Interval Estimation Procedures and Information Inequalities

Masafumi Akahira
Institute of Mathematics, University of Tsukuba, Ibaraki, Japan, 305-8571
e-mail: akahira@math.tsukuba.ac.jp

Abstract

In the case when a parameter is assumed to be nonnegative or positive valued, a combined Bayesian-frequentist approach to confidence intervals is adopted, and its comparison with ordinary and Bayesian ones are done in normal cases. In the interval estimation problem on the difference between two means in exponential and normal cases, a systematic method of the construction of a confidence interval is proposed, and its application to the Behrens-Fisher type problem is given. Next, we consider a family of distributions for which the second order Bhattacharyya bound becomes sharp, and a necessary and sufficient condition for the second order one to be sharp is given for a family of linear combinations of distributions which belong to the exponential family. From the Bayesian viewpoint, we construct an estimator which minimizes locally the variance of any estimator satisfying weaker conditions than the unbiasedness from which an information inequality is derived.

Key Words: Confidence intervals, Combined Bayesian-frequentist approach, Behrens-Fisher type problem, Bhattacharyya bound

1. Introduction

If we adopt the non-Bayesian standpoint, interval estimation procedure is usually formalized as a procedure with a fixed confidence coefficient covering the true value of the parameter with probability of preassigned value. Such a procedure is mathematically obtained from a class of test procedures.

In the case when a parameter is assumed to be nonnegative or positive-valued, we consider an interval estimation problem on an unknown parameter based on the observations including errors. On such a problem, various methods to construct confidence intervals are proposed by many physicists and others (see, e.g. Feldman and Cousins (1998) and Mandelkern (2002)). In such cases, a combined Bayesian-frequentist approach to confidence intervals are constructed, and its comparison with ordinary and Bayesian confidence intervals are done in normal cases (Akahira et al. (2005)).

In the interval estimation problem on the difference between two means in exponential and normal cases, the possibility of extending the definition of confidence intervals is discussed by Weerahandi (1995) and others. In such a problem, a systematic method of the construction of a confidence interval is proposed, and its application to the Behrens-Fisher type problem is given (Akahira (2002)).

There are various information inequalities in statistical estimation. For example, the Cramér-Rao inequality, the Bhattacharyya one, etc. are well known as the fact that the variance of all unbiased estimators can not be smaller than the lower bound under suitable regularity conditions (see, e.g. Zacks (1971)). We consider a family of distributions for which the second order Bhattacharyya bound becomes sharp, and a necessary and sufficient condition for the second order one to be sharp is given for a family of linear combinations of distributions which belong to the exponential family (Tanaka and Akahira (2003)).

From the Bayesian viewpoint, we consider the information inequality applicable to the non-regular case when the regularity conditions do not necessarily hold, and construct an estimator which minimizes locally the variance of any estimator satisfying weaker conditions than the unbiasedness from which an information inequality is derived (Akahira and Ohyauchi (2007)).

2. Ordinary, likelihood ratio (LR), Bayesian and combined Bayesian-frequentist confidence intervals

In this section we discuss ordinary, LR and Bayesian confidence intervals, adopt the Bayesian-frequentist confidence interval and can compare it with others, according to Akahira et.al (2005). Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with the normal distribution $N(\mu, \sigma_0^2)$, where $\mu > 0$ and σ_0^2 is known. Since the pivotal quantity $T(\bar{X}, \mu) := \sqrt{n}(\bar{X} - \mu)/\sigma_0$ is distributed according to $N(0,1)$, it follows that the interval

$$I(\bar{X}) := \left[\max\left\{0,\ \bar{X} - u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right\},\ \max\left\{0,\ \bar{X} + u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right\}\right]$$

is the ordinary confidence interval (c.i.) for μ of confidence coefficient (c.c.) $1-\alpha$, where $\bar{X} := (1/n)\sum_{i=1}^n X_i$ and $u_{\alpha/2}$ is the upper $100(\alpha/2)$ percentile. Note that, for $\bar{X} \le -1$, the c.i. $I(\bar{X})$ is degenerate in the case of 68.27% c.c., hence there is still room for improvement.

