Improved Estimation of State of Stochastic Systems via Invariant Embedding Technique

NICHOLAS A. NECHVAL, GUNDARS BERZINS, MARIS PURGAILIS
Department of Mathematical Statistics
University of Latvia
Raina Blvd 19, LV-1050, Riga
LATVIA
nechval@junik.lv

KONSTANTIN N. NECHVAL
Department of Applied Mathematics
Transport and Telecommunication Institute
Lomonosov Street 1, LV-1019, Riga
LATVIA
konstan@tsi.lv

Abstract: - In the present paper, for constructing the minimum risk estimators of state of stochastic systems, 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. It allows one to eliminate unknown parameters from the problem and to find the best invariant estimator, which has smaller risk than any of the well-known estimators. Also the problem of how to select the total number of the observations optimally when a constant cost is incurred for each observation taken is discussed. To illustrate the proposed technique, examples are given.

Key-Words: - Stochastic system; State; Estimation; Invariant embedding technique

1 Introduction

The state estimation of discrete-time systems in the presence of random disturbances and measurement noise is an important field in modern control theory. A significant research effort has been devoted to the problem of state estimation for stochastic systems. Since Kalman's noteworthy paper [1], the problem of state estimation in linear and nonlinear systems has been treated extensively and various aspects of the problem have been analyzed [2-8].

The problem of determining an optimal estimator of the state of stochastic system in the absence of complete information about the distributions of random disturbances and measurement noise is seen to be a standard problem of statistical estimation. Unfortunately, the classical theory of statistical estimation has little to offer in general type of situation of loss function. The bulk of the classical theory has been developed about the assumption of a quadratic, or at least symmetric and analytically simple loss structure. In some cases this assumption is made explicit, although in most it is implicit in the search for estimating procedures that have the "nice" statistical properties of unbiasedness and minimum variance. Such procedures are usually satisfactory if the estimators so generated are to be used solely for the purpose of reporting information to another party for an unknown purpose, when the loss structure is not easily discernible, or when the number of observations is large enough to support Normal approximations and asymptotic results. Unfortunately, we seldom are fortunate enough to be in asymptotic situations. Small sample sizes are generally the rule when estimation of system states and the small sample properties of estimators do not appear to have been thoroughly investigated. Therefore, the above procedures of the state estimation have long been recognized as deficient, however, when the purpose of estimation is the making of a specific decision (or sequence of decisions) on the basis of a limited amount of information in a situation where the losses are clearly asymmetric – as they are here.

There exists a class of control systems where observations are not available at every time due to either physical impossibility and/or the costs involved in taking a measurement. For such systems it is realistic to derive the optimal policy of state estimation with some constraints imposed on the observation scheme.

It is assumed in this paper that there is a constant cost associated with each observation taken. The optimal estimation policy is obtained for a discrete-time deterministic plant observed through noise. It is shown that there is an optimal number of observations to be taken.

The outline of the paper is as follows. A formulation of the problem is given in Section 2. Section 3 is devoted to characterization of estimators. A comparison of estimators is discussed in Section 4. An invariant embedding technique is described in Section 5. A general problem analysis is presented in Section 6. An example is given in Section 7.

2 Problem Statement

To make the above introduction more precise, consider the discrete-time system, which in particular is described by vector difference equations of the following form:

$$\mathbf{x}(k+1) = \mathbf{A}(k+1,k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k), \tag{1}$$

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{w}(k), \quad k = 1,2,3,\ldots, \tag{2}$$

where $\mathbf{x}(k+1)$ is an n vector representing the state of the system at the $(k+1)$th time instant with initial condition $\mathbf{x}(1)$; $\mathbf{z}(k)$ is an m vector (the observed signal) which can be termed a measurement of the system at the kth instant; $\mathbf{H}(k)$ is an $m \times n$ matrix; $\mathbf{A}(k+1,k)$ is a transition matrix of dimension $n \times n$, and $\mathbf{B}(k)$ is an $n \times p$ matrix, $\mathbf{u}(k)$ is a p vector, the control vector of the system; $\mathbf{w}(k)$ is a random vector of dimension m (the measurement noise). By repeated use of (1) we find

$$\mathbf{x}(k) = \mathbf{A}(k,j)\mathbf{x}(j) + \sum_{i=j}^{k-1} \mathbf{A}(k,i+1)\mathbf{B}(i)\mathbf{u}(i), \quad j \le k, \tag{3}$$

where the discrete-time system transition matrix satisfies the matrix difference equation,

$$\mathbf{A}(k+1,j) = \mathbf{A}(k+1,k)\mathbf{A}(k,j), \quad \forall k,j, \tag{4}$$

$$\mathbf{A}(k,k) = \mathbf{I}, \quad \mathbf{A}(k,j) = \prod_{i=j}^{k-1} \mathbf{A}(i+1,i). \tag{5}$$

From these properties, it immediately follows that

$$\mathbf{A}^{-1}(k,j) = \mathbf{A}(j,k), \quad \forall k,j, \tag{6}$$

$$\mathbf{A}(\alpha,\beta)\mathbf{A}(\beta,\gamma) = \mathbf{A}(\alpha,\gamma), \quad \forall \alpha,\beta,\gamma. \tag{7}$$

Thus, for $j \le k$,

$$\mathbf{x}(j) = \mathbf{A}(j,k)\mathbf{x}(k) - \sum_{i=j}^{k-1} \mathbf{A}(j,i+1)\mathbf{B}(i)\mathbf{u}(i). \tag{8}$$

The problem to be considered is the estimation of the state of the above discrete-time system. This problem may be stated as follows. Given the observed sequence, $\mathbf{z}(1)$, …, $\mathbf{z}(k)$, it is required to obtain an estimator $\mathbf{d}$ of $\mathbf{x}(l)$ based on all available observed data $\mathbf{Z}^k=\{\mathbf{z}(1), \ldots, \mathbf{z}(k)\}$ such that the expected losses (risk function)

$$R(\theta,\mathbf{d}) = \mathrm{E}_\theta\{r(\theta,\mathbf{d})\} \tag{9}$$

is minimized, where $r(\theta,\mathbf{d})$ is a specified loss function at decision point $\mathbf{d}\equiv\mathbf{d}(\mathbf{Z}^k)$, $\theta=(\mathbf{x}(l),\omega)$, ω is an unknown parametric vector of the probability distribution of $\mathbf{w}(k)$, $k \le l$.

If it is assumed that a constant cost $c > 0$ is associated with each observation taken, the criterion function for the case of k observations is taken to be

$$r_k(\theta,\mathbf{d}) = r(\theta,\mathbf{d}) + ck. \tag{10}$$

In this case, the optimization problem is to find

$$\min_k \min_{\mathbf{d}} \mathrm{E}_\theta\{r_k(\theta,\mathbf{d})\}, \tag{11}$$

where the inner minimization operation is with respect to $\mathbf{d}\equiv\mathbf{d}(\mathbf{Z}^k)$, when the k observations have been taken, and where the outer minimization operation is with respect to k.

3 Characterization of Estimators

For any statistical decision problem, an estimator (a decision rule) $\mathbf{d}_1$ is said to be equivalent to an estimator (a decision rule) $\mathbf{d}_2$ if $R(\theta,\mathbf{d}_1)=R(\theta,\mathbf{d}_2)$ for all $\theta\in\Theta$, where $R(.)$ is a risk function, Θ is a parameter space,. An estimator $\mathbf{d}_1$ is said to be uniformly better than an estimator $\mathbf{d}_2$ if $R(\theta,\mathbf{d}_1) < R(\theta,\mathbf{d}_2)$ for all $\theta\in\Theta$. An estimator $\mathbf{d}_1$ is said to be as good as an estimator $\mathbf{d}_2$ if $R(\theta,\mathbf{d}_1) \le R(\theta,\mathbf{d}_2)$ for all $\theta\in\Theta$. However, it is also possible that we may have

"$\mathbf{d}_1$ and $\mathbf{d}_2$ are incomparable", that is, $R(\theta,\mathbf{d}_1) < R(\theta,\mathbf{d}_2)$ for at least one $\theta\in\Theta$, and $R(\theta,\mathbf{d}_1) > R(\theta,\mathbf{d}_2)$ for at least one $\theta\in\Theta$. Therefore, this ordering gives a partial ordering of the set of estimators.

An estimator **d** is said to be uniformly non-dominated if there is no estimator uniformly better than **d**. The conditions that an estimator must satisfy in order that it might be uniformly non-dominated are given by the following theorem.

Theorem 1 (*Uniformly non-dominated estimator*). Let (ξ_τ; τ=1,2, ...) be a sequence of the prior distributions on the parameter space Θ. Suppose that ($\mathbf{d}_\tau$;τ=1,2, ...) and ($Q(\xi_\tau,\mathbf{d}_\tau)$; τ=1,2, ...) are the sequences of Bayes estimators and prior risks, respectively. If there exists an estimator $\mathbf{d}^*$ such that its risk function $R(\theta,\mathbf{d}^*)$, $\theta\in\Theta$, satisfies the relationship

$$\lim_{\tau\to\infty}\left[Q(\xi_\tau,\mathbf{d}^*)-Q(\xi_\tau,\mathbf{d}_\tau)\right]=0, \tag{12}$$

where

$$Q(\xi_\tau,\mathbf{d})=\int_\Theta R(\theta,\mathbf{d})\xi_\tau(d\theta), \tag{13}$$

then $\mathbf{d}^*$ is an uniformly non-dominated estimator.

