Optimization of Statistical Decisions via an Invariant Embedding Technique

Nicholas A. Nechval, Maris Purgailis, Kaspars Cikste, Gundars Berzins, and Konstantin N. Nechval

Abstract— **In the present paper, for optimization of statistical decisions under parametric uncertainty, a new technique of invariant embedding of sample statistics in a performance index 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 decision rule, which has smaller risk than any of the well-known decision rules. To illustrate the proposed technique, application examples are given.**

Index Terms— **Optimization, parametric uncertainty, statistical decision rule, technique of invariant embedding.**

I. Introduction

Most of the operations research and management science literature assumes that the true distributions are specified explicitly. However, in many practical situations, the true distributions are not known, and the only information available may be a time-series (or random sample) of the past data.

Analysis of decision-making problems with unknown distribution is not new. Several important papers have appeared in the literature. When the true distribution is unknown, one may either use a parametric approach (where it is assumed that the true distribution belongs to a parametric family of distributions) or a non-parametric approach (where no assumption regarding the parametric form of the unknown distribution is made).

Under the parametric approach, one may choose to estimate the unknown parameters or choose a prior distribution for the unknown parameters and apply the Bayesian approach to incorporating the past data available. Parameter estimation is first considered in [1] and recent development is reported in [2]. Scarf [3] considers a Bayesian framework for the unknown demand distribution. Specifically, assuming that the demand distribution belongs to the family of exponential distributions, the demand process is characterized by the prior distribution on the unknown parameter. Further extension of this approach is presented in [4].

Within the non-parametric approach, either the empirical distribution [2] or the bootstrapping method (e.g. see [5]) can be applied with the available past data to obtain a statistical decision rule.

A third alternative to dealing with the unknown distribution is when the random variable is partially characterized by its moments. When the unknown demand distribution is characterized by the first two moments, Scarf [6] derives a robust min–max inventory control policy. Further development and review of this model is given in [7].

In the present paper we consider the case, where it is known that the true distribution function belongs to a parametric family of distributions. It will be noted that, in this case, most stochastic models to solve the problems of control and optimization of system and processes are developed in the extensive literature under the assumptions that the parameter values of the underlying distributions are known with certainty. In actual practice, such is simply not the case. When these models are applied to solve real-world problems, the parameters are estimated and then treated as if they were the true values. The risk associated with using estimates rather than the true parameters is called estimation risk and is often ignored. When data are limited and (or) unreliable, estimation risk may be significant, and failure to incorporate it into the model design may lead to serious errors. Its explicit consideration is important since decision rules that are optimal in the absence of uncertainty need not even be approximately optimal in the presence of such uncertainty.

The problem of determining an optimal decision rule in the absence of complete information about the underlying distribution, i.e., when we specify only the functional form of the distribution and leave some or all of its parameters unspecified, 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

Manuscript received March 12, 2010. This work was supported in part by Grant No. 06.1936, Grant No. 07.2036, Grant No. 09.1014, and Grant No. 09.1544 from the Latvian Council of Science and the National Institute of Mathematics and Informatics of Latvia.

N. A. Nechval is with the Statistics Department, EVF Research Institute, University of Latvia, Riga LV-1050, Latvia (e-mail: nechval@junik.lv).

M. Purgailis is with the University of Latvia, Riga LV-1050, Latvia (e-mail: maris.purgailis@lu.lv).

K. Cikste is with the University of Latvia, Riga LV-1050, Latvia (e-mail: kaspars.cikste@lu.lv).

G. Berzins is with the University of Latvia, Riga LV-1050, Latvia (e-mail: gundars.berzins@lu.lv).

K. N. Nechval is with the Applied Mathematics Department, Transport and Telecommunication Instutute, Riga LV-1019, Latvia (e-mail: konstan@tsi.lv).

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 statistical 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.

In this paper, we propose a new technique to solve optimization problems of statistical decisions under parametric uncertainty. The technique is based on the constructive use of the invariance principle for improvement (or optimization) of statistical decisions. It allows one to yield an operational, optimal information-processing rule and may be employed for finding the effective statistical decisions for many problems of the operations research and management science, the illustrative application examples of which are given below.

