Optimization of Interval Estimators via Invariant Embedding Technique

Nicholas A. Nechval, Konstantin N. Nechval and Edgars K.Vasermanis

Department of Mathematical Statistics
University of Latvia
Raina Blvd 19, LV-1586 Riga, Latvia
Fax: +371-7100660, E-mail: nechval@junik.lv

Abstract

In the present paper, for optimization of interval estimators, a new technique of invariant embedding of sample statistics in a loss function is proposed. This technique represents a simple and computationally attractive statistical method based on the constructive use of the invariance principle in mathematical statistics. Unlike the Bayesian approach, an invariant embedding technique is independent of the choice of priors. The aim of the paper is to show how the invariance principle may be employed in the particular case of finding the interval estimators that are uniformly best invariant. The technique proposed here is a special case of more general considerations applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space. This technique may be used for constructing the minimum risk estimators of state of computing anticipatory systems. To illustrate the proposed technique, examples are given.

Keywords: Interval estimation; Risk function; Minimization; Best invariant estimator

1 Introduction

Estimation, in both theory and practice, tends to be divided into two almost distinct branches. On the one hand, point estimation, by providing a working value for the unknown true parameter value, is undoubtedly useful in further investigations of the given situation although its reliability may be difficult to assess in real terms through the medium of the estimated standard error. On the other hand, confidence interval estimation, by presenting a whole set of more or less plausible values of the parameter, is often less easy to apply but has a more direct assessment of reliability through the confidence coefficient. The dilemma in estimation arises essentially from this need for a balance between usefulness and reliability – usefulness through the practical advantages of narrowing down the set of plausible values and the increasing reliability, which attends the enlarging of the set. This dilemma is one of the sources of awkward

International Journal of Computing Anticipatory Systems, Volume 9, 2001
Edited by D. M. Dubois, CHAOS, Liège, Belgium, ISSN 1373-5411 ISBN 2-9600262-2-5

questions from users about estimation. What confidence coefficient should be used? Why not use the whole parameter space as confidence interval and so ensure 100% confidence? Why use a point estimate when it is almost certainty not the true parameter value? While answers to such naive questions can be expressed in terms of the distribution of the estimator or the class of confidence intervals associated with various confidence coefficients, they are not the only solution to the dilemma, and, in our view, are not readily appreciated by users.

Any serious attempt to take account of the consequences of unreliability in not capturing the true parameter value and of lack of usefulness in excessive width should, we feel, involve the specification of some reasonable loss function and the subsequent examination of the problem in terms of decision theory. Such an attempt is usually beset by the well-known difficulties of the non-existence of a uniformly best solution in a frequentist approach and of the assessment of the prior distribution in a Bayesian approach. There is, however, one unexploited specification of the decision-theory approach to estimation which has some degree of realism and for which a satisfactory frequentist solution can be readily obtained.

This paper considers the consequences of adopting a piecewise-linear loss function for a situation where interval estimators are required for location or scale parameters. Under certain circumstances intervals that are uniformly best invariant, with respect to the group of location and scale changes, can be found for this frequentist decision problem. Such interval estimators, while of interest in their own right as solutions of the particular decision problem, may of course be regarded as confidence intervals in the conventional sense and the appropriate equivalent confidence coefficient evaluated. This confidence coefficient is simply related to the constants specifying the loss function, and has an extremely simple approximation that provides some insight into the concept of confidence interval estimation. When the interval collapses to a point we can relate the variance of the point estimator to the constants of the loss function. We have thus a decision framework from which either point or interval estimators may arise depending on the constants of the loss function.

The outline of the paper is as follows. An invariant embedding technique is presented in Section 2. Formulation of the problem and interval estimation of a location parameter is given in Section 3. Section 4 is devoted to interval estimation of a scale parameter.

2 Invariant Embedding Technique

This paper is concerned with the implications of group theoretic structure for invariant loss functions. We present an invariant embedding technique (Nechval, 1982, 1984, 2000) based on the constructive use of the invariance principle in mathematical statistics. This technique allows one to solve many problems of the theory of statistical inferences in a simple way. The aim of the paper is to show how the invariance principle may be employed in the particular case of finding the interval estimators that are uniformly best invariant. The technique used here is a special case of more general considerations applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space.

2.1 Preliminaries

Our underlying structure consists of a class of probability models ($\mathscr{X}$, $\mathscr{A}$, $\mathscr{P}$), a one-one mapping ψ taking $\mathscr{P}$ onto an index set Θ, a measurable space of actions ($\mathscr{D}$, $\mathscr{B}$), and a real-valued loss function r defined on $\Theta \times \mathscr{D}$. We assume that a group G of one-one $\mathscr{A}$ - measurable transformations acts on $\mathscr{X}$ and that it leaves the class of models ($\mathscr{X}$, $\mathscr{A}$, $\mathscr{P}$) invariant. We further assume that homomorphic images $\overline{G}$ and $\widetilde{G}$ of G act on Θ and $\mathscr{D}$, respectively. ($\overline{G}$ may be induced on Θ through ψ; $\widetilde{G}$ may be induced on $\mathscr{D}$ through r). We shall say that r is invariant if for every $(\theta,d) \in \Theta \times \mathscr{D}$

$$r(\overline{g}\theta, \widetilde{g}d) = r(\theta,d), \quad g \in G. \tag{1}$$

Given the structure described above there are aesthetic and sometimes admissibility grounds for restricting attention to estimators (decision rules) φ: $\mathscr{X} \to \mathscr{D}$ which are $(G, \widetilde{G})$ equivariant in the sense that

$$\varphi(gX) = \widetilde{g}\varphi(X), \quad X \in \mathscr{X}, \quad g \in G. \tag{2}$$