In order to improve the above, Feldman and Cousins (1998) consider the c.i. based on the acceptance region of the LR test as follows. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with the normal distribution $N(\mu, 1)$, where $\mu > 0$. Since the likelihood function L of μ, given $\bar{X} = \bar{x} := (1/n)\sum_{i=1}^n x_i$, is

$$L(\mu|\bar{x}) = (2\pi)^{-n/2}\exp\left[-\frac{1}{2}\left\{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right\}\right]$$

for $\mu > 0$, it follows that the maximum likelihood estimator (MLE) is given by $\hat{\mu}_{ML} := \max\{\bar{X}, 0\}$. Then the LR is

$$R_\mu(\bar{X}) := \frac{L(\mu|\bar{x})}{L(\hat{\mu}_{ML}|\bar{x})} = \begin{cases} \exp\left\{-\frac{n}{2}(\bar{x} - \mu)^2\right\} & \text{for } \bar{x} \ge 0, \\ \exp\left\{n\left(\mu\bar{x} - \frac{\mu^2}{2}\right)\right\} & \text{for } \bar{x} < 0. \end{cases}$$

If we can obtain $a(\mu)$ and $b(\mu)$ such that $R(a(\mu)) = R(b(\mu))$ and $1 - \alpha = \Phi(\sqrt{n}(b(\mu) - \mu)) - \Phi(\sqrt{n}(a(\mu) - \mu))$, then the interval $[a(\mu), b(\mu)]$ is an acceptance interval, where Φ is the cumulative distribution function (c.d.f.) of $N(0,1)$. If $S(\bar{X}) := \{\mu | \bar{X} \in (a(\mu), b(\mu))\}$ is an interval, then it is the LR c.i. for μ of c.c. $1-\alpha$.

Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables with the normal distribution $N(\mu, 1)$, where $\mu > 0$. Let $\pi(\mu)$ be the improper prior distribution $\pi(\mu) = 1$ for $\mu > 0$; $= 0$ for $\mu \le 0$. Since the posterior density of μ given $\bar{X} = \bar{x}$ is $f(\mu|\bar{x}) = [\sqrt{n}\exp\{-(n/2)(\mu - \bar{x})^2\}]/\{\sqrt{2\pi}\Phi(\sqrt{n}\bar{x})\}$, it follows that the interval $[\max\{\bar{X}-d,0\},\bar{X}+d]$ is the Bayesian c.i. for μ of c.c. $1-\alpha$, where

$$d=\begin{cases}\frac{1}{\sqrt{n}}\Phi^{-1}\left(1-\alpha\Phi(\sqrt{n}\bar{X})\right) & \text{for } \bar{X}\le x_0,\\ \frac{1}{\sqrt{n}}\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2}(1-\alpha)\Phi(\sqrt{n}\bar{X})\right) & \text{for } \bar{X}>x_0\end{cases}$$

with $x_0=(1/\sqrt{n})\Phi^{-1}(1/(1+\alpha))$ (see Mandelkern (2002)).

Under the same setup as the above, we construct confidence intervals for μ by the combined Bayesian-frequentist approach. We consider the problem of testing hypothesis H: $\mu=\mu_0(>0)$ against K: $\mu\sim\pi(\mu)$, i.e. μ is distributed as the same improper prior as the above. Then the acceptance region of the most powerful (MP) test is of the form $T(\bar{X},\mu_0):=\Phi(\sqrt{n}\bar{X})/\{\sqrt{n}\phi(\sqrt{n}(\bar{X}-\mu_0))\}\le\lambda$, where ϕ is a density of $N(0,1)$, and, for a given α $(0<\alpha<1)$, λ is determined by

$$\alpha=\int_{\{y|\Phi(\sqrt{n}(y+\mu_0))/(\sqrt{n}\phi(\sqrt{n}y))>\lambda\}}\sqrt{n}\phi(\sqrt{n}u)du.$$

Let $\underline{y}(\mu_0)$ and $\bar{y}(\mu_0)$ be solutions of the equation $\Phi(\sqrt{n}(y+\mu_0))/\{\sqrt{n}\phi(\sqrt{n}y)\}=\lambda$. Putting $\bar{z}=\sqrt{n}\bar{y}$, $\underline{z}=\sqrt{n}\underline{y}$ and $m:=\sqrt{n}\mu_0$, we have the acceptance region $[(\underline{z}+m)/\sqrt{n},\ (\bar{z}+m)/\sqrt{n}]$, hence we construct a confidence interval by the combined Bayesian-frequentist approach.