II. 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 for decision-making. 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 improvement or optimization of 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 [8-15]

A. Preliminaries

In the general formulation of decision theory, we observe a random variable $\mathbf{X}$ (which may be multivariate) with distribution function $F(\mathbf{x}|\boldsymbol{\theta})$ where a parameter $\boldsymbol{\theta}$ (in general, vector) is unknown, $\boldsymbol{\theta}\in\Theta$, and if we choose decision d from the set of all possible decisions $\mathcal{D}$, then we suffer a loss $l(d,\boldsymbol{\theta})$. A "decision rule" is a method of choosing d from $\mathcal{D}$ after observing $\mathbf{x}\in\mathcal{X}$, that is, a function $u(\mathbf{x})=d$. Our average loss (called risk) $E_{\mathbf{x}}\{l(u(\mathbf{X}),\boldsymbol{\theta})\}$ is a function of both $\boldsymbol{\theta}$ and the decision rule $u(\cdot)$, called the risk function $r(u,\boldsymbol{\theta})$, and is the criterion by which rules are compared. Thus, the expected loss (gains are negative losses) is a primary consideration in evaluating decisions. We will now define the major quantities just introduced.

Definition 1. A general statistical decision problem is a triplet $(\Theta,\mathcal{D},l)$ and a random variable $\mathbf{X}$. The random variable $\mathbf{X}$ (called the data) has a distribution function $F(\mathbf{x}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is unknown but it is known that $\boldsymbol{\theta}\in\Theta$. $\mathcal{X}$ will denote the set of possible values of the random variable $\mathbf{X}$. $\boldsymbol{\theta}$ is called the state of nature, while the nonempty set Θ is called the parameter space. The nonempty set $\mathcal{D}$ is called the decision space or action space. Finally, l is called the loss function and to each $\boldsymbol{\theta}\in\Theta$ and $d\in\mathcal{D}$ it assigns a real number $l(d,\boldsymbol{\theta})$.

Definition 2. For a statistical decision problem $(\Theta,\mathcal{D},l)$, $\mathbf{X}$, a (nonrandomized) decision rule is a function $u(\cdot)$ which to each $\mathbf{x}\in\mathbf{X}$ assigns a member d of $\mathcal{D}$: $u(\mathbf{X})=d$.

Definition 3. The risk function $r(u,\boldsymbol{\theta})$ of a decision rule $u(\mathbf{X})$ for a statistical decision problem $(\Theta,\mathcal{D},l)$, $\mathbf{X}$ (the expected loss or average loss when $\boldsymbol{\theta}$ is the state of nature and a decision is chosen by rule $u(\cdot)$) is $r(u,\boldsymbol{\theta})=E_{\mathbf{x}}\{l(u(\mathbf{X}),\boldsymbol{\theta})\}$.

This paper is concerned with the implications of group theoretic structure for invariant loss functions. 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 $(\mathcal{D}, \mathcal{B})$, and a real-valued loss function

$$l(d,\boldsymbol{\theta}) = E_x\left\{l^{\circ}(d,X)\right\} \tag{1}$$

defined on $\Theta\times\mathcal{D}$, where $l^{\circ}(d,X)$ is a random loss function with a random variable $X\in(0,\infty)$ (or $(-\infty,\infty)$). 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 $\mathcal{D}$, respectively. ($\overline{G}$ may be induced on Θ through ψ; $\widetilde{G}$ may be induced on $\mathcal{D}$ through l). We shall say that l is invariant if for every $(\boldsymbol{\theta}, d)\in\Theta\times\mathcal{D}$

$$l(\tilde{g}d,\bar{g}\boldsymbol{\theta}) = l(d,\boldsymbol{\theta}),\quad g\in G. \tag{2}$$

A loss function, $l(d,\boldsymbol{\theta})$, can be transformed as follows:

$$l(d,\boldsymbol{\theta}) = l(\tilde{g}_{\hat{\boldsymbol{\theta}}}^{-1}d,\bar{g}_{\hat{\boldsymbol{\theta}}}^{-1}\boldsymbol{\theta}) = l^{\#}(\eta,\mathbf{V}), \tag{3}$$

where $\mathbf{V}=\mathbf{V}(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$ is a pivotal quantity whose distribution does not depend on unknown parameter $\boldsymbol{\theta}$; $\eta=\eta(d,\hat{\boldsymbol{\theta}})$ is an ancillary factor; $\hat{\boldsymbol{\theta}}$ is a maximum likelihood estimator of $\boldsymbol{\theta}$ (or a sufficient statistic for $\boldsymbol{\theta}$). Then the best invariant decision rule (BIDR) is given by

$$u^{\mathrm{BIDR}} \equiv d^{*} = \eta^{-1}(\eta^{*},\hat{\boldsymbol{\theta}}), \tag{4}$$

where

$$\eta^{*} = \arg\inf_{\eta} E_{\mathrm{v}}\left\{l^{\#}(\eta,\mathbf{V})\right\} \tag{5}$$

and a risk function

$$r(u^{\mathrm{BIDR}},\boldsymbol{\theta}) = E_{\hat{\boldsymbol{\theta}}}\left\{l(u^{\mathrm{BIDR}},\boldsymbol{\theta})\right\} = E_{\mathrm{v}}\left\{l^{\#}(\eta^{*},\mathbf{V})\right\} \tag{6}$$

does not depend on $\boldsymbol{\theta}$.