Proof. Suppose $\mathbf{d}^*$ is uniformly dominated. Then there exists an estimator $\mathbf{d}^{**}$ such that $R(\theta,\mathbf{d}^{**}) < R(\theta,\mathbf{d}^*)$ for all $\theta\in\Theta$. Let

$$\varepsilon=\inf_{\theta\in\Theta}\left[R(\theta,\mathbf{d}^*)-R(\theta,\mathbf{d}^{**})\right]>0. \tag{14}$$

Then

$$Q(\xi_\tau,\mathbf{d}^*)-Q(\xi_\tau,\mathbf{d}^{**})\ge\varepsilon. \tag{15}$$

Simultaneously,

$$Q(\xi_\tau,\mathbf{d}^{**})-Q(\xi_\tau,\mathbf{d}_\tau)\ge 0, \tag{16}$$

τ=1,2, ..., and

$$\lim_{\tau\to\infty}\left[Q(\xi_\tau,\mathbf{d}^{**})-Q(\xi_\tau,\mathbf{d}_\tau)\right]\ge 0. \tag{17}$$

On the other hand,

$$Q(\xi_\tau,\mathbf{d}^{**})-Q(\xi_\tau,\mathbf{d}_\tau)$$

$$=\left[Q(\xi_\tau,\mathbf{d}^*)-Q(\xi_\tau,\mathbf{d}_\tau)\right]-\left[Q(\xi_\tau,\mathbf{d}^*)-Q(\xi_\tau,\mathbf{d}^{**})\right]$$

$$\le\left[Q(\xi_\tau,\mathbf{d}^*)-Q(\xi_\tau,\mathbf{d}_\tau)\right]-\varepsilon \tag{18}$$

and

$$\lim_{\tau\to\infty}\left[Q(\xi_\tau,\mathbf{d}^{**})-Q(\xi_\tau,\mathbf{d}_\tau)\right]<0. \tag{19}$$

This contradiction proves that $\mathbf{d}^*$ is an uniformly non-dominated estimator. □

4 Comparison of Estimators

In order to judge which estimator might be preferred for a given situation, a comparison based on some "closeness to the true value" criteria should be made. The following approach is commonly used [9-10]. Consider two estimators, say, $\mathbf{d}_1$ and $\mathbf{d}_2$ having risk function $R(\theta,\mathbf{d}_1)$ and $R(\theta,\mathbf{d}_2)$, respectively. Then the relative efficiency of $\mathbf{d}_1$ relative to $\mathbf{d}_2$ is given by

$$\text{rel.eff.}_R\{\mathbf{d}_1,\mathbf{d}_2;\theta\}=R(\theta,\mathbf{d}_2)/R(\theta,\mathbf{d}_1). \tag{20}$$

When $\text{rel.eff.}_R\{\mathbf{d}_1,\mathbf{d}_2;\theta_0\}<1$ for some θ_0, we say that $\mathbf{d}_2$ is more efficient than $\mathbf{d}_1$ at θ_0. If $\text{rel.eff.}_R\{\mathbf{d}_1,\mathbf{d}_2;\theta\}\le 1$ for all θ with a strict inequality for some θ_0, then $\mathbf{d}_1$ is inadmissible relative to $\mathbf{d}_2$.

5 Invariant Embedding Technique

This paper is concerned with the implications of group theoretic structure for invariant performance indexes. We present an invariant embedding technique 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 present paper is to show how the invariance principle may be employed in the particular case of finding the improved statistical decisions. 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.

5.1 Preliminaries

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

$$r(\overline{g}\theta, \widetilde{g}\mathbf{u}) = r(\theta, \mathbf{u}), g \in G. \qquad (21)$$

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

$$\varphi(g\mathbf{x}) = \widetilde{g}\varphi(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}, \quad g \in G. \qquad (22)$$

If $\overline{G}$ is trivial and (21), (22) hold, we say φ is G-invariant, or simply invariant [11-12].

5.2 Invariant Functions

We begin by noting that r is invariant in the sense of (21) if and only if r is a $G^{\bullet}$-invariant function, where $G^{\bullet}$ is defined on $\Theta \times \mathcal{U}$ as follows: to each $g \in G$, with homomorphic images $\overline{g}, \widetilde{g}$ in $\overline{G}, \widetilde{G}$ respectively, let $g^{\bullet}(\theta,\mathbf{u}) = (\overline{g}\theta, \widetilde{g}\mathbf{u})$, $(\theta,\mathbf{u}) \in (\Theta \times \mathcal{U})$. 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 $\theta, \vartheta \in \Theta$ 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 2 (*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 $\Theta \times \mathcal{U}$ is defined by

$$M(\theta, \mathbf{u}) = \widetilde{g}_{\theta}^{-1}\mathbf{u}, \quad (\theta, \mathbf{u}) \in \Theta \times \mathcal{U}. \qquad (23)$$

Proof. For each $(\theta,\mathbf{u}) \in (\Theta \times \mathcal{U})$ and $\overline{g} \in \overline{G}$

$$M(\overline{g}\theta, \widetilde{g}\mathbf{u}) = (\widetilde{g}_{\overline{g}\theta}^{-1})\widetilde{g}\mathbf{u} = (\widetilde{g}\widetilde{g}_{\theta})^{-1}\widetilde{g}\mathbf{u}$$

$$= \widetilde{g}_{\theta}^{-1}\widetilde{g}^{-1}\widetilde{g}\mathbf{u} = \widetilde{g}_{\theta}^{-1}\mathbf{u} = M(\theta, \mathbf{u}) \qquad (24)$$

by Lemma 1 and the structure preserving properties of homomorphisms. Thus M is $G^{\bullet}$-invariant. To see that M is maximal, let $M(\theta_1,\mathbf{u}_1) = M(\theta_2,\mathbf{u}_2)$. Then $\widetilde{g}_{\theta_1}^{-1}\mathbf{u}_1 = \widetilde{g}_{\theta_2}^{-1}\mathbf{u}_2$ or $\mathbf{u}_1 = \widetilde{g}\,\mathbf{u}_2$, where $\widetilde{g} = \widetilde{g}_{\theta_1}\widetilde{g}_{\theta_2}^{-1}$. Since $\theta_1 = \overline{g}_{\theta_1}\theta_0 = \overline{g}_{\theta_1}\overline{g}_{\theta_2}^{-1}\theta_2 = \overline{g}\theta_2$, $(\theta_1,\mathbf{u}_1) = g^{\bullet}(\theta_2,\mathbf{u}_2)$ for some $g^{\bullet} \in G^{\bullet}$, and the proof is complete. □

Corollary 2.1 (*Invariant embedding*). An invariant function, $r(\theta,\mathbf{u})$, can be transformed as follows:

$$r(\theta, \mathbf{u}) = r(\overline{g}_{\hat{\theta}}^{-1}\theta, \widetilde{g}_{\hat{\theta}}^{-1}\mathbf{u}) = \ddot{r}(\mathbf{v}, \eta), \qquad (25)$$

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

Corollary 2.2 (*Best invariant decision rule*). If $r(\theta,\mathbf{u})$ is an invariant loss function, the best invariant decision rule is given by

$$\varphi^*(\mathbf{x}) = \mathbf{u}^* = \eta^{-1}(\eta^*, \hat{\theta}), \qquad (26)$$

where

$$\eta^* = \arg\inf_{\eta} \mathrm{E}_{\eta}\{\ddot{r}(\mathbf{v}, \eta)\}. \qquad (27)$$

Corollary 2.3 (*Risk*). A risk function (performance index)

$$R(\theta, \varphi(\mathbf{x})) = \mathrm{E}_{\theta}\{r(\theta, \varphi(\mathbf{x}))\} = \mathrm{E}_{\eta_\circ}\{\ddot{r}(\mathbf{v}_\circ, \eta_\circ)\} \quad (28)$$

is constant on orbits when an invariant decision rule $\varphi(\mathbf{x})$ is used, where $\mathbf{v}_\circ = \mathbf{v}_\circ(\theta, \mathbf{x})$ is a function whose distribution does not depend on θ; $\eta_\circ = \eta_\circ(\mathbf{u}, \mathbf{x})$ is an ancillary factor.

For instance, consider the problem of estimating the location-scale parameter of a distribution belonging to a family generated by a continuous cdf F: $\mathcal{P}=\{P_\theta: F((x-\mu)/\sigma), x\in R, \theta\in\Theta\}$, $\Theta=\{(\mu,\sigma): \mu,\sigma \in R, \sigma > 0\} = \mathcal{U}$. 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

$$r(\theta, \varphi(x)) = r[(\varphi_1(x) - \mu)/\sigma, \varphi_2(x)/\sigma], \quad (29)$$

where

$$\theta=(\mu,\sigma) \text{ and } \varphi(x)=(\varphi_1(x),\varphi_2(x)). \quad (30)$$

Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ and $\mathbf{u}=(u_1,u_2)$, then

$$r(\theta, \mathbf{u}) = \ddot{r}(\mathbf{v}, \eta) = \ddot{r}(v_1 + \eta_1 v_2, \eta_2 v_2), \quad (31)$$

where

$$\mathbf{v}=(v_1,v_2),\ v_1=(\hat{\mu} - \mu)/\sigma,\ v_2=\hat{\sigma}/\sigma; \quad (32)$$

$$\eta=(\eta_1,\eta_2),\ \eta_1=(u_1 - \hat{\mu})/\hat{\sigma},\ \eta_2=u_2/\hat{\sigma}. \quad (33)$$

5.3 Illustrative Example 1

Consider an inventory manager faced with a one-period Christmas-tree stocking problem. Assume the decision maker has demand data on the sale of trees over the last n seasons. For the sake of simplicity, we shall consider the case where the demand data can be measured on a continuous scale. We restrict attention to the case where these demand values constitute independent observations from a distribution belonging to invariant family. In particular, we consider a distribution belonging to location-scale family generated by a continuous cdf F: $\mathcal{P}=\{P_\theta: F((x-\mu)/\sigma), x\in R, \theta\in\Theta\}$, $\Theta=\{(\mu,\sigma): \mu,\sigma\in R, \sigma>0\}$, which is indexed by the vector parameter $\theta=(\mu,\sigma)$, where μ and σ (>0) are respectively parameters of location and scale. The group G of location and scale changes leaves the class of models invariant. The purpose in restricting attention to such families of distributions is that for such families the decision problem is invariant, and if the estimators of safety stock levels are equivariant (i.e. the group of location and scale changes leaves the decision problem invariant), then any comparison of estimation procedures is independent of the true values of any unknown parameters. The common distributions used in inventory problems are the normal, exponential, Weibull, and gamma distributions.

