big]$$
$$= k\Big[\log\Big(\frac{a}{b}\Big) + \frac{b}{a} - 1\Big].$$

Similarly, the Kullback-Leibler distance from $U(x)$ to $L(x)$ is

$$D\big(L(x)||U(x)\big) = k\Big[\log\Big(\frac{b}{a}\Big) + \frac{a}{b} - 1\Big],$$

hence, the Resistor-Average distance is

$$R\big(L(x), U(x)\big) = \frac{D\big(L(x)||U(x)\big) \cdot D\big(U(x)||L(x)\big)}{D\big(L(x)||U(x)\big) + D\big(U(x)||L(x)\big)}, \quad (7)$$

The values of a and b need to satisfy $P(a \leq Y \leq b) = 1 - \alpha$. The problem of the shortest confidence interval problem comes down to find the suitable values of a and b which also minimizes equation 7.

Let P_a be the lower cumulative probability of a, and Q_b be the upper cumulative probability of b. A simulation study is carried out to find $R\big(L(x), U(x)\big)$ at $\alpha = 0.05$, such that $P(a \leq Y \leq b) = 1 - \alpha$ or equivalently $P_a + Q_b = \alpha$. Table 2, present some choices of a and b alongside the associated R value, when $k = 3$. The direct application of Theorem 1 suggests that the shortest pivotal confidence Interval is

$$\Big\{\beta : \frac{x}{7.94829} \leq \beta \leq \frac{x}{0.71250}\Big\},$$

with Resistor-average distance value equals 3.77307, meanwhile the confidence interval given by constrained minimization (at $k = 3$) is

$$\Big\{\beta : \frac{x}{11.316} \leq \beta \leq \frac{x}{0.818}\Big\},$$

has Resistor-average distance value equals 4.37050.

Table 3 below, presents the values of a, b and R that insures the shortest confidence interval using Resistor-average distance and constrained minimization. The bold face values of R indicates the smaller values. these results gives direct conclusions of Theorem 1. The SPCI's (using R and constrained minimization) found in Table 3 share approximately 93.9% when $k = 2$, and shares approximately 94%, 93.16%, 93.24% when $k = 3, 4$ or 5, respectively.

Table 2: R of some a and b choices

a	b	P_a	Q_b	R
0.19053	6.32342	0.001	0.049	1.62415
⋮	⋮	⋮	⋮	⋮
0.33786	6.43945	0.005	0.045	5.29878
⋮	⋮	⋮	⋮	⋮
0.43605	6.59891	0.010	0.040	4.62644
⋮	⋮	⋮	⋮	⋮
0.56721	6.98381	0.020	0.030	4.04413
⋮	⋮	⋮	⋮	⋮
0.66480	7.51660	0.030	0.020	3.80969
⋮	⋮	⋮	⋮	⋮
0.70701	7.88869	0.035	0.015	3.77364
⋮	⋮	⋮	⋮	⋮
0.71250	**7.94829**	**0.03568**	**0.01432**	**3.77307**
⋮	⋮	⋮	⋮	⋮
0.71506	7.97728	0.036	0.014	3.77320
⋮	⋮	⋮	⋮	⋮
0.74618	8.40595	0.04	0.01	3.79973
⋮	⋮	⋮	⋮	⋮
0.78293	9.27379	0.045	0.005	3.93716
⋮	⋮	⋮	⋮	⋮
0.81088	11.22887	0.049	0.001	4.37336

Table 3: R for the some a and b choices, $\alpha = 0.05$.

k	a	b	R	a	b	R
2	0.30351	6.40146	**3.73333**	0.358	10.448	4.38977
3	0.71250	7.94829	**3.77307**	0.818	11.316	4.37051
4	1.20701	9.43059	**3.79170**	1.360	12.410	4.31817
5	1.75803	10.86417	**3.80242**	1.954	13.588	4.26492

Example 2: The shortest pivotal confidence interval for Normal distribution.