Consider now a situation described by one of a family of density functions $f(x|\mu,\sigma)$ indexed by the vector parameter $\boldsymbol{\theta}=(\mu,\sigma)$, where μ and σ (>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 $\mathbf{X}=(X_1, \ldots, X_n)$) a sufficient statistic (M,S) for (μ,σ) with density function $h(m,s|\mu,\sigma)$ of the form

$$h(m,s \mid \mu,\sigma) = \sigma^{-2} h_{\bullet}[(m-\mu)/\sigma, s/\sigma] \tag{7}$$

such that

$$h(m,s \mid \mu,\sigma) dm ds = h_{\bullet}(v_1, v_2) dv_1 dv_2, \tag{8}$$

where $V_1=(M-\mu)/\sigma$, $V_2=S/\sigma$. We are thus assuming that for the family of density functions an induced invariance holds under the group G of transformations: $m \to am+b$, $s \to as$ (a>0). The family of density functions $f(x|\mu,\sigma)$ 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.

Below, we give some applications of the invariant embedding technique.

III. Optimization of Statistical Decisions for Newsboy Problem

The classical newsboy problem is reflective of many real life situations and is often used to aid decision-making in the fashion and sporting industries, both at the manufacturing and retail levels (Gallego and Moon [7]). The newsboy problem can also be used in managing capacity and evaluating advanced booking of orders in service industries such as airlines and hotels (Weatherford and Pfeifer [16]). A partial review of the newsboy problem literature has been recently conducted in a textbook by Silver *et al*. [17]. Researchers have followed two approaches to solving the newsboy problems. In the first approach, the expected costs of overestimating and underestimating demand are minimized. In the second approach, the expected profit is maximized. Both approaches yield the same results. We use the first approach in stating the newsboy problem.

For product j, define:

- X_j quantity demanded during the period, a random variable,
- $f_j(x_j|\mu_j,\sigma_j)$ the probability density function of X_j,
- $\boldsymbol{\theta}_j=(\mu_j,\sigma_j)$ the parameter of $f_j(x_j|\mu_j,\sigma_j)$,
- $F_j(x_j|\mu_j,\sigma_j)$ the cumulative distribution function of X_j,
- $c_j^{(1)}$ overage (excess) cost per unit,
- $c_j^{(2)}$ underage (shortage) cost per unit,
- d_j inventory/order quantity, a decision variable.

The cost per period is

$$l_j^{\circ}(d_j, X_j) = \begin{cases} c_j^{(1)}(d_j - X_j), & \text{if } X_j < d_j, \\ c_j^{(2)}(X_j - d_j), & \text{if } X_j \ge d_j. \end{cases} \tag{9}$$

Complete Information. A standard newsboy formulation (see, e.g., Nahmias [18]) is to consider each product j's cost function:

$$l_j^{+}(d_j,\boldsymbol{\theta}_j) = c_j^{(1)} \int_{-\infty}^{d_j} (d_j - x_j) f_j(x_j \mid \mu_j,\sigma_j) dx_j$$

$$+ c_j^{(2)} \int_{d_j}^{\infty} (x_j - d_j) f_j(x_j \mid \mu_j,\sigma_j) dx_j. \tag{10}$$

Expanding (10) gives

$$l_j^{+}(d_j,\boldsymbol{\theta}_j) = -c_j^{(1)} \int_{-\infty}^{d_j} x_j f_j(x_j \mid \mu_j,\sigma_j) dx_j$$

$$+ c_j^{(2)} \int_{d_j}^{\infty} x_j f_j(x_j \mid \mu_j,\sigma_j) dx_j$$

$$+ (c_j^{(1)} + c_j^{(2)}) d_j \left(F_j(d_j \mid \mu_j,\sigma_j) - \frac{c_j^{(2)}}{c_j^{(1)} + c_j^{(2)}} \right). \tag{11}$$