Let us assume that, for one reason or another, a $100\gamma\%$ service level is desired (i.e. the decision maker wants to ensure that at least $100\gamma\%$ of his customers are satisfied). If the demand distribution is completely specified, the appropriate amount of inventory to stock for the season is u satisfying

$$\Pr\{X \le u\} = F\left(\frac{u-\mu}{\sigma}\right) = \gamma \quad (34)$$

or

$$u = \mu + p_\gamma \sigma, \quad (35)$$

where

$$p_\gamma = F^{-1}(\gamma) \quad (36)$$

is the γth percentile of the above distribution. Since the inventory manager does not know μ or σ, the estimator commonly used to estimate u is the maximum likelihood estimator

$$\hat{u} = \hat{\mu} + p_\gamma \hat{\sigma}, \quad (37)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimators of the parameters μ and σ, respectively. This estimator is one possible estimator of u and it may yield poor results.

The correct procedure for estimating u requires establishing a tolerance limit for the percentile. It should be noted that tolerance limits are to percentiles what confidence limits are to parameters. With confidence limits, inferences may be drawn on parameters, whereas with tolerance limits, inferences may be drawn about proportions of a distribution.

There are two criteria for establishing tolerance limits. The first criterion establishes an interval such that the expected percentage of observations falling into the interval just exceeds $100\gamma\%$ [13]. This interval is called the $100\gamma\%$ expectation interval. The second criterion establishes an interval, which ensures that $100\gamma\%$ of the population is covered with confidence $1-\alpha$ [14]. Such an interval is called a $100\gamma\%$ content tolerance interval at level $1-\alpha$. The decision as to which interval to construct depends on the nature of the problem. A precision-instrument manufacturer wanting to construct an interval which, with high confidence, contains 90% of the distribution of diameters, for example, would use a 90% content tolerance interval, whereas an inventory manager wanting to stock sufficient items to ensure that in the long run an average of 95% of demand will be satisfied may find expectation intervals more appropriate. Expectation intervals are only appropriate in inventory problems where average service levels are to be controlled.

Tolerance limits of the types mentioned above are considered in this subsection.

That is, if $f(x;\theta)$ denotes the density function of the parent population under consideration and if S is any statistic obtained from a random sample of that population, then $\widehat{u}^\circ \equiv \widehat{u}^\circ(S)$ is a lower 100(1-γ)% expectation limit if

$$\Pr\left\{X > \widehat{u}^\circ\right\} = \mathrm{E}_\theta\left\{\int_{\widehat{u}^\circ}^{\infty} f(x;\theta)dx\right\}$$

$$= \mathrm{E}_\theta\left\{1 - F\left(\frac{\widehat{u}^\circ - \mu}{\sigma}\right)\right\} = 1-\gamma. \quad (38)$$

This expression represents a risk of $\widehat{u}^\circ$, i.e.

$$R^\circ(\theta, \widehat{u}^\circ) = \Pr\left\{X > \widehat{u}^\circ\right\} = 1-\gamma. \quad (39)$$

A lower 100(1-γ)% content tolerance limit at level 1-α, $\widehat{u}^\bullet \equiv \widehat{u}^\bullet(S)$, is defined by

$$\Pr\left\{\int_{\widehat{u}^\bullet}^{\infty} f(x;\theta)dx \le 1-\gamma\right\} = \Pr\left\{F\left(\frac{\widehat{u}^\bullet - \mu}{\sigma}\right) \ge \gamma\right\}$$

$$= \Pr\left\{\widehat{u}^\bullet \ge \mu + p_\gamma \sigma\right\} = 1-\alpha. \quad (40)$$

A risk of this limit is

$$R^\bullet(\theta, \widehat{u}^\bullet) = 1 - \Pr\left\{\widehat{u}^\bullet \ge \mu + p_\gamma \sigma\right\} = \alpha. \quad (41)$$

Since it is often desirable to have statistical tolerance limits available for the distributions used to describe demand data in inventory control, the problem is to find these limits. We give below a general procedure for obtaining tolerance limits. This procedure is based on the use of an invariant embedding technique given above.

Lower 100(1-γ)% expectation limit. Suppose X_1, ..., X_n are a random sample from the exponential distribution, with pdf

$$f(x;\sigma) = \frac{1}{\sigma}\exp(-x/\sigma), \quad x \ge 0, \quad (42)$$

where $\sigma > 0$ is unknown parameter. Let

$$S_n = \sum_{i=1}^{n} X_i. \quad (43)$$

It can be justified by using the factorization theorem that S_n is a sufficient statistic for σ. We wish, on the basis of the sufficient statistic S_n for σ, to construct the lower 100(1-γ)% expectation limit for a stock level. It follows from (38) that this limit is defined by

$$\Pr\left\{X > \widehat{u}^\circ\right\} = \mathrm{E}_\sigma\left\{\int_{\widehat{u}^\circ}^{\infty} f(x;\sigma)dx\right\}$$

$$= \mathrm{E}_\sigma\left\{\exp(-\widehat{u}^\circ/\sigma)\right\} = 1-\gamma. \quad (44)$$

where $\widehat{u}^\circ \equiv \widehat{u}^\circ(S_n)$.

Using the technique of invariant embedding of S_n in a maximal invariant

$$M = \widehat{u}^\circ/\sigma, \quad (45)$$

we reduce (44) to

$$\mathrm{E}_\sigma\left\{\exp(-\widehat{u}^\circ/\sigma)\right\} = \mathrm{E}\left\{\exp(-\eta^\circ V); \eta^\circ\right\} = 1-\gamma. \quad (46)$$

where

$$V = S_n/\sigma \quad (47)$$

is the pivotal quantity whose distribution does not depend on unknown parameter σ,

$$\eta^\circ = \widehat{u}^\circ/S_n. \quad (48)$$

is an ancillary factor. It is well known that the probability density function of V is given by

$$h(v) = \frac{1}{\Gamma(n)} v^{n-1}\exp(-v), \quad v \ge 0. \quad (49)$$

Thus, for this example, $\widehat{u}^\circ$ can be found explicitly as

$$\hat{u}^{\circ} = \eta^{\circ} S_n, \tag{50}$$

where (see (46))

$$\eta^{\circ} = \left(\frac{1}{1-\gamma}\right)^n - 1. \tag{51}$$

If the parameters μ and σ were known, it follows from (44) that

$$u = p_\gamma \sigma, \tag{52}$$

where

$$p_\gamma = \ln\left(\frac{1}{1-\gamma}\right). \tag{53}$$

The maximum likelihood estimator of u is given by

$$\hat{u} = p_\gamma \hat{\sigma}, \tag{54}$$

where

$$\hat{\sigma} = S_n / n \tag{55}$$

is the maximum likelihood estimator of the parameter σ.

One can see that each of the above estimators is a member of the class

$$\mathscr{C} = \left\{ \hat{d} : \hat{d} = k S_n \right\}, \tag{56}$$

where k is a non-negative real number. A risk of an estimator, which belongs to the class $\mathscr{C}$, is given by

$$R^{\circ}(\sigma, \hat{d}) = \left(\frac{1}{k+1}\right)^n. \tag{57}$$

Then the relative efficiency of $\hat{d}$ relative to $\hat{u}^{\circ}$ is given by

$$\text{rel.eff.}_{\mathrm{R}^{\circ}}\left\{\hat{d}, \hat{u}^{\circ}; \sigma\right\} = \mathrm{R}^{\circ}(\sigma, \hat{u}^{\circ}) \big/ \mathrm{R}^{\circ}(\sigma, \hat{d})$$

$$= (1-\gamma)(1+k)^n. \tag{58}$$

If, say,

$$k = p_\gamma / n = n^{-1} \ln\left(\frac{1}{1-\gamma}\right), \tag{59}$$

n=2 and γ=0.95, then the relative efficiency of the maximum likelihood estimator, $\hat{u}$, relative to $\hat{u}^{\circ}$ is given by

$$\text{rel.eff.}_{R^{\circ}}\left\{\hat{u}, \hat{u}^{\circ}; \sigma\right\} = (1-\gamma)\left[1 + n^{-1} \ln\left(\frac{1}{1-\gamma}\right)\right]^n$$

$$=0.312. \tag{60}$$

Lower 100(1-γ)% content tolerance limit at level 1-α. Now we wish, on the basis of a sufficient statistic S_n for σ, to construct the lower 100(1-γ)% content tolerance limit at level 1-α for the size of the stock in order to ensure an adequate service level. It follows from (40) that this tolerance limit is defined by

$$\Pr\left\{\int_{\hat{u}^{\bullet}}^{\infty} f(x;\sigma)dx \le 1-\gamma\right\} = \Pr\left\{F\left(\frac{\hat{u}^{\bullet}}{\sigma}\right) \ge \gamma\right\}$$

$$= \Pr\left\{\hat{u}^{\bullet} \ge p_\gamma \sigma\right\} = 1-\alpha. \tag{61}$$

By using the technique of invariant embedding of S_n in a maximal invariant

$$M = \hat{u}^{\bullet} / \sigma, \tag{62}$$

we reduce (61) to

$$\Pr\left\{\hat{u}^{\bullet} \ge p_\gamma \sigma\right\} = \Pr\left\{V \ge p_\gamma / \eta^{\bullet}\right\} = 1-\alpha. \tag{63}$$

where $\hat{u}^{\bullet} \equiv \hat{u}^{\bullet}(S_n)$,

$$\eta^{\bullet} = \hat{u}^{\bullet} / S_n \tag{64}$$

is an ancillary factor.

It follows from the above that, in this case, $\hat{u}^{\bullet}$ can be found explicitly as

$$\hat{u}^{\bullet} = \eta^{\bullet} S_n, \tag{65}$$

where

$$\eta^{\bullet} = \frac{2p_\gamma}{\chi_\alpha^2(2n)} = \frac{2\ln\left(\frac{1}{1-\gamma}\right)}{\chi_\alpha^2(2n)}, \tag{66}$$

$\chi_\alpha^2(2n)$ is the 100α% point of the chi-square distribution with 2n degrees of freedom.