If $\overline{G}$ is trivial and (1), (2) hold, we say φ is G-invariant, or simply invariant (Lehmann, 1959).

2.2 Invariant Functions

We begin by noting that r is invariant in the sense of (1) if and only if r is a $G^{\bullet}$-invariant function, where $G^{\bullet}$ is defined on $\Theta \times \mathscr{D}$ as follows: to each $g \in G$, with homomorphic images $\overline{g}, \widetilde{g}$ in $\overline{G}, \widetilde{G}$ respectively, let $g^{\bullet}(\theta,d) = (\overline{g}\theta, \widetilde{g}d)$, $(\theta,d) \in (\Theta \times \mathscr{D})$. It is assumed that $\widetilde{G}$ is a homomorphic image of $\overline{G}$.

Definition 1 (*Transitivity*). A transformation group $\overline{G}$ acting on a set Θ is called (uniquely) transitive if for every θ, $\vartheta \in \Theta$ there exists a (unique) $\overline{g} \in \overline{G}$ such that $\overline{g}\theta = \vartheta$.

When $\overline{G}$ is transitive on Θ we may index $\overline{G}$ by Θ: fix an arbitrary point $\theta \in \Theta$ and define $\overline{g}_{\theta_1}$ to be the unique $\overline{g} \in \overline{G}$ satisfying $\overline{g}\theta = \theta_1$. The identity of $\overline{G}$ clearly corresponds to θ. An immediate consequence is Lemma 1.

Lemma 1 (*Transformation*). Let $\overline{G}$ be transitive on Θ. Fix $\theta \in \Theta$ and define $\overline{g}_{\theta_1}$ as above. Then $\overline{g}_{\overline{q}\theta_1} = \overline{q}\overline{g}_{\theta_1}$ for $\theta \in \Theta$, $\overline{q} \in \overline{G}$.

Proof. The identity $\overline{g}_{\overline{q}\theta_1}\theta = \overline{q}\theta_1 = \overline{q}\overline{g}_{\theta_1}\theta$ shows that $\overline{g}_{\overline{q}\theta_1}$ and $\overline{q}\overline{g}_{\theta_1}$ both take θ into $\overline{q}\theta_1$, and the lemma follows by unique transitivity. □

Theorem 1 (*Maximal Invariant*). Let $\overline{G}$ be transitive on Θ. Fix a reference point $\theta_0 \in \Theta$ and index $\overline{G}$ by Θ. A maximal invariant M with respect to $G^\bullet$ acting on $\Theta \times \mathscr{D}$ is defined by

$$M(\theta, d) = \widetilde{g}_\theta^{-1} d, \quad (\theta, d) \in \Theta \times \mathscr{D}. \tag{3}$$

Proof. For each $(\theta,d) \in (\Theta \times \mathscr{D})$ and $\overline{g} \in \overline{G}$

$$M(\overline{g}\theta, \widetilde{g}d) = (\widetilde{g}_{\overline{g}\theta}^{-1})\widetilde{g}d = (\widetilde{g}\widetilde{g}_\theta)^{-1}\widetilde{g}d = \widetilde{g}_\theta^{-1}\widetilde{g}^{-1}\widetilde{g}d = \widetilde{g}_\theta^{-1}d = M(\theta, d) \tag{4}$$

by Lemma 1 and the structure preserving properties of homomorphisms. Thus M is $G^\bullet$-invariant. To see that M is maximal, let $M(\theta_1,d_1)=M(\theta_2,d_2)$. Then $\widetilde{g}_{\theta_1}^{-1}d_1 = \widetilde{g}_{\theta_2}^{-1}d_2$ or $d_1=\widetilde{g}\, d_2$ where $\widetilde{g} = \widetilde{g}_{\theta_1}\widetilde{g}_{\theta_2}^{-1}$. Since $\theta_1=\overline{g}_{\theta_1}\theta_0 = \overline{g}_{\theta_1}\overline{g}_{\theta_2}^{-1}\theta_2 = \overline{g}\theta_2$, $(\theta_1,d_1)=g^\bullet(\theta_2,d_2)$ for some $g^\bullet \in G^\bullet$, and the proof is complete. □

Corollary 1.1 (*Invariant Embedding*). An invariant loss function, $r(\theta,d)$, can be transformed as follows:

$$r(\theta, d) = r(\overline{g}_{\hat{\theta}}^{-1}\theta, \widetilde{g}_{\hat{\theta}}^{-1}d) = \ddot{r}(v, \eta), \tag{5}$$

where $v=v(\theta,\hat{\theta})$ is a function (it is called a pivotal quantity) such that the distribution of v does not depend on θ; $\eta=\eta(d,\hat{\theta})$ is an ancillary factor; $\hat{\theta}$ is the maximum likelihood estimator of θ (or the sufficient statistic for θ).