Let the superscript * denote optimality. Using Leibniz's rule to obtain the first and second derivatives shows that $l_j^{+}(d_j \mid \boldsymbol{\theta}_j)$ is concave. The sufficient optimality condition is the well-known fractile formula:

$$F_j(d_j^* \mid \mu_j,\sigma_j) = \frac{c_j^{(2)}}{c_j^{(1)} + c_j^{(2)}}. \tag{12}$$

It follows from (12) that

$$d_j^* = F_j^{-1}\left(\frac{c_j^{(2)}}{c_j^{(1)} + c_j^{(2)}} \mid \mu_j,\sigma_j \right). \tag{13}$$

At optimality, substituting (12) into the last (bracketed) term in Eq. (11) gives

$$(c_j^{(1)} + c_j^{(2)}) d_j^* \left(F_j(d_j^* \mid \mu_j,\sigma_j) - \frac{c^{(2)}}{c^{(1)} + c^{(2)}} \right) = 0. \tag{14}$$

Hence (11) reduces to

$$l_j^{+}(d_j^*,\boldsymbol{\theta}_j) = c_j^{(2)} E_{x_j}\{X_j\} - (c_j^{(1)} + c_j^{(2)}) \int_{-\infty}^{d_j^*} x_j f_j(x_j \mid \mu_j,\sigma_j) dx_j. \tag{15}$$

Parametric Uncertainty. Let us assume that the functional form of the probability density function $f_j(x_j|\mu_j,\sigma_j)$ is specified but its parameter $\boldsymbol{\theta}=(\mu_j,\sigma_j)$ is not specified. Let $\mathbf{X}_j=(X_{j1}, \ldots, X_{jn})$ be a random sample of observations on a continuous random variable X_j. We shall assume that there is obtainable from a random sample of observations $\mathbf{X}_j=(X_{j1}, \ldots, X_{jn})$ a sufficient statistic (M_j,S_j) for $\boldsymbol{\theta}=(\mu_j,\sigma_j)$ with density function of the form (7),

$$h_j(m_j,s_j \mid \mu_j,\sigma_j) = \sigma_j^{-2} h_{\bullet j}[(m_j - \mu_j)/\sigma_j, s_j/\sigma_j], \tag{16}$$

and with

$$h_j(m_j, s_j \mid \mu_j, \sigma_j)dm_j ds_j = h_{\bullet}(v_{1j}, v_{2j})dv_{1j}dv_{2j}, \quad (17)$$

where $V_{1j}=(M_j-\mu_j)/\sigma_j$, $V_{2j}=S_j/\sigma_j$.

Using the above-mentioned invariant embedding technique, we transform (10) as follows:

$$l_j^+(d_j, \mathbf{\theta}_j) = \omega_j(\sigma_j) l_j^\#(\eta_j, \mathbf{V}_j), \quad (18)$$

where $\omega_j(\sigma_j)=\sigma_j$,

$$l_j^\#(\eta_j, \mathbf{V}_j) = c_j^{(1)} \int_{-\infty}^{\eta_j V_{2j}+V_{1j}} (\eta_j V_{2j} + V_{1j} - z_j) f_j(z_j) dz_j$$

$$+ c_j^{(2)} \int_{\eta_j V_{2j}+V_{1j}}^{\infty} \left(z_j - \eta_j V_{2j} - V_{1j}\right) f_j(z_j) dz_j, \quad (19)$$

$Z_j=(X_j-\mu_j)/\sigma_j$ is a pivotal quantity, $f_j(z_j)$ is defined by $f_j(x_j|\mu_j,\sigma_j)$, i.e.,

$$f_j(z_j)dz_j = f_j(x_j|\mu_j,\sigma_j)dx_j, \quad (20)$$

$\mathbf{V}_j=(V_{1j},V_{2j})$ is a pivotal quantity, $\eta_j=(d_j-M_j)/S_j$ is an ancillary factor. It follows from (18) that the risk associated with u_j^{BIDR} (or η_j^*) can be expressed as

$$r_j^+(u_j^{\text{BIDR}}, \mathbf{\theta}_j) = E_{m_j,s_j}\left\{l_j^+(u_j^{\text{BIDR}}, \mathbf{\theta}_j)\right\}$$

$$= \omega_j(\sigma_j) E_{\mathbf{V}_j}\left\{l_j^\#(\eta_j^*, \mathbf{V}_j)\right\}, \quad (21)$$

where

$$u_j^{\text{BIDR}} \equiv d_j^* = M_j + \eta_j^* S_j, \quad (22)$$

$$\eta_j^* = \arg\min_{\eta_j} E_{\mathbf{V}_j}\left\{l_j^\#(\eta_j, \mathbf{V}_j)\right\}, \quad (23)$$

$$E_{\mathbf{V}_j}\left\{l_j^\#(\eta_j, \mathbf{V}_j)\right\} = \iint_{v_{1j},v_{2j}} l_j^\#(\eta_j; v_{1j}, v_{2j}) h_{\bullet j}(v_{1j}, v_{2j}) dv_{1j} dv_{2j}. \quad (24)$$