Since the estimator $\hat{u}^{\bullet}$ belongs to the class $\mathscr{C}$, then the relative efficiency of $\hat{d} \in \mathscr{C}$ relative to $\hat{u}^{\bullet}$ is given by

$$\text{rel.eff.}_{R^\bullet}\{\hat{d},\hat{u}^\bullet;\sigma\} = R^\bullet(\sigma,\hat{u}^\bullet)/R^\bullet(\sigma,\hat{d})$$

$$= \alpha\left[1-\Pr\left\{\chi^2(2n) \ge \frac{2p_\gamma}{k}\right\}\right]^{-1}. \quad (67)$$

If, say, k is given by (59), n=2 and α=0.05, then we have that the relative efficiency of the maximum likelihood estimator, $\hat{u}$, relative to $\hat{u}^\bullet$ is given by

$$\text{rel.eff.}_{R^\bullet}\{\hat{u},\hat{u}^\bullet;\sigma\} = \alpha\left[1-\Pr\{\chi^2(2n) \ge 2n\}\right]^{-1}$$

$$=0.084. \quad (68)$$

5.4 Illustrative Example 2

Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(k)}$ be the k smallest observations in a sample of size n from the two-parameter exponential distribution, with density

$$f(x;\theta) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), \quad x \ge \mu, \quad (69)$$

where σ>0 and μ are unknown parameters, $\theta=(\mu,\sigma)$.

Let $Y_{(r)}$ be the rth smallest observation in a future sample of size m from the same distribution. We wish, on the basis of observed $X_{(1)}, \ldots, X_{(k)}$ to construct prediction intervals for $Y_{(r)}$.

Let

$$S_r=(Y_{(r)}-\mu)/\sigma, \quad S_1=(X_{(1)}-\mu)/\sigma \quad (70)$$

and

$$T_1=T/\sigma, \quad (71)$$

where

$$T = \sum_{i=1}^{k}(X_{(i)} - X_{(1)}) + (n-k)(X_{(k)} - X_{(1)}). \quad (72)$$

To construct prediction intervals for $Y_{(r)}$, consider the quantity (invariant statistic)

$$V = n(S_r-S_1)/T_1= n(Y_{(r)}-X_{(1)})/T. \quad (73)$$

It is well known [15] that nS_1 has a standard exponential distribution, that $2T_1 \sim \chi^2_{2k-2}$ and that S_1 and T_1 are independent. Also, S_r is the rth order statistic from a sample of size m from the standard exponential distribution and thus has probability density function [16],

$$f(s_r) = r\binom{m}{r}(1-e^{-s_r})^{r-1}e^{-s_r(m-r+1)}, \quad (74)$$

if s_r>0, and $f(s_r)$=0 for $s_r \le 0$. Using the technique of invariant embedding, we find after some algebra that

$$F(v)=\Pr\{V \le v\}$$

$$= \begin{cases} 1-nr\binom{m}{r} \\ \times\sum_{j=0}^{r-1}\dfrac{\binom{r-1}{j}(-1)^j[1+v(m-r+j+1)/n]^{-k+1}}{(m+n-r+j+1)(m-r+j+1)}, & v>0, \\ m^{(r)}(1-v)^{-k+1}/(m+n)^{(r)}, & v \le 0, \end{cases}$$

(75)

where $m^{(r)}=m(m-1)\cdots(m-r+1)$.

The special case in which r=1 is worth mentioning, since in this case (75) simplifies somewhat. We find here that we can write

$$F(v)=\Pr\{V \le v\}$$

$$= \begin{cases} 1-\dfrac{\vartheta}{\vartheta+1}\left(\dfrac{\vartheta}{\vartheta+v}\right)^{k-1}, & v>0, \\ (\vartheta+1)^{-1}(1-v)^{-k+1}, & v \le 0, \end{cases} \quad (76)$$

where $\vartheta=n/m$.

Consider the ordered data given by Grubbs [17] on the mileages at which nineteen military carriers failed. These were 162, 200, 271, 302, 393, 508, 539, 629, 706, 777, 884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880, and thus constitute a complete sample with k=n=19. We find

$$T = \sum_{i=1}^{19}(X_{(i)} - X_{(1)}) = 15869 \quad (77)$$

and of course $X_{(1)}$=162.

Suppose we wish to set up the shortest-length ($1-\alpha$=0.95) prediction interval for the smallest observation $Y_{(1)}$ in a future sample of size m=5. Consider the invariant statistic

$$V=\frac{n(Y_{(1)}-X_{(1)})}{T}. \qquad (78)$$

Then

$$\Pr\left\{v_1<\frac{n(Y_{(1)}-X_{(1)})}{T}<v_2\right\}$$

$$=\Pr\left\{X_{(1)}+v_1\frac{T}{n}<Y_{(1)}<X_{(1)}+v_2\frac{T}{n}\right\}$$

$$=\Pr\{z_L<Y_{(1)}<z_U\}=1-\alpha, \qquad (79)$$

where

$$z_L=X_{(1)}+v_1T/n \qquad (80)$$

and

$$z_U=X_{(1)}+v_2T/n. \qquad (81)$$

The length of the prediction interval is

$$\Delta_z=z_U-z_L=(T/n)(v_2-v_1). \qquad (82)$$

We wish to minimize Δ_z subject to

$$F(v_2)-F(v_1)=1-\alpha. \qquad (83)$$

It can be shown that the minimum occurs when

$$f(v_1)=f(v_2), \qquad (84)$$

where v_1 and v_2 satisfy (83). The shortest-length prediction interval is given by

$$C^*_{Y_{(1)}}(X_{(1)},T)=\left(X_{(1)}+v_1^*\frac{T}{n},X_{(1)}+v_2^*\frac{T}{n}\right)$$

$$=(10.78,\ 736.62), \qquad (85)$$

where v_1^*= –0.18105 and v_2^*= 0.688. Thus, the length of this interval is Δ_z^*= 736.62 – 10.78 =725.84.

The equal tails prediction interval at the $1-\alpha$=0.95 confidence level is given by

$$C^\circ_{Y_{(1)}}(X_{(1)},T)=\left(X_{(1)}+v_{\alpha/2}\frac{T}{n},X_{(1)}+v_{1-\alpha/2}\frac{T}{n}\right)$$

$$=(57.6,\ 834.34), \qquad (86)$$

where $F(v_\alpha)=\alpha$, $v_{\alpha/2}$= –0.125 and $v_{1-\alpha/2}$= 0.805. The length of this interval is Δ_z°= 834.34 – 57.6 = 776.74.

The relative efficiency of $C^\circ_{Y_{(1)}}(X_{(1)},T)$ relative to $C^*_{Y_{(1)}}(X_{(1)},T)$, taking into account Δ_z is given by

$$\text{rel.eff.}_{E_\theta\{\Delta_z\}}\left(C^\circ_{Y_{(1)}}(X_{(1)},T),C^*_{Y_{(1)}}(X_{(1)},T)\right)$$

$$=\frac{\Delta_z^*}{\Delta_z^\circ}=\frac{v_2^*-v_1^*}{v_{1-\alpha/2}-v_{\alpha/2}}=0.934. \qquad (87)$$

One may also be interested in predicting the mean

$$\overline{Y}=\sum_{j=1}^{m}Y_j/m \qquad (88)$$

or total lifetime in a future sample. Consider the quantity

$$V=n(\overline{Y}-X_{(1)})/T. \qquad (89)$$

Using the invariant embedding technique, we find after some algebra that

$$F(v)=\Pr\{V\le v\}$$

$$=\begin{cases}1-\displaystyle\sum_{j=0}^{m-1}\binom{k+j-2}{j}\frac{(v/\vartheta)^j[1-(1+\vartheta)^{-m+j}]}{(1+v/\vartheta)^{k+j-1}}, & v>0,\\ (1+\vartheta)^{-m}(1-v)^{-k+1}, & v\le 0.\end{cases} \qquad (90)$$

Probability statements about V lead to prediction intervals for $\overline{Y}$ or

$$\sum_{j=1}^{m}Y_j=m\overline{Y}. \qquad (91)$$

5.5 Illustrative Example 3

Suppose that $X_1, ..., X_n$ and $Y_{1i}, ..., Y_{mi}$ (i=1, ..., k) denote $n+km$ independent and identically distributed random variables from a two-parameter exponential distribution with pdf (69), where σ >0 and μ are unknown parameters.

Let $X_{(1)}$ be the smallest observation in the initial sample of size n and

$$S_n = \sum_{j=1}^{n} \left(X_j - X_{(1)} \right). \tag{92}$$

It can be justified by using the factorization theorem that $(X_{(1)},S_n)$ is a sufficient statistic for (μ,σ). Let $Y_{(1i)}$ be the smallest observation in the ith future sample of size m, $\forall i=1(1)k$. We wish, on the basis of a sufficient statistic $(X_{(1)},S_n)$ for (μ,σ), to construct simultaneous lower one-sided β-content tolerance limits at level γ for $Y_{(1i)}$, i=1, ..., k. It can be shown that this problem is reduced to the problem of constructing a lower one-sided β-content tolerance limit at level γ, $L \equiv L(X_{(1)},S_n)$, for

$$Y_{(1)} = \min_{1 \le i \le k} Y_{(1i)}. \tag{93}$$

This tolerance limit is defined by

$$\Pr\left\{ \int_L^{\infty} f(y_{(1)};\mu,\sigma) dy_{(1)} \ge \beta \right\}$$

$$= \Pr\left\{ \frac{L-\mu}{\sigma} \le -\ln \beta^{\frac{1}{km}} \right\} = \gamma. \tag{94}$$

By using the technique of invariant embedding of $(X_{(1)},S_n)$ into a maximal invariant $M=(L-\mu)/\sigma$, we reduce (94) to

$$\Pr\left\{ V_1 \le -\eta V_2 - \ln \beta^{\frac{1}{km}} \right\} = \gamma, \tag{95}$$

where

$$V_1 = \frac{X_{(1)} - \mu}{\sigma}, \quad V_2 = \frac{S_n}{\sigma} \tag{96}$$

are the pivotal quantities,

$$\eta = \frac{L - X_{(1)}}{S_n} \tag{97}$$

is the ancillary factor. It follows from (95) that

$$1 - \frac{\beta^{\frac{n}{km}}}{(1-\eta n)^{n-1}} = \gamma. \tag{98}$$

Therefore, in this case, L can be found explicitly as

$$L = X_{(1)} + \frac{S_n}{n} \left[1 - \left(\frac{\beta^{\frac{n}{km}}}{1-\gamma} \right)^{\frac{1}{n-1}} \right]. \tag{99}$$

For instance, let us suppose that shipments of a lot of electronic systems of a specified type are made to each of 3 customers. Further suppose each customer selects a random sample of 5 systems and accepts his shipment only if no failures occur before a specified time has elapsed. The manufacturer wishes to take a random sample and to calculate the simultaneous lower one-sided β-content tolerance limits so that all shipments will be accepted with a probability of γ at least for 100β% of the future cases of such k shipments, where β=0.95, γ=0.95, and k=3.