Corollary 1.2 (*Best Invariant Estimator*). The best invariant estimator (decision rule) is given by

$$\varphi^*(X) = d^* = \eta^{-1}(\eta^*, \hat{\theta}), \tag{6}$$

where

$$\eta^* = \arg\inf_\eta E_\eta\{\ddot{r}(v, \eta)\}. \tag{7}$$

Corollary 1.3 (*Risk*). A risk function

$$R(\theta,\varphi(X)) = E_\theta\{r(\theta,\varphi(X))\} = E_{\eta_\circ}\{\ddot{r}(v_\circ,\eta_\circ)\} \tag{8}$$

is constant on orbits when an invariant estimator (decision rule) $\varphi(X)$ is used, where $v_\circ = v_\circ(\theta,X)$ is a function (pivotal quantity) whose distribution does not depend on θ; $\eta_\circ = \eta_\circ(d,X)$ is an ancillary factor.

Consider, for instance, the problem of estimating the location-scale parameter of a distribution belonging to a family generated by a continuous cdf F: $\mathcal{P}=\{P_\theta: F((x-\mu)/\sigma), x\in R, \theta\in\Theta\}$, $\Theta=\{(\mu,\sigma): \mu,\sigma\in R, \sigma>0\}=\mathcal{D}$. The group G of location and scale changes leaves the class of models invariant. Since $\overline{G}$ induced on Θ by $P_\theta \to \theta$ is uniquely transitive, we may apply Theorem 1 and obtain invariant loss functions of the form

$$r(\theta,\varphi(X)) = r[(\varphi_1(X)-\mu)/\sigma, \varphi_2(X)/\sigma] \tag{9}$$

if $\theta=(\mu,\sigma)$ and $\varphi(X)=(\varphi_1(X), \varphi_2(X))$. Let $\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2)$, $d=(d_1,d_2)$, then

$$r(\theta,d) = r[(d_1-\mu)/\sigma, d_2/\sigma] = r(v_1+\eta_1 v_2, \eta_2 v_2) = \ddot{r}(v,\eta), \tag{10}$$

where $v=(v_1,v_2)$, $v_1=(\hat{\theta}_1-\mu)/\sigma$, $v_2=\hat{\theta}_2/\sigma$; $\eta=(\eta_1,\eta_2)$, $\eta_1=(d_1-\hat{\theta}_1)/\hat{\theta}_2$, $\eta_2=d_2/\hat{\theta}_2$.

3 Interval Estimation of a Location Parameter

3.1 Problem Statement

Consider a situation described by one of a family of density functions, indexed by the vector parameter $\theta=(\mu,\sigma)$, where μ and $\sigma(>0)$ are respectively parameters of location and scale. For this family, invariant under the group of positive linear transformations: $x\to ax+b$ with $a>0$, we shall assume that there is obtainable from some informative experiment (a random sample of observations $X=(x_1, \ldots ,x_n)$) a sufficient statistic (m_n,s_n) for (μ,σ) with density function $p(m_n,s_n;\mu,\sigma)$ of the form

$$p(m_n,s_n;\mu,\sigma) = \sigma^{-2} f[(m_n-\mu)/\sigma, s_n/\sigma]. \tag{11}$$

We are thus assuming that for the family of density functions an induced invariance holds under the group G of transformations: $m_n\to am_n+b$, $s_n\to as_n$ ($a>0$). The family of density functions satisfying the above conditions is, of course, the limited one of normal, negative exponential, Weibull and gamma (with known index) density functions. The structure of the problem is, however, more clearly seen within the general framework.

Suppose that we assert that an interval (d_1,d_2) contains μ. If, as is usually the case, the purpose of this interval statement is to convey useful information we incur penalties if d_1 lies above μ or if d_2 falls below μ. Suppose that these penalties are $c_1(d_1-\mu)$ and $c_2(\mu-d_2)$, losses proportional to the amounts by which μ escapes the interval. Since c_1 and c_2 may be different the possibility of differential losses associated with the interval overshooting and undershooting the true μ is allowed. In addition to these losses there will be a cost attaching to the length of interval used. For example, it will be more difficult and more expensive to design or plan when the interval $d=(d_1,d_2)$ is wide. Suppose that the cost associated with the interval is proportional to its length, say $c(d_2-d_1)$. In the specification of the loss function, σ is clearly a 'nuisance parameter' and no alteration to the basic decision problem is caused by multiplying all loss factors by $1/\sigma$. Thus we are led to investigate the piecewise-linear loss function

$$r(\theta,d)=\begin{cases}\dfrac{c_1(d_1-\mu)}{\sigma}+\dfrac{c(d_2-d_1)}{\sigma} & (\mu<d_1),\\ \dfrac{c(d_2-d_1)}{\sigma} & (d_1\le\mu\le d_2),\\ \dfrac{c(d_2-d_1)}{\sigma}+\dfrac{c_2(\mu-d_2)}{\sigma} & (\mu>d_2).\end{cases} \tag{12}$$