Let $X_1, X_2, ..., X_n$ be *iid* from $N(\mu, \sigma^2)$, where σ^2 in known, its quite straight forward to get,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

a pivotal quantity with standard normal. Then, the pivotal $(1-\alpha)100\%$ confidence interval is given by,

$$\left\{\mu : \bar{x} - b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} - a\frac{\sigma}{\sqrt{n}}\right\},$$

where a and b need to satisfy $P(a \leq Z \leq b) = 1 - \alpha$.
To obtain the Shortest Pivotal confidence interval, we need to find the values a and b that minimizes $b - a$ and to satisfy $P(a \leq Z \leq b) = 1 - \alpha$. (the interval length is $(b-a)\sigma/\sqrt{n}$, but since the factor $\sigma/\sqrt{n}$ is common in the interval bounds, it can be ignored).
It is easy to show that the values $a = -z_{\frac{\alpha}{2}}$ and $b = z_{\frac{\alpha}{2}}$ is optimal for minimum length. Hence, the SPCI is

$$\left\{\mu : \bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}.$$

To apply the resistor-average distance here. Note that, the lower bound $L(x) = \bar{x} - b \cdot \sigma/\sqrt{n}$ has a normal distribution with mean $\mu - b \cdot \sigma/\sqrt{n}$ and standard deviation $\sigma/\sqrt{n}$.
Similarly, the upper bound; $U(x) \sim N(\mu - a \cdot \sigma/\sqrt{n}, \sigma^2/n)$.
Now, the Kullback-Leibler distance from $L(x)$ to $U(x)$ is,

$$\begin{aligned}
D\Big(U(x)||L(x)\Big) &= \int U(x)\log\left(\frac{U(x)}{L(x)}\right) \\
&= \frac{(\mu_U - \mu_L)^2}{2\sigma_U^2} + \frac{1}{2}\left\{\frac{\sigma_U^2}{\sigma_L^2} - \log\frac{\sigma_U^2}{\sigma_L^2} - 1\right\} \\
&= \frac{(\mu - a \cdot \sigma/\sqrt{n} - (\mu - b \cdot \sigma/\sqrt{n}))^2}{2\sigma^2/n} + \frac{1}{2}\left\{\frac{\sigma^2/n}{\sigma^2/n} - \log\frac{\sigma^2/n}{\sigma^2/n} - 1\right\} \\
&= \frac{(-a \cdot \sigma/\sqrt{n} + b \cdot \sigma/\sqrt{n})^2}{2\sigma^2/n} \\
&= \frac{((-\sigma/\sqrt{n}) \cdot (b-a))^2}{2\sigma^2/n} = \frac{(b-a)^2}{2}.
\end{aligned}$$

Similarly, $D\big(L(x)||U(x)\big) = \frac{(b-a)^2}{2}$. Hence, the Resistor-Average distance is $R\Big(L(x), U(x)\Big) = \frac{D\big(L(x)||U(x)\big)}{2}$, (Proposition 2.1).
Equivalently,

$$R\Big(L(x), U(x)\Big) = \frac{D\big(L(x)||U(x)\big) \cdot D\big(U(x)||L(x)\big)}{D\big(L(x)||U(x)\big) + D\big(U(x)||L(x)\big)} = \frac{\frac{(b-a)^2}{2} \cdot \frac{(b-a)^2}{2}}{\frac{(b-a)^2}{2} + \frac{(b-a)^2}{2}} = \frac{(b-a)^2}{4}.$$

The shortest confidence interval is obtained by minimizing $\frac{(b-a)^2}{4}$ subject to $P(a \leq Z \leq b) = 1 - \alpha$.
A simulation study is carried out to find $R\big(L(x), U(x)\big)$ at $\alpha = 0.05$, the results are given in Table 4. The results in bold face presents the values associated with the smallest R,

Table 4: R for the some a and b choices, $\alpha = 0.05$.