The resulting failure times (rounded off to the nearest hour) of an initial sample of size 20 from a population of such electronic systems are: 3149, 3407, 3215, 3296, 3095, 3563, 3178, 3112, 3086, 3160, 3155, 3742, 3143, 3240, 3184, 3621, 3125, 3109, 3118, 3127.

It is assumed that the failure times follow a two-parameter exponential distribution with unknown parameters μ and σ. Thus, for this example, n=20, k=3, m=5, β=0.95 γ=0.95, $X_{(1)}$=3086, and S_n=3105.

The manufacturer finds from (99) that

$$L = 3086 + \frac{3105}{20} \left[1 - \left(\frac{(0.95)^{20/15}}{1-0.95} \right)^{\frac{1}{19}} \right] = 3060. \tag{100}$$

and he has 95% assurance that no failures will occur in each shipment (i.e. each shipment will be accepted) before L=3060 hours at least for 95% of the future cases of such shipments of a lot of electronic systems which will be made to each of three firms.

6 General Problem Analysis

6.1 Inner Minimization

First consider the inner minimization, i.e., k (Section 2) is held fixed for the time being. Then the term ck does not affect the result of this minimization. Consider a situation of state estimation described by one of a family of density functions, indexed by the vector parameter $\theta=(\mu,\sigma)$, where $\mu\equiv x(k)$ and $\sigma\equiv\omega(>0)$ are respectively parameters of location and scale. For this family, invariant under the group of positive linear transformations: $z\to az+b$ with $a>0$, we shall assume that there is obtainable from some informative experiment (a random sample of observations $\mathbf{z}^k=\{z(0), \ldots, z(k)\}$) a sufficient statistic (m_k,s_k) for (μ,σ) with density function $p_k(m_k,s_k;\mu,\sigma)$ of the form

$$p_k(m_k, s_k; \mu, \sigma) = \sigma^{-2} f_k[(m_k - \mu)/\sigma, s_k/\sigma]. \quad (101)$$

We are thus assuming that for the family of density functions an induced invariance holds under the group G of transformations: $m_k\to am_k+b$, $s_k\to as_k$ ($a>0$). 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 loss incurred by making decision d when $\mu\equiv x(l)$ is the true parameter is given by the piecewise-linear loss function

$$r(\theta, d) = \begin{cases} \dfrac{c_1(d-\mu)}{\sigma} & (\mu \le d), \\ \dfrac{c_2(\mu-d)}{\sigma} & (\mu > d). \end{cases} \quad (102)$$

The decision problem specified by the informative experiment density function (101) and the loss function (102) is invariant under the group G of transformations. Thus, the problem is to find the best invariant estimator of μ,

$$d^* = \arg\min_{d\in\mathcal{D}} \mathrm{R}(\theta, d), \quad (103)$$

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

6.2 Best Invariant Estimator

It can be shown by using the invariant embedding technique that an invariant loss function, $r(\theta,d)$, can be transformed as follows:

$$r(\theta, d) = \ddot{r}(\mathbf{v}, \eta), \quad (104)$$

where

$$\ddot{r}(\mathbf{v}, \eta) = \begin{cases} c_1(v_1 + \eta v_2) & (v_1 \ge -\eta v_2), \\ -c_2(v_1 + \eta v_2) & (v_1 < -\eta v_2), \end{cases} \quad (105)$$

$\mathbf{v}=(v_1,v_2)$, $v_1=(m_k-\mu)/\sigma$, $v_2=s_k/\sigma$, $\eta=(d-m_k)/s_k$.

It follows from (104) that the risk associated with d and θ can be expressed as

$$R(\theta, d) = \mathrm{E}_\theta\{r(\theta, d)\} = \mathrm{E}_k\{\ddot{r}(v, \eta)\}$$

$$= c_1 \int_0^\infty dv_2 \int_{-\eta v_2}^{\infty} (v_1 + \eta v_2) f_k(v_1, v_2) dv_1$$

$$- c_2 \int_0^\infty dv_2 \int_{-\infty}^{-\eta v_2} (v_1 + \eta v_2) f_k(v_1, v_2) dv_1, \quad (106)$$

which is constant on orbits when an invariant estimator (decision rule) d is used, where $f_k(v_1,v_2)$ is defined by (101). The fact that the risk (106) is independent of θ means that a decision rule d, which minimizes (106), is uniformly best invariant. The following theorem gives the central result in this section.

Theorem 3 (*Best invariant estimator of* μ). Suppose that (v_1,v_2) is a random vector having density function

$$v_2 f_k(v_1, v_2)\left[\int_0^\infty v_2 dv_2 \int_{-\infty}^{\infty} f_k(v_1, v_2) dv_1\right]^{-1} \quad (v_1 \text{ real}, v_2>0), \quad (107)$$

where f_k is defined by (101), and let G_k be the distribution function of v_1/v_2. Then the uniformly best invariant linear-loss estimator of μ is given by

$$d^* = m_k + \eta^* s_k, \quad (108)$$

where

$$G_k(-\eta^*) = c_1/(c_1 + c_2). \quad (109)$$

Proof. From (106)

$$\frac{\partial E_k\{\ddot{r}(\mathbf{v},\eta)\}}{\partial \eta}$$

$$= c_1\int_0^\infty v_2 dv_2 \int_{-\eta v_2}^\infty f_k(v_1,v_2)dv_1 - c_2\int_0^\infty v_2 dv_2 \int_{-\infty}^{-\eta v_2} f_k(v_1,v_2)dv_1$$

$$= \int_0^\infty v_2 dv_2 \int_{-\infty}^\infty f_k(v_1,v_2)dv_1[c_1 P_k\{(v_1,v_2): v_1+\eta v_2>0\}$$

$$- c_2 P_k\{(v_1,v_2): v_1+\eta v_2<0\}]$$

$$= \int_0^\infty v_2 dv_2 \int_{-\infty}^\infty f_k(v_1,v_2)dv_1[c_1(1-G_k(-\eta)) - c_2 G_k(-\eta)]. \tag{110}$$

Then the minimum of $E_k\{\ddot{r}(\mathbf{v},\eta)\}$ occurs for η^* being determined by setting $\partial E_k\{\ddot{r}(\mathbf{v},\eta)\}/\partial\eta = 0$ and this reduces to

$$c_1[1-G_k(-\eta^*)] - c_2 G_k(-\eta^*) = 0, \tag{111}$$

which establishes (109). □

Corollary 3.1 (*Minimum risk of the best invariant estimator of μ*). The minimum risk is given by

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

$$= c_1\int_0^\infty dv_2 \int_{-\eta^* v_2}^\infty v_1 f_k(v_1,v_2)dv_1 - c_2\int_0^\infty dv_2 \int_{-\infty}^{-\eta^* v_2} v_1 f_k(v_1,v_2)dv_1 \tag{112}$$

with η^* as given by (109).

Proof. These results are immediate from (104) when use is made of $\partial E_k\{\ddot{r}(\mathbf{v},\eta)\}/\partial\eta = 0$. □

6.3 Outer Minimization

The results obtained above can be further extended to find the optimal number of observations. Now

$$E_\theta\{r_k(\theta,d^*)\} = E_\theta\{r(\theta,d^*)+ck\} = E_k\{\ddot{r}(\mathbf{v},\eta^*)+ck\}$$

$$= c_1\int_0^\infty dv_2 \int_{-\eta^* v_2}^\infty v_1 f_k(v_1,v_2)dv_1$$

$$- c_2\int_0^\infty dv_2 \int_{-\infty}^{-\eta^* v_2} v_1 f_k(v_1,v_2)dv_1 + ck \tag{113}$$

is to be minimized with respect to k. It can be shown that this function (which is the constant risk corresponding to taking a sample of fixed sample size k and then estimating $x(l)$ by the expression (108) with k for k^*) has at most two minima (if there are two, they are for successive values of k; moreover, there is only one minimum for all but a denumerable set of values of c). If there are two minima, at k^* and k^*+1, one may randomize in any way between the decisions to take k^* or k^*+1 observations.

7 Example

Consider the one-dimensional discrete-time system, which is described by scalar difference equations of the form (1)-(2), and the case when the measurement noises $w(k)$, $k = 1, 2, \ldots$ (see (2)) are independently and identically distributed random variables drawn from the exponential distribution with the density

$$f(w;\sigma) = (1/\sigma)\exp(-w/\sigma), \quad w\in(0,\infty), \tag{114}$$

where the parameter $\sigma>0$ is unknown. It is required to find the best invariant estimator of $x(l)$ on the basis of the data sample $\mathbf{z}^k=(z(1), \ldots, z(k))$ relative to the piecewise linear loss function

$$r(\theta,d) = \begin{cases} c_1(d-\mu)/\sigma, & d\ge\mu, \\ c_2(\mu-d)/\sigma, & \text{otherwise,} \end{cases} \tag{115}$$

where $\theta=(\mu,\sigma)$, $\mu\equiv x(l)$, $c_1>0$, $c_2=1$.