Next, as a proper prior distribution, we consider the exponential distribution with the density $\pi_\theta(\mu)=(1/\theta)e^{-\mu/\theta}$ for $\mu>0$; $=0$ for $\mu\le 0$, where $\theta>0$. Note that the proper prior $\pi_\theta(\mu)$ converges to the improper uniform prior of type $\pi(\mu)$ as $n\to\infty$ in the sense that $\theta\pi_\theta(\mu)\to 1$ as $\theta\to\infty$. In a similar way to the above, it is shown that the acceptance region of the MP test is of the form

$$T(\bar{X},\mu_0):=\frac{1}{\theta}\Phi\left(\sqrt{n}\bar{X}-\frac{1}{\sqrt{n}\theta}\right)\exp\left\{-\frac{n}{2}\left(\frac{2\bar{X}}{n\theta}-\frac{1}{n^2\theta^2}\right)\right\}\Big/\{\sqrt{n}\phi(\sqrt{n}(\bar{X}-\mu_0))\}\le\lambda.$$

Here, for given α $(0<\alpha<1)$, λ is determined by

$$\alpha=\int_{\{\bar{x}|T(\bar{x},\mu_0)>\lambda\}}\sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n}{2}(\bar{x}-\mu_0)^2\right\}d\bar{x}.$$

Letting $t_1(\mu_0)$ and $t_2(\mu_0)$ $(t_1(\mu_0)<t_2(\mu_0))$ be solutions of the equation $T(\bar{x},\mu_0)=\lambda$, we have the acceptance region $[t_1(\mu_0),\ t_2(\mu_0)]$, hence we conctruct a confidence interval by the Bayesian-frequentist approach.

We can numerically compare ordinary, LR, Bayesian and combined Bayesian confidence limits.

3. Confidence intervals for the difference of means

In this section, we propose a systematic method of the construction of a confidence interval for the difference between two means in exponential and gamma cases, and apply a similar method to the Behrens-Fisher type problem, according to Akahira (2002). Suppose that $X_1,\cdots,X_n$ are i.i.d. random variables according to the exponential distribution $Exp(\theta_1)$ with a density $f(x,\theta_1)=(1/\theta_1)e^{-x/\theta_1}$ for $x\ge 0$; $=0$ for $x<0$, where $\theta_1>0$, and that $Y_1,\cdots,Y_n$ are i.i.d. random variables according to the exponential distribution $Exp(\theta_2)$, where θ_1 and θ_2 are unknown. We also assume that $X_1,\cdots,X_n,\ Y_1,\cdots,Y_n$ are independent. Let $Z_x:=\sum_{i=1}^n X_i$ and $Z_y:=\sum_{i=1}^n Y_i$. Then Z_x and Z_y are independently distributed as the gamma distributions

$G(n,\theta_1)$ and $G(n,\theta_2)$ with densities $f(z,\theta_j) = \theta_j^{-n}\{\Gamma(n)\}^{-1}z^{n-1}e^{-z/\theta_j}$ for $z \geq 0$; $= 0$ for $z < 0$, for $j = 1,\ 2$, respectively, where $\theta_j > 0$ $(j = 1, 2)$ and $n > 0$. For any α $(0 < \alpha < 1)$, let $[b(Z_x, Z_y),\ a(Z_x, Z_y)]$ be a confidence interval for $n(\theta_1 - \theta_2)$ of c.c. $1 - \alpha$, i.e.

$$P_{\theta_1,\theta_2}\{b(Z_x, Z_y) \leq n(\theta_1 - \theta_2) \leq a(Z_x, Z_y)\} \geq 1 - \alpha \tag{1}$$

for all θ_1 and θ_2. Here, in order that the confidence interval bound on (Z_x, Z_y) coincides with that based on (Z_y, Z_x), it is necessary that

$$b(Z_x, Z_y) = -a(Z_y, Z_x) \ \text{ a.e.} \tag{2}$$

So, we assume that (2) holds. Let $T_x := Z_x/\theta_1$ and $T_y := Z_y/\theta_2$. Then T_x and T_y are i.i.d. according to the gamma distribution $G(n, 1)$. From (1) and (2) we have

$$P_{\theta_1,\theta_2}\{a(\theta_1 T_x, \theta_2 T_y) < n(\theta_1 - \theta_2)\} + P_{\theta_1,\theta_2}\{a(\theta_2 T_y, \theta_1 T_x) < n(\theta_2 - \theta_1)\} \leq \alpha. \tag{3}$$