The decision problem specified by the informative experiment density function (11) and the loss function (12) is invariant under the group G of transformations. Thus, the problem is to find the best invariant interval estimator of μ,

$$d^{*}=\arg\min_{d\in\mathcal{D}} R(\theta,d), \tag{13}$$

where $\mathcal{D}$ is a set of invariant interval estimators of μ, $R(\theta,d)=E_\theta\{r(\theta,d)\}$ is a risk function.

3.2 Best Invariant Estimator

It follows from Corollary 1.1 that an invariant loss function, $r(\theta,d)$, can be transformed as follows:

$$r(\theta,d)=\ddot{r}(v,\eta), \tag{14}$$

where

$$\ddot{r}(v,\eta)=\begin{cases}c_1(v_1+\eta_1v_2)+c(\eta_2-\eta_1)v_2 & (v_1>-\eta_1v_2),\\ c(\eta_2-\eta_1)v_2 & (-\eta_1v_2\ge v_1\ge-\eta_2v_2),\\ -c_2(v_1+\eta_2v_2)+c(\eta_2-\eta_1)v_2 & (v_1<-\eta_2v_2).\end{cases} \tag{15}$$

$v=(v_1,v_2)$, $v_1=(m_n-\mu)/\sigma$, $v_2=s_n/\sigma$; $\eta=(\eta_1,\eta_2)$, $\eta_1=(d_1-m_n)/s_n$, $\eta_2=d_2/s_n$.

It follows from (15) and Corollary 1.3 that the risk associated with d and θ can be expressed as

$$R(\theta,d)=E_\theta\{r(\theta,d)\}=E_\eta\{\ddot{r}(v,\eta)\}=c_1\int_0^\infty dv_2\int_{-\eta_1 v_2}^{\infty}(v_1+\eta_1 v_2)f(v_1,v_2)dv_1$$

$$-c_2\int_0^\infty dv_2\int_{-\infty}^{-\eta_2 v_2}(v_1+\eta_2 v_2)f(v_1,v_2)dv_1+c(\eta_2-\eta_1)\int_0^\infty v_2dv_2\int_{-\infty}^{\infty}f(v_1,v_2)dv_1, \qquad (16)$$

which is constant on orbits when an invariant estimator (decision rule) d is used, where $f(v_1,v_2)$ is defined by (11). The fact that the risk (16) is independent of θ means that a decision function $\eta=(\eta_1,\eta_2)$ which minimizes (16) is uniformly best invariant. The following theorem gives the central result in this section.

Theorem 2 (*Best Invariant Estimator of μ*). Suppose that (u_1,u_2) is a random vector having density function

$$u_2f(u_1,u_2)\left[\int_0^\infty u_2du_2\int_{-\infty}^{\infty}f(u_1,u_2)du_1\right]^{-1} \quad (u_1 \text{ real},\ \ u_2>0), \qquad (17)$$

where f is defined by (11), and let Q be the distribution function of u_1/u_2.

(i) If $c/c_1+c/c_2<1$ then the uniformly best invariant linear-loss interval estimator of μ is $d^*=(m_n+\eta_1 s_n,\ m_n+\eta_2 s_n)$, where

$$Q(-\eta_1)=1-c/c_1, \quad Q(-\eta_2)=c/c_2. \qquad (18)$$

(ii) If $c/c_1+c/c_2\geq 1$ then the uniformly best invariant linear-loss interval estimator degenerates into a point estimator $m_n+\eta_\bullet s_n$, where

$$Q(-\eta_\bullet)=c_1/(c_1+c_2). \qquad (19)$$

Proof. From (16)

$$\frac{\partial E_\eta\{\ddot{r}(v,\eta)\}}{\partial\eta_1}=c_1\int_0^\infty v_2dv_2\int_{-\eta_1 v_2}^{\infty}f(v_1,v_2)dv_1-c\int_0^\infty v_2dv_2\int_{-\infty}^{\infty}f(v_1,v_2)dv_1$$

$$=\int_0^\infty v_2dv_2\int_{-\infty}^{\infty}f(v_1,v_2)dv_1[c_1P\{(u_1,u_2):u_1+\eta_1u_2>0\}-c]$$

$$= \int_{0}^{\infty} v_2 dv_2 \int_{-\infty}^{\infty} f(v_1, v_2) dv_1 [c_1(1 - Q(-\eta_1)) - c]. \tag{20}$$

Similarly

$$\frac{\partial E_\eta \{\ddot{r}(v,\eta)\}}{\partial \eta_2} = \int_{0}^{\infty} v_2 dv_2 \int_{-\infty}^{\infty} f(v_1, v_2) dv_1 [c - c_2 Q(-\eta_2)]. \tag{21}$$

Now $\partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_1 = \partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_2 = 0$ if and only if (18) hold. We thus obtain one stationary value for $E_\eta\{\ddot{r}(v,\eta)\}$ provided (18) has a solution with $\eta_1<\eta_2$ and this is so if $1-c/c_1>c/c_2$. It is easily confirmed that this $\eta=(\eta_1,\eta_2)$ gives the minimum value of $E_\eta\{\ddot{r}(v,\eta)\}$. Thus (i) is established.