The likelihood function of $\mathbf{z}^k$ is

$$L(\mathbf{z}^k;\mu,\sigma) = \frac{1}{\sigma^k}\exp\left[-\sum_{j=1}^k\left(z(j)-H(j)x(j)\right)/\sigma\right]$$

$$= \frac{1}{\sigma^k} \exp\left[-\sum_{j=1}^{k} a(j)\big(y(j)-\mu\big)/\sigma \right], \qquad (116)$$

where

$$y(j) = [a(j)]^{-1}\left(z(j) + H(j)\sum_{i=j}^{l-1} A(j,i+1)B(i)u(i) \right),$$

$$j \le l, \qquad (117)$$

$$y(j) = [a(j)]^{-1}\left(z(j) - H(j)\sum_{i=l}^{j-1} A(j,i+1)B(i)u(i) \right),$$

$$j > l, \qquad (118)$$

if $l < k$ (estimation of the past state of the system), and

$$y(j) = \frac{z(j)+b(j)}{a(j)}, \qquad (119)$$

$$a(j) = H(j)A(j,k_1), \qquad (120)$$

$$b(j) = H(j)\sum_{i=j}^{l-1} A(j,i+1)B(i)u(i), \qquad (121)$$

if either $l = k$ (estimation of the current state of the system) or $l > k$ (prediction of the future state of the system).

It can be justified by using the factorization theorem that (m_k,s_k) is a sufficient statistic for $\theta=(\mu,\sigma)$, where

$$m_k = \min_{1\le j\le k} y(j), \quad s_k = \sum_{j=1}^{k} a(j)[y(j)-m_k]. \qquad (122)$$

The probability density function of (m_k,s_k) is given by

$$p_k(m_k, s_k; \mu,\sigma)$$

$$= \frac{n(k)}{\sigma} e^{-\frac{n(k)[m_k-\mu]}{\sigma}} \frac{1}{\Gamma(k-1)\sigma^{k-1}} s_k^{k-2} e^{-\frac{s_k}{\sigma}},$$

$$m_k > \mu, \quad s_k > 0, \qquad (123)$$

where

$$n(k) = \sum_{j=1}^{k} a(j). \qquad (124)$$

Since the loss function (115) is invariant under the group G of location and scale changes, it follows that

$$r(\boldsymbol{\theta}, d) = \ddot{r}(\mathbf{v},\eta) = \begin{cases} c_1(v_1+\eta v_2)/\sigma, & v_1 \ge -\eta v_2, \\ -(v_1+\eta v_2)/\sigma, & \text{otherwise}, \end{cases} \qquad (125)$$

where $\mathbf{v}=(v_1,v_2)$,

$$v_1 = \frac{m_k-\mu}{\sigma}, \quad v_2 = \frac{s_k}{\sigma}, \quad \eta = \frac{d-m_k}{s_k}. \qquad (126)$$

Thus, using (108) and (109), we find that the best invariant estimator (BIE) of μ is given by

$$d_{\text{BIE}} = m_k + \eta^* s_k, \qquad (127)$$

where

$$\eta^* = [1-(c_1+1)^{1/k}]/n(k) = \arg\inf_{\eta} \mathrm{E}_k\{\ddot{r}(\mathbf{v},\eta)\}, \qquad (128)$$

$$\mathrm{E}_k\{\ddot{r}(\mathbf{v},\eta)\} = [(c_1+1)(1-\eta n(k))^{-(k-1)} - 1]/n(k)$$

$$-\eta(k-1). \qquad (129)$$

The risk of this estimator is

$$R(\boldsymbol{\theta}, d_{\text{BIE}}) = \mathrm{E}_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta}, d_{\text{BIE}})\}$$

$$= \mathrm{E}_k\{\ddot{r}(\mathbf{v},\eta^*)\} = k[(c_1+1)^{1/k} - 1]/n(k). \qquad (130)$$

Here the following theorem holds.

Theorem 4 (*Characterization of the estimator* d_{BIE}). For the loss function (115), the best invariant estimator of μ, d_{BIE}, given by (127) is uniformly non-dominated.

Proof. The proof follows immediately from Theorem 1 if we use the prior distribution on the parameter space Θ,

$$\xi_\tau(d\theta) = \frac{1}{\tau\sigma} e^{-\frac{\tau-\mu}{\tau\sigma}} \frac{1}{\Gamma(1/\tau)\sigma^{1/\tau+1}} \left(\frac{1}{\tau}\right)^{1/\tau} e^{-\frac{1}{\tau\sigma}} d\mu d\sigma,$$

$$\mu \in (-\infty, \tau), \quad \sigma \in (0, \infty). \tag{131}$$

This ends the proof. □

Consider, for comparison, the following estimators of μ (state of the system):

The maximum likelihood estimator (MLE):

$$d_{\text{MLE}} = m_k; \tag{132}$$

The minimum variance unbiased estimator (MVUE):

$$d_{\text{MVUE}} = m_k - \frac{s_k}{(k-1)n(k)}; \tag{133}$$

The minimum mean square error estimator (MMSEE):

$$d_{\text{MMSEE}} = m_k - \frac{s_k}{kn(k)}; \tag{134}$$

The median unbiased estimator (MUE):

$$d_{\text{MUE}} = m_k - (2^{1/(k-1)} - 1)\frac{s_k}{n(k)}. \tag{135}$$

Each of the above estimators is readily seen to be of a member of the class

$$\mathscr{C} = \{d : d = m_k + \eta s_k\}, \tag{136}$$

where η is a real number. A risk of an estimator, which belongs to the class $\mathscr{C}$, is given by (129). If, say, k=3 and c_1=26, then we have that

$$\text{rel.eff.}_R\{d_{\text{MLE}}, d_{\text{BIE}}; \theta\} = 0.231, \tag{137}$$

$$\text{rel.eff.}_R\{d_{\text{MVUE}}, d_{\text{BIE}}; \theta\} = 0.5, \tag{138}$$

$$\text{rel.eff.}_R\{d_{\text{MMSEE}}, d_{\text{BIE}}; \theta\} = 0.404, \tag{139}$$

$$\text{rel.eff.}_R\{d_{\text{MUE}}, d_{\text{BIE}}; \theta\} = 0.45. \tag{140}$$

In this case (113) becomes

$$\text{E}_\theta\left\{r_k(\theta, d^*)\right\} = \text{E}_\theta\left\{r(\theta, d_{\text{BIE}}) + ck\right\}$$

$$= \text{E}_k\left\{\ddot{r}(\mathbf{v}, \eta^*) + ck\right\}$$

$$= k[(c_1+1)^{1/k} - 1]/n(k) + ck \equiv J_k. \tag{141}$$

Now (141) is to be minimized with respect to k. It is easy to see that

$$J_k - J_{k-1} = -\Big((k-1)[(c_1+1)^{1/(k-1)} - 1]/n(k-1)$$

$$- k[(c_1+1)^{1/k} - 1]/n(k)\Big) + c. \tag{142}$$

Define

$$\varphi(k) = (k-1)[(c_1+1)^{1/(k-1)} - 1]/n(k-1)$$

$$- k[(c_1+1)^{1/k} - 1]/n(k). \tag{143}$$

Thus

$$c \gtrless \varphi(k) \Leftrightarrow J_k \gtrless J_{k-1}. \tag{144}$$

By plotting $\varphi(k)$ versus k the optimal number of observations k^* can be determined.

For each value of c, we can find an equilibrium point of k, i.e., $c=\varphi(k^\bullet)$. The following two cases must be considered:

1) $k^\bullet$ is not an integer. We have $k^{(1)}<k^\bullet<k^{(1)}+1=k^{(2)}$, where $k^{(1)}$ and $k^{(2)}$ are neighboring integers. Since $\varphi(k)$ is monotonically decreasing, we know that $\varphi(k^{(1)})>c$ and $\varphi(k^{(2)})<c$. Then, by using these properties, (133) becomes

$$J_{k^{(1)}} - J_{k^{(1)}-1} = -\varphi(k^{(1)}) + c < 0, \tag{145}$$

$$J_{k^{(2)}} - J_{k^{(1)}} = -\varphi(k^{(2)}) + c > 0, \tag{146}$$

Thus

$$J_{k^{(2)}} > J_{k^{(1)}} < J_{k^{(1)}-1}. \qquad (147)$$

Therefore, $k^{(1)}$ is the optimal number of observations. We conclude that the optimal number k^* is equal to the largest integer below the equilibrium point.

2) $k^\bullet$ is an integer. By the same sort of argument, we know that $k^\bullet$ is as good as $k^\bullet$-1. Consequently, both $k^\bullet$ and $k^\bullet$-1 are the optimal number of observations. Notice that in this case, J_{k^*} can be computed directly and precisely from (132).

8 Conclusions and Directions for Future Research

In this paper we construct the minimum risk estimators of state of stochastic systems. 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.

For a class of state estimation problems where observations on system state vectors are constrained, i.e., when it is not feasible to make observations at every moment, the question of how many observations to take must be answered. This paper models such a class of problems by assigning a fixed cost to each observation taken. The total number of observations is determined as a function of the observation cost.

Extension to the case where the observation cost is an explicit function of the number of observations taken is straightforward. A different way to model the observation constraints should be investigated.

More work is needed, however, to obtain improved decision rules for the problems of unconstrained and constrained optimization under parameter uncertainty when: (i) the observations are from general continuous exponential families of distributions, (ii) the observations are from discrete exponential families of distributions, (iii) some of the observations are from continuous exponential families of distributions and some from discrete exponential families of distributions, (iv) the observations are from multiparameter or multidimensional distributions, (v) the observations are from truncated distributions, (vi) the observations are censored, (vii) the censored observations are from truncated distributions.