a	b	P_a	Q_b	R
-1.654	3.090	0.0490	0.0010	5.628425
⋮	⋮	⋮	⋮	⋮
-1.685	2.576	0.0450	0.0050	4.560845
⋮	⋮	⋮	⋮	⋮
-1.751	2.326	0.0400	0.0100	4.155551
⋮	⋮	⋮	⋮	⋮
-1.881	2.054	0.0300	0.0200	3.870156
⋮	⋮	⋮	⋮	⋮
-1.960	**1.960**	**0.0250**	**0.0250**	**3.841459**
⋮	⋮	⋮	⋮	⋮
-2.054	1.881	0.0200	0.0300	3.870156
⋮	⋮	⋮	⋮	⋮
-2.326	1.751	0.0100	0.0400	4.155551
⋮	⋮	⋮	⋮	⋮
-3.719	1.646	0.0050	0.0450	4.560845

the resistor-average distance technique gives the optimal values of a and b, as the classical way did.

4 Conclusions and Remarks

This article aimed to present the shortest confidence interval using the Resistor-Average distance based on the belief that entropy measures hold information and reduce uncertainty in making inferences. We do not believe that classical methods of finding the shortest pivotal confidence intervals give less accurate estimates or misleading infereces, but when the classical methods are not applicable, we suggest the entropy based measure for obtaining such intervals. Theorem 1 and Proposition 2.1 gives the requirements of the interval bounds that insures shortest distance between them, without affecting the coverage probability of the interval.
Two examples were introduced in section 4, we applied a classical method and the en-

tropy based method and the results were compared in Tables 3 and 4. We found that the Resistor-Average distance performed better (in the sense of uncertainty, see Example 1) or it will be as good as the classical method (see, Example 2).

Acknowledgment

We have to express out appreciation to the Editor in chief and the anonymous reviewers for their time and valuable comments which led to a better presentation of this article.

References

Abbas, A., Cadenbach, A., and Salimi, E. (2017). A kullback–leibler view of maximum entropy and maximum log-probability methods. *Entropy*, 19(5):232.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Bhattacharya, B. and Al-Talib, M. (2017). A minimum relative entropy based correlation model between the response and covariates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1095–1118.

Bhattacharya, B. and Dykstra, R. (1997). A fenchel duality aspect of iterative i-projection procedures. *Annals of the Institute of Statistical Mathematics*, 49(3):435–446.

Casella, G. and Berger, R. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cover, T. and Thomas, J. (2012). *Elements of information theory.* John Wiley & Sons.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318.

Dykstra, R. (1985). An iterative procedure for obtaining i-projections onto the intersection of convex sets. *The annals of Probability*, pages 975–984.

Ferentinos, K. and Karakostas, K. (2006). More on shortest and equal tails confidence intervals. *Communications in Statistics-Theory and Methods*, 35(5):821–829.

Gao, S., Zhang, Z., and Cao, C. (2012). Particle swarm optimization algorithm for the shortest confidence interval problem. *JCP*, 7(8):1809–1816.

Good, I. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *The Annals of Mathematical Statistics*, pages 911–934.

Guenther, W. (1969). Shortest confidence intervals. *The American Statistician*, 23(1):22–25.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, pages 927–953.

Johnson, D. and Sinanovic, S. (2001). Symmetrizing the kullback-leibler distance. *IEEE Transactions on Information Theory.*

Juola, R. (1993). More on shortest confidence intervals. *The American Statistician*, 47(2):117–119.

Kullback, S. (1997). *Information theory and statistics.* Courier Corporation.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Meeker, W., Hahn, G., and Escobar, L. (2017). *Statistical intervals: a guide for practitioners and researchers*, volume 541. John Wiley & Sons.

Nakagawa, S. and Cuthill, I. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews*, 82(4):591–605.

Penny, W. (2001). Kullback-liebler divergences of normal, gamma, dirichlet and wishart densities. *Wellcome Department of Cognitive Neurology, London.*