The fact that (24) is independent of $\mathbf{\theta}_j$ means that an ancillary factor η_j^*, which minimizes (24), is uniformly best invariant. Thus, d_j^* given by (22) is the best invariant decision rule.

A. Numerical Example

Complete Information. Assuming that the demand for product *j*, X_j, is exponentially distributed with the probability density function,

$$f_j(x_j|\sigma_j)=(1/\sigma_j)\exp(-x_j/\sigma_j) \ \ (x_j>0), \quad (25)$$

it follows from (10), (13) and (15) that

$$l_j^+(d_j, \sigma_j) = c_j^{(1)}(d_j - \sigma_j) + (c_j^{(1)} + c_j^{(2)})\sigma_j \exp\left(-\frac{d_j}{\sigma_j}\right), \quad (26)$$

$$d_j^* = \sigma_j \ln\left(1 + \frac{c_j^{(2)}}{c_j^{(1)}}\right), \quad (27)$$

and

$$l_j^+(d_j^*, \sigma_j) = c_j^{(1)} \sigma_j \ln\left(1 + \frac{c_j^{(2)}}{c_j^{(1)}}\right), \quad (28)$$

respectively.

Parametric Uncertainty. Consider the case when the parameter σ_j is unknown. Let $\mathbf{X}_j=(X_{j1}, \ldots, X_{jn})$ be a random sample of observations (each with density function (25)) on a continuous random variable X_j. Then

$$S_j = \sum_{i=1}^{n} X_{ji}, \quad (29)$$

is a sufficient statistic for σ_j; S_j is distributed with

$$h_j(s_j \mid \sigma_j) = \frac{1}{\Gamma(n)\sigma_j^n} s_j^{n-1} \exp\left(-\frac{s_j}{\sigma_j}\right) \quad (s_j > 0), \quad (30)$$

so that

$$h_{\bullet j}(v_{2j}) = \frac{1}{\Gamma(n)} v_{2j}^{n-1} e^{-v_{2j}} \quad (v_{2j}>0). \quad (31)$$

It follows from (21) and (26) that

$$r_j^+(u_j^{\text{BIDR}}, \sigma_j) = E_{s_j}\left\{l_j^+(u_j^{\text{BIDR}}, \sigma_j)\right\}$$

$$= \sigma_j \int_0^{\infty} l_j^\#(\eta_j^*, v_{2j}) h_{\bullet j}(v_{2j}) dv_{2j}$$

$$= \sigma_j \left[c_j^{(1)}(n\eta_j^* - 1) + \frac{c_j^{(1)} + c_j^{(2)}}{(1+\eta_j^*)^n}\right], \quad (32)$$

where

$$u_j^{\text{BIDR}} = \eta_j^* S_j, \quad (33)$$

$$\eta_j^* = \arg\min_{\eta_j} \sigma_j \left[c_j^{(1)}(n\eta_j - 1) + \frac{c_j^{(1)} + c_j^{(2)}}{(1+\eta_j)^n}\right]$$

$$= \left[1 + \frac{c_j^{(2)}}{c_j^{(1)}}\right]^{1/(n+1)} - 1. \quad (34)$$

Comparison of Decision Rules. For comparison, consider the maximum likelihood decision rule (MLDR) that may be obtained from (27),

$$u_j^{\text{MLDR}} = \hat{\sigma}_j \ln\left(1 + \frac{c_j^{(2)}}{c_j^{(1)}}\right) = \eta_j^{\text{MLDR}} S_j, \quad (35)$$

where $\hat{\sigma}_j = S_j/n$ is the maximum likelihood estimator of σ_j. Since u_j^{BIDR} and u_j^{MLDR} belong to the same class

$$\mathscr{C} = \{u_j : u_j = \eta_j S_j\}, \quad (36)$$

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

it follows from the above that u_j^{MLDR} is inadmissible in relation to u_j^{BIDR}.