Appendix: Further Applications of the Invariant Embedding Technique

A.1 Finding Shortest-Length Confidence Intervals for System Availability

Availability analysis is performed to verify that an item has a satisfactory probability of being operational, so it can achieve its intended objective. An item's availability can be considered as combination of its reliability and maintainability. Accordingly, when no maintenance repair is performed (e.g., in nonrepairable items), reliability can be considered as instantaneous availability. Availability is very important to users of repairable products and systems, such as computer networks, manufacturing systems, power plants, transportation vehicles, and fire-protection systems.

Mathematically, the availability of an item is a measure of the fraction of time that the item is in operating condition in relation to total or calendar time, i.e., availability indicates the percent of the time that products are expected to operate satisfactory. There are several measures of availability, namely, inherent availability, achieved availability, and operational availability. For further definition of these availability measures, see [18]. Here, we consider inherent availability, which is the most common definition used in the literature. This availability, A, is the designed-in capability of a product and is defined by [19]

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}, \qquad \text{(A1)}$$

where MTTR is the Mean Time To Repair (more generally, the mean time that the process is inoperable when it is down for maintenance or because of a breakdown) and MTBF is the Mean Time Between Failures (more generally, the mean operating time between one downtime and the next, where each downtime can be due to maintenance or a breakdown).

Actually the true inherent availability is rarely known. Usually, it is estimated from the few collected data on the operating (up) times and repair/replace (down) times. The point estimate of the availability is then given by

$$\widehat{A} = \frac{\widehat{\text{MTBF}}}{\widehat{\text{MTBF}} + \widehat{\text{MTTR}}}, \qquad \text{(A2)}$$

where $\widehat{A}$ is an estimate of the inherent availability, $\widehat{\text{MTBF}}$ is an estimate of MTBF from sample data, $\widehat{\text{MTTR}}$ is an estimate of MTTR from sample data.

Obviously, this point estimate is a function of the sample data and the sample size. Different samples will result in different estimates. The sample error affects the quantification of the calculated availability. If the estimates were based on one failure and one repair only, it would be quite risky [20]. We would feel more confident if we had more data (more failures and repairs). The question is how good the estimated inherent availability is. The answer is to attach a confidence level to the calculated availability, or give the confidence limits on the availability at a chosen confidence level. The most interesting confidence limits would be the shortest-length confidence limits on the true availability at a given confidence level.

In a wide variety of inference problems one is not interested in estimating the parameter or testing some hypothesis concerning it. Rather, one wishes to establish a lower or an upper bound, or both, for the real-valued parameter. For example, if X is the time to failure of a piece of equipment, one is interested in a lower bound for the mean of X. If the rv X measures the toxicity of a drug, the concern is to find an upper bound for the mean. Similarly, if the rv X measures the nicotine content of a certain brand of cigarettes, one is interested in determining an upper and a lower bound for the average nicotine content of these cigarettes.

The following result provides a general method of finding shortest-length confidence intervals and covers most cases in practice.

Let $S=s(\mathbf{X})$ be a statistic, based on a random sample $\mathbf{X}$. Let F be the distribution function of the pivotal quantity $V(S,A) \equiv A$ and let v_L, v_U be such that

$$F(v_U) - F(v_L) = \Pr\{v_L < V < v_U\} = 1-\alpha. \quad \text{(A3)}$$

It will be noted that the distribution of V does not depend on any unknown parameter. A $100(1-\alpha)\%$ confidence interval of A is $(A_L(S,v_L,v_U),A_U(S,v_L,v_U))$ and the length of this interval is $\Delta(S, v_L, v_U) = A_U - A_L$. We want to choose v_1, v_2, minimizing $A_U - A_L$ and satisfying (A3). Thus, we consider the problem:

Minimize:

$$\Delta(S, v_L, v_U) = A_U - A_L, \quad \text{(A4)}$$

Subject to:

$$F(v_U) - F(v_L) = 1-\alpha. \quad \text{(A5)}$$

The search for the shortest-length confidence interval $\Delta = A_U - A_L$ is greatly facilitated by the use of the following theorem.

Theorem A1. Under appropriate derivative conditions, there will be a pair (v_L, v_U) giving rise to the shortest-length confidence interval $\Delta(S, v_L, v_U) = A_U - A_L$ for A as a solution to the simultaneous equations:

$$\frac{\partial \Delta}{\partial v_L} + \frac{\partial \Delta}{\partial v_U}\frac{F'(v_L)}{F'(v_U)} = 0, \quad \text{(A6)}$$

$$F(v_U) - F(v_L) = 1-\alpha. \quad \text{(A7)}$$

Proof. Note that (A5) forces v_U to be a function of v_L (or visa-versa). Take $\Delta(S,v_L,v_U)$ as a function of v_L, say $\Delta(S,v_L,v_U(v_L))$. Then, by using the method of Lagrange multipliers, the proof follows immediately. □

A.1.1 Example 1

Consider the problem of constructing the shortest-length confidence interval for system availability from time-to-failure and time-to-repair test data. It is assumed that X_1 (time-to-failure) and X_2 (time-to-repair) are stochastically independent random variables with probability density functions

$$f_1(x_1;\theta_1) = \frac{1}{\theta_1}e^{-x_1/\theta_1}, \quad x_1 \in (0,\infty), \quad \theta_1 > 0, \quad \text{(A8)}$$

and

$$f_2(x_2;\theta_2) = \frac{1}{\theta_2}e^{-x_2/\theta_2}, \quad x_2 \in (0,\infty), \quad \theta_2 > 0. \quad \text{(A9)}$$

Availability is usually defined as the probability that a system is operating satisfactorily at any point in time. This probability can be expressed mathematically as

$$A = \frac{\theta_1}{\theta_1 + \theta_2}, \quad \text{(A10)}$$

where θ_1 is a system mean-time-to-failure, θ_2 is a system mean-time-to-repair.

Consider a random sample $\mathbf{X}_1=(X_{11},\ldots,X_{1n_1})$ of n_1 times-to-failure and a random sample $\mathbf{X}_2=(X_{21},\ldots,X_{2n_2})$ of n_2 times-to-repair drawn from the populations described by (A8) and (A9) with sample means

$$\overline{X}_1 = \sum_{i=1}^{n_1} X_{1i}/n_1, \qquad \overline{X}_2 = \sum_{i=1}^{n_2} X_{2i}/n_2. \quad \text{(A11)}$$

It is well known that $2n_1\overline{X}_1/\theta_1$ and $2n_2\overline{X}_2/\theta_2$ are chi-square distributed variables with $2n_1$ and $2n_2$ degrees of freedom, respectively. They are independent due to the independence of the variables X_1 and X_2.

It follows from (A10) that

$$\frac{A}{1-A}=\frac{\theta_1}{\theta_2}. \qquad (A12)$$

Using the invariant embedding technique, we obtain from (A12) a pivotal quantity

$$V(S,A)=S\frac{A}{1-A}=\frac{\overline{X}_2}{\overline{X}_1}\frac{\theta_1}{\theta_2}$$

$$=\left(\frac{2n_2\overline{X}_2/\theta_2}{2n_2}\right)\Bigg/\left(\frac{2n_1\overline{X}_1/\theta_1}{2n_1}\right), \qquad (A13)$$

which is F-distributed with $(2n_2,2n_1)$ degrees of freedom, and

$$S=\overline{X}_2/\overline{X}_1. \qquad (A14)$$

Thus, (A13) allows one to find a $100(1-\alpha)\%$ confidence interval for A from

$$\Pr\{A_L<A<A_U\}=1-\alpha, \qquad (A15)$$

where

$$A_L=\frac{v_L}{v_L+S} \quad \text{and} \quad A_U=\frac{v_U}{v_U+S}. \qquad (A16)$$

It follows from Theorem A1 that the shortest-length confidence interval for A is given by

$$C_A^*=(A_L,A_U) \qquad (A17)$$

with

$$\Delta^*(S,v_L,v_U)=A_U-A_L, \qquad (A18)$$

where v_L and v_U are a solution of

$$(v_L+S)^2f(v_L)=(v_U+S)^2f(v_U) \qquad (A19)$$

(f is the pdf of an F-distributed rv with $(2n_2,2n_1)$ d.f.) and

$$\Pr\{v_L<V<v_U\}=\Pr\{v_L<F(2n_2,2n_1)<v_U\}=1-\alpha. \qquad (A20)$$

In practice, the simpler equal tails confidence interval for A,

$$C_A=(A_L,A_U)=\left(\frac{v_L}{v_L+S},\frac{v_U}{v_U+S}\right) \qquad (A21)$$

with

$$\Delta(S,v_L,v_U)=A_U-A_L, \qquad (A22)$$

is employed, where

$$v_L=F_{\alpha/2}(2n_2,2n_1), \quad v_U=F_{1-\alpha/2}(2n_2,2n_1), \qquad (A23)$$

and

$$\Pr\{F(2n_2,2n_1)>F_{\alpha/2}(2n_2,2n_1)\}\le 1-\alpha/2. \qquad (A24)$$

Consider, for instance, the following case. A total of 400 hours of operating time with 2 failures, which required an average of 20 hours of repair time, were observed for aircraft air-conditioning equipment. What is the confidence interval for the inherent availability of this equipment at the 90% confidence level?

The point estimate of the inherent availability is

$$\widehat{A}=\frac{200}{200+20}=0.909, \qquad (A25)$$

and the confidence interval for the inherent availability, at the 90% confidence level, is found as follows.

From (A21), the simpler equal tails confidence interval is

$$C_A=\left(\frac{F_{0.05}(4,4)}{F_{0.05}(4,4)+1/\widehat{A}-1},\frac{F_{0.95}(4,4)}{F_{0.95}(4,4)+1/\widehat{A}-1}\right)$$

$$=(0.61,0.985), \qquad (A26)$$

i.e.,

$$\Delta(S,v_L,v_U)=A_U-A_L=0.375. \qquad (A27)$$

From (A17), the shortest-length confidence interval is

$$C_A^*=\left(\frac{v_L}{v_L+S},\frac{v_U}{v_U+S}\right)=(0.707,0.998), \qquad (A28)$$

where v_L and v_U are a solution of (A19) and (A20). Thus,

$$\Delta^*(S,v_L,v_U)=A_U-A_L=0.291. \qquad (A29)$$

The relative efficiency of C_A relative to C_A^* is given by

$$\text{rel.eff.}_{\text{C}}\left(C_A, C_A^*\right) = \frac{\Delta^*(S, v_L, v_U)}{\Delta(S, v_L, v_U)} = \frac{0.291}{0.375} = 0.776. \quad \text{(A30)}$$

A.1.2 Example 2

Consider the problem of constructing the shortest-length confidence interval for system availability from time-to-failure and time-to-repair test data when the operating time distribution is lognormal and repair time distribution is Inverse Gaussian.