If $c/c_1+c/c_2\geq 1$ then the minimum of $E_\eta\{\ddot{r}(v,\eta)\}$ in the region $\eta_2\geq\eta_1$ occurs where $\eta_1=\eta_2=\eta_\bullet$, $\eta_\bullet$ being determined by setting $\partial E_\eta\{\ddot{r}(v,(\eta_\bullet,\eta_\bullet))\}/\partial\eta_\bullet=0$ and this reduces to

$$c_1[1 - Q(-\eta_\bullet)] - c_2 Q(-\eta_\bullet) = 0, \tag{22}$$

which establishes (ii). □

Corollary 2.1 (***Minimum Risk of the Best Invariant Estimator of*** μ). The minimum risk is given by

$$R(\theta, d^*) = E_\theta\{r(\theta, d^*)\} = E_\eta\{\ddot{r}(v,\eta)\}$$

$$= c_1 \int_{0}^{\infty} dv_2 \int_{-\eta_1 v_2}^{\infty} v_1 f(v_1, v_2) dv_1 - c_2 \int_{0}^{\infty} dv_2 \int_{-\infty}^{-\eta_2 v_2} v_1 f(v_1, v_2) dv_1 \tag{23}$$

for case (i) with $\eta=(\eta_1,\eta_2)$ as given by (18) and for case (ii) with $\eta_1=\eta_2=\eta_\bullet$ as given by (19).

Proof. These results are immediate from (16) when use is made of $\partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_1 = \partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_2 = 0$ in case (i) and $\partial E_\eta\{\ddot{r}(v,(\eta_\bullet,\eta_\bullet))\}/\partial\eta_\bullet=0$ in case (ii). □

The underlying reason why $c/c_1+c/c_2$ acts as a separator of interval and point estimation is that for $c/c_1+c/c_2\geq 1$ every interval estimator is inadmissible, there existing some point estimator with uniformly smaller risk.

3.3 Equivalent Confidence Coefficient

For case (i) when we obtain an interval estimator for μ we may regard the interval as a confidence interval in the conventional sense and evaluate its confidence coefficient. The general result is contained in the following theorem.

Theorem 3 (*Equivalent Confidence Coefficient*). Suppose that $v=(v_1,v_2)$ is a random vector having density function $f(v_1,v_2)$ (v_1 real, $v_2>0$) where f is defined by (11) and let F be the distribution function of v_1/v_2. Then the confidence coefficient associated with the optimum interval $d^*=(d_1,d_2)$, where $d_1=m_n+\eta_1 s_n$, $d_2=m_n+\eta_2 s_n$, is

$$P\{d^*: d_1 < \mu < d_2; \mu,\sigma\} = F\left[Q^{-1}(1-c/c_1)\right] - F\left[Q^{-1}(c/c_2)\right] \tag{24}$$

Proof. The confidence coefficient corresponding to (μ,σ) is given by

$$P\{(m_n,s_n): m_n+\eta_1 s_n < \mu < m_n+\eta_2 s_n; \mu,\sigma\} = \int_0^\infty ds_n \int_{\mu-\eta_2 s_n}^{\mu-\eta_1 s_n} p(m_n,s_n;\mu,\sigma)dm_n$$

$$= P\{(v_1,v_2): -\eta_2 < v_1/v_2 < -\eta_1\}$$

$$= F(-\eta_1) - F(-\eta_2) = F\left[Q^{-1}(1-c/c_1)\right] - F\left[Q^{-1}(c/c_2)\right] \tag{25}$$

This is independent of (μ,σ). □

Later when we examine specific forms for $p(m_n,s_n;\mu,\sigma)$ we shall compute this equivalent confidence coefficient. The way in which (24) varies with c, c_1 and c_2, and the fact that c_1 and c_2 are the factors of proportionality associated with losses from overshooting and undershooting relative to loss involved in increasing the length of interval, provides an interesting interpretation of confidence interval estimation.