If, say, n=1 and $c_j^{(2)}/c_j^{(1)}$=100, we have that

$$\text{rel.eff.}_{r_j^+}\{(u_j^{\text{MLDR}}, u_j^{\text{BIDR}}, \sigma_j\}$$

$$= r_j^+(u_j^{\text{BIDR}}, \sigma_j) \Big/ r_j^+(u_j^{\text{MLDR}}, \sigma_j)$$

$$= \left(n\eta_j^* - 1 + \frac{1 + c_j^{(2)}/c_j^{(1)}}{(1+\eta_j^*)^n} \right) \left(n\eta_j^{\text{MLDR}} - 1 + \frac{1 + c_j^{(2)}/c_j^{(1)}}{(1+\eta_j^{\text{MLDR}})^n} \right)^{-1}$$

$$=0.838. \tag{37}$$

Thus, in this case, the use of u_j^{BIDR} leads to a reduction in the risk of about 16.2 % as compared with u_j^{MLDR}. The absolute risk will be proportional to σ_j and may be considerable.

IV. Shortest-Length Prediction Interval

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

$$f(x;\boldsymbol{\theta}) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), \quad x \ge \mu, \tag{38}$$

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

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

$$S_r=(Y_{(r)}-\mu)/\sigma, \quad S_1=(X_{(1)}-\mu)/\sigma \tag{39}$$

and

$$T_1=T/\sigma, \tag{40}$$

where

$$T = \sum_{i=1}^{k}(X_{(i)} - X_{(1)}) + (n-k)(X_{(k)} - X_{(1)}). \tag{41}$$

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

$$V = n(S_r-S_1)/T_1 = n(Y_{(r)}-X_{(1)})/T. \tag{42}$$

It is well known [19] 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 [20],

$$f(s_r) = r\binom{m}{r}(1-e^{-s_r})^{r-1}e^{-s_r(m-r+1)}, \tag{43}$$

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

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

$$= \begin{cases} 1 - nr\dbinom{m}{r}\displaystyle\sum_{j=0}^{r-1}\frac{\dbinom{r-1}{j}(-1)^j[1+v(m-r+j+1)/n]^{-k+1}}{(m+n-r+j+1)(m-r+j+1)}, & v > 0, \\ m^{(r)}(1-v)^{-k+1}/(m+n)^{(r)}, & v \le 0, \end{cases} \tag{44}$$

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

The special case in which r=1 is worth mentioning, since in this case (44) simplifies somewhat. We find here that we can write

$$F(v) = \Pr\{V \le v\} = \begin{cases} 1 - \dfrac{\vartheta}{\vartheta+1}\left(\dfrac{\vartheta}{\vartheta+v}\right)^{k-1}, & v > 0, \\ (\vartheta+1)^{-1}(1-v)^{-k+1}, & v \le 0, \end{cases} \tag{45}$$

where $\vartheta=n/m$.

Consider the ordered data given by Grubbs [21] 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

$$T = \sum_{i=1}^{19}(X_{(i)} - X_{(1)}) = 15869 \tag{46}$$

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

$$V = \frac{n(Y_{(1)} - X_{(1)})}{T}. \tag{47}$$

Then

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

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

$$= \Pr\left\{ z_L < Y_{(1)} < z_U \right\} = 1-\alpha, \tag{48}$$

where

$$z_L = X_{(1)} + v_1T/n \quad\text{and}\quad z_U = X_{(1)} + v_2T/n. \tag{49}$$

The length of the prediction interval is

$$\Delta_z = z_U - z_L = (T/n)(v_2 - v_1). \tag{50}$$

We wish to minimize Δ_z subject to

$$F(v_2)-F(v_1)=1-\alpha. \tag{51}$$

It can be shown that the minimum occurs when

$$f(v_1)=f(v_2), \tag{52}$$

where v_1 and v_2 satisfy (51).

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The shortest-length prediction interval is given by

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

$$=(10.78,\ 736.62), \qquad (53)$$

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−α=0.95 confidence level is given by

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

$$=(57.6,\ 834.34), \qquad (54)$$

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.