Further we assume that $a(z_1, z_2) = z_1\tilde{a}(z_1/(z_1 + z_2))$ for almost all $z_1 > 0$ and $z_2 > 0$, where $\tilde{a}(\cdot)$ is a positive-valued function defined on the interval $(0, 1)$. Letting $U := T_x/(T_x + T_y)$ and $V := T_x + T_y$, we see that U and V are independent, U is distributed as the beta distribution $Be(n, n)$ and V is distributed as $Be(2n, 1)$. Let $\delta := \theta_2/\theta_1$ and F_V be a c.d.f. of V. From (3) we have the following.

Theorem 1 Under the above conditions, let

$$\begin{aligned} p(\delta) :=& E_\delta\left[F_V\left(\frac{n(1-\delta)}{U\tilde{a}(U/(U+\delta(1-U)))}\right)\right]\chi_{(0,1]}(\delta) \\ &+ E_\delta\left[F_V\left(\frac{n(1-(1/\delta))}{U\tilde{a}((1-(1/\delta))/(U+(1/\delta)(1-U)))}\right)\right]\chi_{(1,\infty)}(\delta) \leq \alpha, \end{aligned} \tag{4}$$

where $\chi_A(\cdot)$ denotes the indicator of a set A. If $\tilde{a}(\cdot)$ satisfies (4) uniformly in δ, then

$$P_{\theta_1,\theta_2}\{-a(Z_y, Z_x) \leq n(\theta_1 - \theta_2) \leq a(Z_x, Z_y)\} \geq 1 - \alpha,$$

that is, $[-a(Z_y, Z_x),\ a(Z_x, Z_y)]$ is the confidence interval for $n(\theta_1 - \theta_2)$ of c.c. $1 - \alpha$.

Next, suppose that $X_1, \cdots, X_n$ are i.i.d. random variables according to the normal distribution $N(\mu_1, \sigma_1^2)$ and $Y_1, \cdots, Y_n$ are random variables with the normal distribution $N(\mu_2, \sigma_2^2)$, where μ_1, μ_2, σ_1^2 and σ_2^2 are unknown. Then we consider the Behrens-Fisher type problem, i.e. a confidence interval for the difference $\mu_1 - \mu_2$ of normal means. Let $\bar{X} := (1/n)\sum_{i=1}^n X_i$, $\bar{Y} := (1/n)\sum_{i=1}^n Y_i$, $S_x^2 := \sum_{i=1}^n (X_i - \bar{X})^2$, $S_y^2 := \sum_{i=1}^n (Y_i - \bar{Y})^2$. For any α $(0 < \alpha < 1)$, let

$$P_\theta\left\{\bar{X} - \bar{Y} + h(S_x^2, S_y^2) \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + g(S_x^2, S_y^2)\right\} \geq 1 - \alpha$$

for all θ, where $\theta := (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. Here, in order to that the confidence interval based on $(\bar{X}, \bar{Y}, S_x^2, S_y^2)$ coincides with that based on $(\bar{Y}, \bar{X}, S_y^2, S_x^2)$, it is necessary that $h(x, y) = -g(y, x)$ a.e. which is assumed here. We also assume that $g(c^2x, c^2y) = cg(x, y)$ for any positive constant c. Let $W_x := S_x^2/\sigma_1^2$ and $W_y := S_y^2/\sigma_2^2$.