3.4 Examples

Example 1 (*Two-Parameter Exponential Distribution*). Suppose that $X=(x_1, \ldots ,x_n)$ is a sample of independent random variables each with density function $(1/\sigma)\exp[-(x-\mu)/\sigma]$ $(x>\mu)$. Then (m_n,s_n), where

$$m_n = \min(x_1,\ldots,x_n), \quad s_n = n^{-1}\left[\sum_{i=1}^{n}(x_i - m_n)\right], \tag{26}$$

is a sufficient statistic for (μ,σ); also m_n and s_n are independently distributed, with

$$p(m_n,s_n;\mu,\sigma) = \frac{n}{\sigma}\exp\left[-\frac{n(m_n-\mu)}{\sigma}\right]\left(\frac{n}{\sigma}\right)^{n-1}\frac{s_n^{n-2}}{\Gamma(n-2)}\exp\left(-\frac{ns_n}{\sigma}\right) \quad (m_n>\mu, s_n>0), \tag{27}$$

so that

$$f(v_1,v_2) = ne^{-nv_1}n^{n-1}v_2^{n-2}e^{-nv_2}/\Gamma(n-1) \quad (v_1>0, v_2>0). \tag{28}$$

The distribution functions of Theorem 1 and 2 are

$$F(x) = 1-(1+x)^{-(n-1)}, \quad Q(x) = 1-(1+x)^{-n} \quad (x>0). \tag{29}$$

It follows immediately from (18) that, when $c/c_1+c/c_2<1$, the best invariant interval estimator is

$$d^* = \left(m_n - [(c_1/c)^{1/n} - 1]s_n, m_n - [(1-c/c_2)^{-1/n} - 1]s_n\right) \tag{30}$$

The equivalent confidence coefficient is also easily calculated as

$$P\{d^*: d_1 < \mu < d_2; \mu, \sigma\} = \left(1 - \frac{c}{c_2}\right)^{1-1/n} - \left(\frac{c}{c_1}\right)^{1-1/n}. \tag{31}$$

If, say, $c=5$, $c_1=c_2=200$, and $n=5$, then $P\{d^*:d_1<\mu<d_2;\mu,\sigma\}=0.928$.

When $c/c_1+c/c_2\geq 1$ the point estimator is, by (19),

$$m_n - \left[\left(1 + \frac{c_1}{c_2}\right)^{1/n} - 1\right]s_n. \tag{32}$$

Only for large n, or for c_1 small compared with c_2, will this be near m_n.

Example 2 (*Normal Distribution*). In many situations the assumption of a normal distribution with unknown mean μ (location parameter) and unknown standard deviation σ (scale parameter) is made. There will then usually be available a sufficient statistic (m_n,s_n) for (μ,σ) from the informative $N(\mu,\sigma^2)$ experiment (sample of independent random variables $X=(x_1, \dots ,x_n)$), where

$$m_n = n^{-1}\sum_{i=1}^{n} x_i, \quad s_n = \left[(n-1)^{-1}\sum_{i=1}^{n}(x_i - m_n)^2\right]^{1/2} \tag{33}$$

are independently distributed, m_n with distribution $N(\mu,\sigma^2/n)$ and s_n with distribution $[\sigma/\sqrt{(n-1)}]\sqrt{\chi^2(n-1)}$, so that the function f of (11) is given for v_1 real and $v_2>0$ by

$$f(v_1,v_2) = \frac{1}{\sqrt{2\pi/n}}\exp\left(\frac{v_1^2}{2n}\right)\frac{(n-1)^{(n-1)/2}v_2^{n-2}\exp[-(n-1)v_2^2/2]}{2^{(n-3)/2}\Gamma((n-1)/2)}. \tag{34}$$

A simple calculation shows that the F and Q of Theorem 2 and 3 are the distribution functions of $n^{-1/2}t(n-1)$ and $[(n-1)^{1/2}/n]t(n)$ random variables respectively. It follows immediately that when $c/c_1+c/c_2<1$ the best invariant interval estimator is

$$d^* = \left(m_n - [(n-1)^{1/2}/n]t(n;1-c/c_1)s_n,\ m_n - [(n-1)^{1/2}/n]t(n;c/c_2)s_n\right), \tag{35}$$

where t(n;p) denotes the p quantile of the Student t(n) distribution.

The equivalent confidence coefficient is easily calculated from (24) as

$$P\{d^* : d_1 < \mu < d_2; \mu, \sigma\} = F\left[Q^{-1}(1-c/c_1)\right] - F\left[Q^{-1}(c/c_2)\right]$$

$$= T_{n-1}\left([(n-1)/n]^{1/2}t(n;1-c/c_1)\right) - T_{n-1}\left([(n-1)/n]^{1/2}t(n;c/c_2)\right), \tag{36}$$

where T_{n-1} is the distribution function of t(n-1). If, for instance, c=5, $c_1=c_2=200$, and n=21, then $P\{d^*:d_1<\mu<d_2;\mu,\sigma\}=0.944$.

When $c/c_1+c/c_2\geq 1$ the appropriate point estimator of μ is

$$m_n - [(n-1)^{1/2}/n]\, t[n; c_1/(c_1+c_2)]s_n. \tag{37}$$

When overshooting and undershooting are equally punishable ($c_1=c_2$) the point estimator becomes m_n.