Let us assume that the operating time X has a lognormal distribution with probability density function (pdf) given by

$$f(x;a,b) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right],$$

$$0 < x < \infty, \ \ \mu \in (-\infty, \infty), \ \sigma > 0 \quad \text{(A31))}$$

and is denoted by $\Lambda(\mu,\sigma^2)$. Note that $E\{X\}$ $=\exp(\mu+\sigma^2/2)$. The repair time Y has an Inverse Gaussian distribution with a pdf given by

$$g(y;a,\lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[-\frac{\lambda(y-a)^2}{2a^2 y}\right],$$

$$0 < y < \infty, \ \ a, \lambda > 0 \quad \text{(A32)}$$

and is denoted by IG(λ,a). It may be noted that $E\{Y\}=a$. The parameter σ of the lognormal distribution and only the ratio λ/a of the IG distribution are known.

Let $X_1, X_2, \ldots, X_m$ be a random sample of times to operate with the pdf (A31). It can be shown that

$$U = e^{-\mu}\left(\prod_{i=1}^{m} X_i\right)^{1/m} = e^{-\mu}G \sim \Lambda\left(0, \frac{\sigma^2}{m}\right), \quad \text{(A33)}$$

where G is the sample geometric mean of time to operate.

Let $Y_1, Y_2, \ldots, Y_n$ be another random sample of times to repair with the pdf (A32). It can be shown that

$$W = \frac{\lambda}{a^2}\sum_{j=1}^{n} Y_j \sim \text{IG}\left(\frac{n^2\lambda^2}{a^2}, \frac{n\lambda}{a}\right). \quad \text{(A34)}$$

If the pivotal quantities U and W are independent, then the joint density of (U, W) is given by

$$f(u,w;m,n,\sigma,\lambda/a) = \frac{\sqrt{m}(n\lambda/a)}{(2\pi\sigma)uw\sqrt{w}}$$

$$\times \exp\left[-\frac{1}{2}\left(\frac{m(\ln u)^2}{\sigma^2} + \frac{1}{w}\left(w - \frac{n\lambda}{a}\right)^2\right)\right], \ \ 0 < u, w < \infty. \quad \text{(A35)}$$

By making transformations $V=U/W$ and $Z=W$, it can be shown that the pdf of V is

$$f(v;m,n,\sigma,\lambda/a)$$

$$= \vartheta\int_0^\infty \frac{1}{vz\sqrt{z}} \exp\left[-\frac{1}{2}\left(\frac{m(\ln vz)^2}{\sigma^2} + \frac{1}{z}\left(z - \frac{n\lambda}{a}\right)^2\right)\right] dz,$$

$$0 < v < \infty, \quad \text{(A36)}$$

where

$$\vartheta = \frac{\sqrt{m}(n\lambda/a)}{2\pi\sigma}. \quad \text{(A37)}$$

Now we can determine v_{L} and v_{U} such that

$$\Pr\{v_{\text{L}} < V < v_{\text{U}}\} = \vartheta\int_{v_{\text{L}}}^{v_{\text{U}}}\int_0^\infty \frac{1}{vz\sqrt{z}}$$

$$\times \exp\left[-\frac{1}{2}\left(\frac{m(\ln vz)^2}{\sigma^2} + \frac{1}{z}\left(z - \frac{n\lambda}{a}\right)^2\right)\right] dz dv = 1-\alpha. \quad \text{(A38)}$$

It follows from (A1) that

$$\frac{1-A}{A} = \frac{a}{\exp(\mu + \sigma^2/2)}. \quad \text{(A39)}$$

Using the invariant embedding technique, we obtain from (A39) a pivotal quantity

$$V(S_\bullet, A) = S_\bullet\frac{1-A}{A} = \frac{e^{-\mu}G}{(\lambda/a^2)n\bar{Y}} = \frac{U}{W}, \quad \text{(A40)}$$

where

$$S_\bullet = \frac{Ga\exp(\sigma^2/2)}{\lambda n\bar{Y}}. \quad \text{(A41)}$$

Thus, (A40) allows one to find a 100(1–α)% confidence interval for A from

$$\Pr\{A_{\text{L}} < A < A_{\text{U}}\} = 1-\alpha, \quad \text{(A42)}$$

where

$$A_{\mathrm{L}} = \frac{S_\bullet}{v_{\mathrm{U}} + S_\bullet} \quad \text{and} \quad A_{\mathrm{U}} = \frac{S_\bullet}{v_{\mathrm{L}} + S_\bullet}. \qquad \text{(A43)}$$

Clearly, (A42) is an exact confidence interval for A, if the parameter σ for lognormal distribution and only the ratio λ/a of IG distribution are known.

It follows from Theorem A1 that the shortest-length confidence interval for A is given by

$$C_A^* = (A_{\mathrm{L}}, A_{\mathrm{U}}) \qquad \text{(A44)}$$

with

$$\Delta^*(S_\bullet, v_{\mathrm{L}}, v_{\mathrm{U}}) = A_{\mathrm{U}} - A_{\mathrm{L}}, \qquad \text{(A45)}$$

where v_{L} and v_{U} are a solution of

$$(v_{\mathrm{L}} + S_\bullet)^2 f(v_{\mathrm{L}}; m, n, \sigma, \lambda/a)$$

$$= (v_{\mathrm{U}} + S_\bullet)^2 f(v_{\mathrm{U}}; m, n, \sigma, \lambda/a). \qquad \text{(A46)}$$

and

$$\Pr\{v_{\mathrm{L}} < V < v_{\mathrm{U}}\} = \int_{v_{\mathrm{L}}}^{v_{\mathrm{U}}} f(v; m, n, \sigma, \lambda/a)\,dv = 1 - \alpha. \qquad \text{(A47)}$$

Acknowledgments

This work was supported in part by Research Grant No. 06.1936 and Research Grant No. 07.2036 from the Latvian Council of Science and the National Institute of Mathematics and Informatics of Latvia.

The authors are also very grateful to three anonymous referees for their detailed comments that have helped in improving an earlier version of this paper.

References:

[1] R. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Trans. ASME, J. Basic Eng.*, Vol. 82, 1960, pp. 34-45,

[2] T. P. McGarty, *Stochastic Systems and State Estimation*, New York: John Wiley and Sons, Inc., 1974.

[3] A. Savkin and I. Petersen, Robust State Estimation and Model Validation for Discrete-time Uncertain System with a Deterministic Description of Noise and Uncertainty, *Automatica*, Vol. 34, 1998, pp. 271–274.

[4] M. Norgaard, N. K. Poulsen and O. Ravn, New Developments in State Estimation for Nonlinear Systems, *Automatica*, Vol. 36, 2000, pp. 1627-1638.

[5] J. Yan and R. R. Bitmead, Incorporating State Estimation into Model Predictive Control and its Application to Network Traffic Control, *Automatica*, Vol. 41, 2005, pp. 595 – 604.

[6] T. Alamo, J. Bravo and E. Camacho, Guaranteed State Estimation by Zonotopes, *Automatica*, Vol. 41, 2005, pp. 1035–1043.

[7] S. Gillijns and B. De Moor, Unbiased Minimum-variance Input and State Estimation for Linear Discrete-time Systems, *Automatica*, Vol. 43, 2007, pp. 111 – 116.

[8] S. Ko and R. R. Bitmead, State Estimation for Linear Systems with State Equality Constraints, *Automatica*, Vol. 43, 2007, pp. 1363 – 1368.

[9] N.A. Nechval, *Modern Statistical Methods of Operations Research*, Riga: RCAEI, 1982.

[10] N.A. Nechval, *Theory and Methods of Adaptive Control of Stochastic Processes*, Riga: RCAEI, 1984.

[11] N.A. Nechval, K.N. Nechval and E.K. Vasermanis, Optimization of Interval Estimators via Invariant Embedding Technique, *IJCAS* (*An International Journal of Computing Anticipatory Systems*), Vol. 9, 2001, pp. 241-255.

[12] E.K. Vasermanis and N.A. Nechval, *Improved Decisions in Statistics*, Riga: SIA, 2004.

[13] G.J. Hahn and W. Nelson, A Survey of Prediction Intervals and their Applications, *J. Qual. Tech.*, Vol. 5, 1973, pp. 178-188.

[14] R.E. Barlow and F. Proshan, Tolerance and Confidence Limits for Classes of Distributions Based on Failure Rate, *Ann. Math. Stat.*, Vol. 37, 1966, pp. 1593-1601.

[15] B. Epstein and M. Sobel, Some Theorems Relevant to Life Testing from an Exponential Population, *Ann. Math. Statist.*, Vol. 25, 1954, pp. 373-381.

[16] M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1 (3rd edition), London: Griffin, 1969.

[17] F.E. Grubbs, Approximate Fiducial Bounds on Reliability for the Two Parameter Negative Exponential Distribution, *Technometrics*, Vol. 13, 1971, pp. 873-876.

[18] W.G. Ireson and C.F. Coombs, eds, *Handbook of Reliability Engineering and Management*, New York: McGraw-Hill, 1988.

[19] M. Ben-Daya, S.O. Duffuaa and A. Raouf, eds, *Maintenance, Modeling and Optimization,* Norwell, Massachusetts: Kluwer Academic Publishers, 2000.

[20] A. Coppola, Some Observations on Demonstrating Availability, *RAC Journal*, Vol. 6, 1997, pp. 17-18.