Comparison of Prediction Intervals. 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

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

$$=\frac{\Delta_z^*}{\Delta_z^\circ}=\frac{v_2^*-v_1^*}{v_{1-\alpha/2}-v_{\alpha/2}}=0.934. \qquad (55)$$

V. Shortest-Length Confidence Interval for System Availability

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

$$f_1(x_1;\theta_1)=\frac{1}{\theta_1}e^{-x_1/\theta_1},\quad x_1\in(0,\infty),\ \ \theta_1>0, \qquad (56)$$

and

$$f_2(x_2;\theta_2)=\frac{1}{\theta_2}e^{-x_2/\theta_2},\quad x_2\in(0,\infty),\ \ \theta_2>0. \qquad (57)$$

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

$$A=\theta_1/(\theta_1+\theta_2), \qquad (58)$$

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 (56) and (57) with sample means

$$\overline{X}_1=\sum_{i=1}^{n_1}X_{1i}/n_1,\quad \overline{X}_2=\sum_{i=1}^{n_2}X_{2i}/n_2. \qquad (59)$$

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 (58) that

$$\frac{A}{1-A}=\frac{\theta_1}{\theta_2}. \qquad (60)$$

Using the invariant embedding technique, we obtain from (60) a pivotal quantity

$$V(S,A)=S\frac{A}{1-A}=\frac{\overline{X}_2}{\overline{X}_1}\frac{\theta_1}{\theta_2}=\left(\frac{2n_2\overline{X}_2/\theta_2}{2n_2}\right)\Bigg/\left(\frac{2n_1\overline{X}_1/\theta_1}{2n_1}\right), \qquad (61)$$

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

$$S=\overline{X}_2/\overline{X}_1. \qquad (62)$$

Thus, (61) allows one to find a 100(1−α)% confidence interval for A from

$$\Pr\{A_L<A<A_U\}=1-\alpha, \qquad (63)$$

where

$$A_L=\frac{v_L}{v_L+S}\quad\text{and}\quad A_U=\frac{v_U}{v_U+S}. \qquad (64)$$

It can be shown that the shortest-length confidence interval for A is given by

$$C_A^*=(A_L,A_U) \qquad (65)$$

with

$$\Delta^*(S,v_L,v_U)=A_U-A_L, \qquad (66)$$

where v_L and v_U are a solution of

$$(v_L+S)^2f(v_L)=(v_U+S)^2f(v_U) \qquad (67)$$

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

$$\Pr\{v_L<V<v_U\}=\Pr\{v_L<F(2n_2,2n_1)<v_U\}=1-\alpha. \qquad (68)$$

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

$$C_A=(A_L,A_U)=\left(\frac{v_L}{v_L+S},\frac{v_U}{v_U+S}\right) \qquad (69)$$

with

$$\Delta(S,v_L,v_U)=A_U-A_L, \qquad (70)$$

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

is employed, where

$$v_{\mathrm{L}} = F_{\alpha/2}(2n_2, 2n_1), \quad v_{\mathrm{U}} = F_{1-\alpha/2}(2n_2, 2n_1), \qquad (71)$$

and

$$\Pr\{F(2n_2,2n_1) > F_{\alpha/2}(2n_2,2n_1)\} \le 1-\alpha/2. \qquad (72)$$

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 $\widehat{A} = 200/(200+20) = 0.909$, and the confidence interval for the inherent availability, at the 90% confidence level, is found as follows.

From (69), the simpler equal tails confidence interval is

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

$$= (0.61, 0.985), \qquad (73)$$

i.e.,

$$\Delta(S, v_{\mathrm{L}}, v_{\mathrm{U}}) = A_{\mathrm{U}} - A_{\mathrm{L}} = 0.375. \qquad (74)$$

From (65), the shortest-length confidence interval is

$$C_A^* = \left(\frac{v_{\mathrm{L}}}{v_{\mathrm{L}} + S}, \frac{v_{\mathrm{U}}}{v_{\mathrm{U}} + S} \right) = (0.707, 0.998), \qquad (75)$$

where v_{L} and v_{U} are a solution of (67) and (68). Thus,

$$\Delta^*(S, v_{\mathrm{L}}, v_{\mathrm{U}}) = A_{\mathrm{U}} - A_{\mathrm{L}} = 0.291. \qquad (76)$$

Comparison of Confidence Intervals for Availability. The relative efficiency of C_{A} relative to C_A^* is given by

$$\text{rel.eff.}_{\mathrm{C}}\left(C_A, C_A^*\right) = \frac{\Delta^*(S, v_L, v_U)}{\Delta(S, v_L, v_U)} = \frac{0.291}{0.375} = 0.776. \qquad (77)$$

VI. Conclusions and Directions for Future Research

In this paper, we propose a new technique to improve or optimize statistical decisions under parametric uncertainty. The method used is that of the invariant embedding of sample statistics in a performance index in order to form pivotal quantities, which make it possible to eliminate unknown parameters (i.e., parametric uncertainty) from the problem. It is especially efficient when we deal with asymmetric performance indexes and small data samples.