4 Interval Estimation of a Scale Parameter

4.1 Problem Statement

We are now interested in scale-invariant linear-loss functions. An argument similar to that of Section 3 leads to intervals scale-invariant under the loss function

$$r(\sigma,d) = \begin{cases} \dfrac{c_1(d_1-\sigma)}{\sigma} + \dfrac{c(d_2-d_1)}{\sigma} & (\sigma < d_1), \\ \dfrac{c(d_2-d_1)}{\sigma} & (d_1 \leq \sigma \leq d_2), \\ \dfrac{c(d_2-d_1)}{\sigma} + \dfrac{c_2(\sigma-d_2)}{\sigma} & (\sigma > d_2). \end{cases} \tag{38}$$

The decision problem specified by the informative experiment density function (11) and the loss function (38) is invariant under the group G of transformations. Thus, the problem is to find the best invariant interval estimator of σ,

$$d^* = \arg\min_{d\in\mathcal{D}} R(\sigma,d), \tag{39}$$

where $\mathcal{D}$ is a set of invariant interval estimators of σ, $R(\sigma,d)=E_\sigma\{r(\sigma,d)\}$ is a risk function.

4.2 Best Invariant Estimator

It follows from Corollary 1.1 that an invariant loss function, r(σ,d), can be transformed as follows:

$$r(\sigma,d)=\ddot{r}(v_2,\eta), \tag{40}$$

where

$$\ddot{r}(v_2,\eta)=\begin{cases} c_1(\eta_1 v_2-1)+c(\eta_2-\eta_1)v_2 & (v_2>1/\eta_1), \\ c(\eta_2-\eta_1)v_2 & (1/\eta_1\ge v_2\ge 1/\eta_2), \\ c_2(1-\eta_2 v_2)+c(\eta_2-\eta_1)v_2 & (v_2<1/\eta_2). \end{cases} \tag{41}$$

$v_2=s_n/\sigma$; $\eta=(\eta_1,\eta_2)$, $\eta_1=d_1/s_n$, $\eta_2=d_2/s_n$.

It follows from (41) and Corollary 1.3 that the risk associated with d and σ can be expressed as

$$R(\sigma,d)=E_\sigma\{r(\sigma,d)\}=E_\eta\{\ddot{r}(v_2,\eta)\}=c_1\int_{1/\eta_1}^{\infty}(\eta_1 v_2-1)f(v_2)dv_2$$

$$+c_2\int_0^{1/\eta_2}(1-\eta_2 v_2)f(v_2)dv_2+c(\eta_2-\eta_1)\int_0^{\infty}v_2 f(v_2)dv_2, \tag{42}$$

which is constant on orbits when an invariant estimator (decision rule) d is used, where

$$f(v_2)=\int_{-\infty}^{\infty}f(v_1,v_2)dv_1 \tag{43}$$

is the marginal distribution associated with (11). The fact that the risk (42) is independent of σ means that a decision function $\eta=(\eta_1,\eta_2)$ which minimizes (42) is uniformly best invariant.

The results corresponding to Theorem 2 can be expressed in terms of the marginal distribution (43).

Theorem 4 (*Best Invariant Estimator of σ*). Let Q be the distribution function corresponding to the density function

$$u_2 f(u_2)\left[\int_0^{\infty}u_2 f(u_2)du_2\right]^{-1} \quad (u_2>0), \tag{44}$$

where f is defined by (43).

(i) If $c/c_1+c/c_2<1$ then the uniformly best invariant linear-loss interval estimator of σ is $d^*=(\eta_1 s_n,\eta_2 s_n)$, where

$$Q(1/\eta_1)=1-c/c_1, \quad Q(1/\eta_2)=c/c_2. \tag{45}$$

(ii) If $c/c_1+c/c_2\geq 1$ then the uniformly best invariant linear-loss interval estimator degenerates into a point estimator $\eta_\bullet s_n$, where

$$Q(1/\eta_\bullet) = c_1/(c_1+c_2). \tag{46}$$

Proof. The proof is similar to that of Theorem 2. □

Corollary 4.1 (*Minimum Risk of the Best Invariant Estimator of σ*). The minimum risk is given by

$$\begin{aligned} R(\sigma,d^*) &= E_\theta\{r(\sigma,d^*)\} = E_\eta\{\ddot{r}(v_2,\eta)\} \\ &= -c_1\int_{1/\eta_1}^{\infty} f(v_2)dv_2 + c_2\int_0^{1/\eta_2} f(v_2)dv_2 \end{aligned} \tag{47}$$

for case (i) with $\eta=(\eta_1,\eta_2)$ as given by (45) and for case (ii) with $\eta_1=\eta_2=\eta_\bullet$ as given by (46).

Proof. These results are immediate from (42) when use is made of $\partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_1 = \partial E_\eta\{\ddot{r}(v,\eta)\}/\partial\eta_2 = 0$ in case (i) and $\partial E_\eta\{\ddot{r}(v,(\eta_\bullet,\eta_\bullet))\}/\partial\eta_\bullet=0$ in case (ii). □

4.3 Equivalent Confidence Coefficient

The evaluation of the equivalent confidence coefficient for the case $c/c_1+c/c_2<1$ is the content of Theorem 5.