Theorem 2 Assume that the above conditions hold. Let F_T be the c.d.f. of the t-distribution with $2(n-1)$ degrees of freedom and B be a random variable according to the Beta distribution $Be((n-1)/2, (n-1)/2)$. If g satisfies

$$r(\delta) := E_\delta\left[F_T\left(\sqrt{2(n-1)}g(n\delta B, n(1-\delta)(1-B))\right)\right]$$

$$+ E_\delta \left[F_T \left(\sqrt{2(n-1)} g(n(1-\delta)(1-B), n\delta B) \right) \right] \geq 2 - \alpha$$

uniformly in δ $(0 < \delta < 1)$, then $\left[\bar{X} - \bar{Y} - g(S_y^2, S_x^2),\ \bar{X} - \bar{Y} + g(S_x^2, S_y^2)\right]$ is a confidence interval for $\mu_1 - \mu_2$ of c.c. $1 - \alpha$.

Remark The coverage probability of the above c.i. is given by

$$P_\theta \left\{ \bar{X} - \bar{Y} - g(S_y^2, S_x^2) \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + g(S_x^2, S_y^2) \right\} = r(\delta) - 1,$$

where $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ and $\delta = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$.

Assume that $g(x, y) = \sqrt{x+y}\,\tilde{g}(x/(x+y))$ for almost all $x > 0$, $y > 0$, where $\tilde{g}$ is a real-valued function defined on the interval $(0, 1)$. Here, we consider $\tilde{g}(z) = a(z-b)^2 + c$ for $0 < z < 1$, where a, b, and c are constants with $a \geq 0$ and $0 < b < 1$. If we can find the constants a, b, and c satisfying $r(\delta) \geq 2 - \alpha$ uniformly in δ $(0 < \delta < 1)$, then, from Theorem 2

$$\left[\bar{X} - \bar{Y} - \sqrt{S_x^2 + S_y^2} \left\{ a \left(\frac{S_y^2}{S_x^2 + S_y^2} - b \right)^2 + c \right\}, \right.$$
$$\left. \bar{X} - \bar{Y} + \sqrt{S_x^2 + S_y^2} \left\{ a \left(\frac{S_x^2}{S_x^2 + S_y^2} - b \right)^2 + c \right\} \right]$$

is a confidence interval for $\mu_1 - \mu_2$ of c.c. $1 - \alpha$.

4. Information inequalities

In this section, first, a family of distributions for which an unbiased estimator of a function $g(\theta)$ of a real parameter θ can attain the second order Bhattacharyya lower bound is derived according to Tanaka and Akahira (2003). Suppose that $(\mathcal{X}, \mathcal{B})$ is a sample space and a family $\mathcal{P} = \{P_\theta :\ \theta \in \Theta\}$ is dominated with respect to a σ-finite measure μ, where Θ is an open interval of $\mathbf{R}^1$. Denote by $f(x, \theta) = dP_\theta/d\mu$ $(\theta \in \Theta)$ a probability density function (p.d.f.). We consider an estimation problem of the U-estimable function $g(\theta)$, i.e. the function $g(\theta)$ for which its unbiased estimator with a finite variance exists, based on a sample X. Here, we assume the following conditions (A1) to (A4).

(A1) For almost all $x[\mu]$, there exist $(\partial^i/\partial\theta^i) f(x, \theta)$ for $i = 1, \cdots, k$.

(A2) For each $i = 1, \cdots, k$, there exist $\mathcal{B}$-measurable functions $M_i(x) > 0$ such that $|(\partial^i/\partial\theta^i) f(x, \theta)| \leq M_i(x)$ for all $\theta \in \Theta$, and $\int M_i(x) d\mu(x) < \infty$.

(A3) $\int |((\partial^i/\partial\theta^i) f(x, \theta)(\partial^j/\partial\theta^j) f(x, \theta))/f(x, \theta)| d\mu(x) < \infty$ for $i,\ j = 1, \cdots, k$ and for all $\theta \in \Theta$.

(A4) For almost all $x[\mu]$ and for all $\theta \in \Theta$, $f(x, \theta) > 0$.