More work is needed, however, to obtain improved or optimal 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 multiparametric or multidimensional distributions, (v) the observations are from truncated distributions, (vi) the observations are censored, (vii) the censored observations are from truncated distributions.

References

[1] S. A. Conrad, "Sales data and the estimation of demand," *Oper. Res. Quart.*, vol. 27, 1976, pp. 123–127.

[2] L. H. Liyanage, and J. G. Shanthikumar, "A practical inventory policy using operational statistics," *Operations Research Letters*, vol. 33, 2005, pp. 341–348.

[3] H. Scarf, "Bayes solutions of statistical inventory problem," *Ann. Math. Statist*., vol. 30, 1959, pp. 490–508.

[4] L. Y. Chu, J. G. Shanthikumar, and Z. J. M Shen, "Solving operational statistics via a Bayesian analysis," *Operations Research Letters*, vol. 36, 2008, pp. 110–116.

[5] J. H. Bookbinder, and A. E. Lordahl, "Estimation of inventory reorder level using the bootstrap statistical procedure," *IIE Trans*., vol. 21, 1989, pp. 302–312.

[6] H. Scarf, "A min–max solution of an inventory problem. studies in the mathematical theory of inventory and production (Chapter 12)," Stanford, CA: Stanford University Press, 1958.

[7] G. Gallego, and I. Moon, "The distribution free newsvendor problem: review and extensions," *J. Oper. Res. Soc*., vol. 44, 1993, pp. 825–834.

[8] N. A. Nechval, K. N. Nechval, and E. K. Vasermanis, "Optimization of interval estimators via invariant embedding technique," *IJCAS (The International Journal of Computing Anticipatory Systems)*, vol. 9, 2001, pp. 241–255.

[9] N. A. Nechval, K. N. Nechval, and E. K. Vasermanis, "Effective state estimation of stochastic systems," *Kybernetes*, vol. 32, 2003, pp. 666–678.

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

[11] N. A. Nechval, and K. N. Nechval, "Constrained optimization in newsboy problems under uncertainty via statistical inference equivalence principle," in *Proceedings of the 12th International Conference on Analytical and Stochastic Modelling Techniques and Applications (ASMTA'05)*, K. Al-Begain, G. Bolch, and M. Telek, Eds. Riga, Latvia: RTU, 2005, pp. 166–171.

[12] K. N. Nechval, N. A. Nechval, G. Berzins, M. Purgailis, and U. Rozevskis, "Dual control of education process," in *Computing Anticipatory Systems*, AIP Conference Proceedings, vol. 1051, D. M. Dubois, Ed. Melville, New York: American Institute of Physics, 2008, pp. 366–379.

[13] N. A. Nechval, G. Berzins, M. Purgailis, and K. N. Nechval, "Improved estimation of state of stochastic systems via invariant embedding technique," *WSEAS Transactions on Mathematics*, vol. 7, 2008, pp. 141–159.

[14] N. A. Nechval, K. N. Nechval, G. Berzinsh, M. Purgailis, and U. Rozevskis, "Stochastic fatigue models for efficient planning inspections in service of aircraft structures," in *Analytical and Stochastic Modeling Techniques and Applications*, LNCS, vol. 5055, K. Al-Begain, A. Heindl, and M. Telek, Eds. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 114–127.

[15] N. A. Nechval, G. Berzins, M. Purgailis, K. N. Nechval, and N. Zolova, "Improved adaptive control of stochastic systems," *Advances in Systems Science and Applications*, vol. 9, 2009, pp. 11–20.

[16] L. R. Weatherford, and P. E. Pfeifer, "The economic value of using advance booking of orders," *Omega*, vol. 22, 1994, pp. 105–111.

[17] E. A. Silver, D. F. Pyke, and R. P. Peterson, *Inventory Management and Production Planning and Scheduling*. New York: John Wiley, 1998.

[18] S. Nahmias, *Production and Operations Management*. Boston: Irwin, 1996.

[19] B. Epstein, and M. Sobel, "Some theorems relevant to life testing from an exponential population," *Ann. Math. Statist*., vol. 25, 1954, pp. 373–381.

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

[21] F. E. Grubbs, "Approximate fiducial bounds on reliability for the two parameter negative exponential distribution," *Technometrics*, vol. 13, 1971, pp. 873–876.

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)