Theorem 5 (*Equivalent Confidence Coefficient*). Let F be the distribution function corresponding to the density function $f(v_2)$ ($v_2>0$). The confidence coefficient associated with the optimum interval $d^*=(d_1,d_2)$, where $d_1=\eta_1 s_n$, $d_2=\eta_2 s_n$, is

$$P\{d^*: d_1<\sigma<d_2;\sigma\} = F\left[Q^{-1}(1-c/c_1)\right] - F\left[Q^{-1}(c/c_2)\right] \tag{48}$$

Proof. The confidence coefficient corresponding to σ is given by

$$\begin{aligned} P\{s_n: \eta_1 s_n<\sigma<\eta_2 s_n;\sigma\} &= \int_{\sigma/\eta_2}^{\sigma/\eta_1} ds_n \int_{-\infty}^{\infty} p(m_n,s_n;\mu,\sigma)dm_n \\ &= P\{v_2: 1/\eta_2<v_2<1/\eta_1\} \\ &= F(1/\eta_1)-F(1/\eta_2) = F\left[Q^{-1}(1-c/c_1)\right] - F\left[Q^{-1}(c/c_2)\right] \end{aligned} \tag{49}$$

This is independent of σ. □

4.4 Examples

Example 3 (*Two-Parameter Exponential Distribution*). The relevant formulae for the exponential case of Example 1 can be expressed in terms of incomplete gamma functions I(a,b) and their inverses (Pearson, 1922) since

$$f(v_2) = n^{n-1} v_2^{n-2} e^{-nv_2} / \Gamma(n-1) \quad (v_2 > 0). \tag{50}$$

When $c/c_1 + c/c_2 < 1$ the best interval estimator of σ is

$$d^* = (d_1, d_2) = (\eta_1 s_n, \eta_2 s_n), \tag{51}$$

where

$$I\left(\frac{\sqrt{n}}{\eta_1}, n-1\right) = 1 - c/c_1, \quad I\left(\frac{\sqrt{n}}{\eta_2}, n-1\right) = c/c_2. \tag{52}$$

The equivalent confidence coefficient is given by

$$P\{d^* : d_1 < \sigma < d_2; \sigma\} = I\left(\frac{n}{\eta_1 \sqrt{n-1}}, n-2\right) - I\left(\frac{n}{\eta_2 \sqrt{n-1}}, n-2\right). \tag{53}$$

If, say, $c=5$, $c_1=c_2=200$, and $n=5$, then $P\{d^*: d_1<\sigma<d_2; \sigma\}=0.910$.

Example 4 (*Normal Distribution*). For the normal situation described in Example 2 the density function $f(v_2)$ is given by

$$f(v_2) = \frac{(n-1)^{(n-1)/2} v_2^{n-2} \exp[-(n-1)v_2^2/2]}{2^{(n-3)/2} \Gamma((n-1)/2)} \quad (v_2 > 0), \tag{54}$$

so that F and Q of Theorem 4 and 5 are the distribution functions of $[\chi^2(n-1)/(n-1)]^{1/2}$ and $[\chi^2(n)/(n-1)]^{1/2}$ random variables. We thus have the best interval estimator of σ, when $c/c_1 + c/c_2 < 1$,

$$d^* = (d_1, d_2), \tag{55}$$

where

$$d_1 = \frac{s_n \sqrt{n-1}}{\sqrt{\chi^2(n; 1-c/c_1)}}, \quad d_2 = \frac{s_n \sqrt{n-1}}{\sqrt{\chi^2(n; c/c_2)}}, \tag{56}$$

$\chi^2(n;p)$ denotes the p quantile of the $\chi^2(n)$ distribution.

The equivalent confidence coefficient is given by

$$P\{d^* : d_1 < \sigma < d_2; \sigma\} == T_{n-1}\left[\chi^2(n; 1 - c/c_1)\right] - T_{n-1}\left[\chi^2(n; c/c_2)\right] \qquad (57)$$

If, for instance, c=5, c_1=c_2=200, and n=21, then $P\{d^*{:}d_1{<}\sigma{<}d_2;\sigma\}=0.945$.

5 Conclusions

In this paper we construct the best invariant interval estimators. The method used is that of the invariant embedding of sample statistics in a loss function in order to form pivotal quantities which make it possible to eliminate unknown parameters from the problem. This method is a special case of more general considerations applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space. It is easy to see that the method given in this paper is simpler compared to any other method available so far, for solving the above problem.

Acknowledgments

This work was supported in part by Grant GR-96.0213 from the Latvian Council of Sciences and the National Institute of Mathematics and Informatics of Latvia. The authors would like to express their appreciation to the referee and the editor for their helpful comments.

References

Lehmann, H. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

Nechval, N.A. (1982). *Modern Statistical Methods of Operations Research*. RCAEI, Riga.

Nechval, N.A. (1984). *Theory and Methods of Adaptive Control of Stochastic Processes*. RCAEI, Riga.

Nechval, N.A. and Nechval, K.N. (2000). State estimation of stochastic systems via invariant embedding technique. In: *Cybernetics and Systems 2000* (edited by R. Trappl). Vienna: Austrian Society for Cybernetic Studies, Vol. 1, pp. 96-101.

Pearson, K. (1922). *Tables of Incomplete Γ-Function*. Cambridge University Press.