Theorem 3 (Bhattacharyya (1946), Zacks (1971)). Suppose that the conditions (A1) to (A4) hold. Assume that $g(\theta)$ is a U-estimable function which is k-times differentiable over Θ. Let $\mathbf{g}(\theta) = {}^t(g^{(1)}(\theta), \cdots, g^{(k)}(\theta))$, where $g^{(i)}(\theta)$ is the i-th order derivative of $g(\theta)$. Let $\hat{g}(X)$ be an unbiased estimator of $g(\theta)$ having a finite variance, and assume that, for $i = 1, \cdots, k$, there exists a function $N_i(x)$ such that $|\hat{g}(x)(\partial^i/\partial\theta^i) f(x, \theta)| \leq N_i(x)$ for all $\theta \in \Theta$, and $\int N_i(x) d\mu(x) < \infty$. Furthermore, let $I(\theta)$ be a $k \times k$ non-negative definite matrix with elements $I_{ij}(\theta) = E_\theta[\{\partial^i f(X, \theta)/\partial\theta^i\}\{\partial^j f(X, \theta)/\partial\theta^j\}/\{f(X, \theta)\}^2]$, $(i,\ j = 1, \cdots, k)$. Then, if $I(\theta)$ is non-singular over Θ,

$$V_\theta(\hat{g}(X)) \geq {}^t\mathbf{g}(\theta) I(\theta)^{-1} \mathbf{g}(\theta) =: B_k(\theta). \tag{5}$$

And the equality holds in (5) if and only if

$$\hat{g}(x) - g(\theta) = \sum_{i=1}^{k} a_i(\theta) \frac{\partial^i f(x,\theta)/\partial\theta^i}{f(x,\theta)} \quad \text{a.a. } x[\mu] \tag{6}$$

for all $\theta \in \Theta$, where ${}^t(a_1(\theta), \cdots, a_k(\theta)) = I(\theta)^{-1t}(g^{(1)}(\theta), \cdots, g^{(k)}(\theta))$.

Theorem 4 Suppose that the conditions (A1) to (A4) hold. Assume that $\mu(\{x \in \mathcal{X} | \hat{g}(x) = r\}) = 0$ for all $r \in \mathbf{R}^1$. Let $k = 2$ and $a_2(\theta) \neq 0$ for all $\theta \in \Theta$. Then the solution of (6) is expressed by a linear combination of distributions from the exponential family if and only if the following (i) and (ii) hold.

(i) There are a function $t(x)$ and constants C_0, C_1 and C_2 such that C_2 has the same sign as $a_2(\theta)$, and $\hat{g}(x)$ is of the form $\hat{g}(x) = C_2 t^2(x) + C_1 t(x) + C_0$.

(ii) For C_0, C_1 and C_2 given in (i), $g(\theta)$ has the form

$$g(\theta) = \frac{a_1^2(\theta)}{4a_2(\theta)} - \frac{a_1(\theta)a_2'(\theta)}{2a_2(\theta)} + \frac{a_1'(\theta)}{2} - \frac{a_2''(\theta)}{4} + \frac{3(a_2'(\theta))^2}{16a_2(\theta)} + C_0 - \frac{C_1^2}{4C_2}.$$

Next, according to Akahira and Ohyauchi (2007), we consider the information inequality from the Bayesian viewpoint. Then it is shown that an estimator minimizing locally the variance of any estimator with weaker conditions than the unbiasedness is constructed, and also the lower bound for the variance of estimators can be expressed by the information inequality which involves the Hammersley-Chapman-Robbins type inequality as its special case.

References

Akahira, M. (2002). Confidence intervals for the difference of means: application to the Behrens-Fisher type problem. *Statistical Papers* **43**, 273–284.

Akahira, M. and Ohyauchi, N. (2007). A Bayesian view of the Hammersley-Chapman-Robbins-type inequality. *Statistics* **41**(2), 137–144.

Akahira, M., Shimizu, A. and Takeuchi, K. (2005). The construction of combined Bayesian-frequentist confidence intervals for a positive parameter. *Statistics* **65**(4), 656–670.

Bhattacharyya, A. (1946). On some analogues of the information and their use in statistical estimation. *Sankhyā*, **8**, 1–14.

Feldman, G. J. and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Review*, **D57**, 3873–3889.

Mandelkern, M. (2002). Setting confidence intervals for bounded parameters. *Statistical Science* **17**, 149–172.

Tanaka, H. and Akahira, M. (2003). On a family of distributions attaining the Bhattacharyya bound. *Ann. Inst. Statist. Math.* **55**(2), 309–317.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis.* Springer, New York.

Zacks, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.