Optimal Prediction Intervals for Order Statistics Coming from Location-Scale Families

Nicholas A. Nechval, *Member, IAENG*, Konstantin N. Nechval, Maris Purgailis, and Uldis Rozevskis

***Abstract*—Prediction, by interval or point, of an unobserved random variable is a fundamental problem in statistics. This paper deals with obtaining a prediction interval on a future observation X_l in an ordered sample of size *m* from an underlying distribution, which belongs to the location-scale family of distributions, for the situation where the first *k* observations $X_1 < X_2 < \cdots < X_k$, $1 \le k < l \le m$, have been observed. Prediction intervals for future order statistics are widely used for reliability problems and other related problems. But the optimality property of these intervals has not been fully explored. To compare prediction intervals, we introduce a piecewise-linear loss function. The interval which minimizes a risk, associated with this piecewise-linear loss function, among the class of invariant prediction intervals is obtained. The technique used here for optimization of prediction intervals based on censored data emphasizes pivotal quantities relevant for obtaining ancillary statistics. It allows one to solve the optimization problems in a simple way. Illustrative examples are given for the Gumbel and two-parameter exponential distributions. The results can be also applied to related distributions.**

***Index Terms* — Location-scale distribution, future order statistic, prediction interval, risk function, optimization**

I. Introduction

TRADITIONAL statistical analysis uses the information contained in a sample to make inferences about the population where this sample was taken from. Usually, these inferences are based on estimates, confidence intervals or hypothesis tests for parameters of a specified model. If this model describes adequately the population, the analyses containing these inferences are appropriate for most scientific problems. On the other hand, we also encounter problems where the understanding of the population's behavior is not of interest by itself; it is a means of foretelling future events. We call such problems prediction problems. Since prediction problems are rarely a case of logical deduction, the use of probabilistic and statistical tools are inevitable in any scientific approach used to solve them. Many early papers deal with prediction; for example Pearson [1], Baker [2], or Wilks [3]. However, even if prediction problems are often encountered, statisticians are now devoting most of their attention to inferential problems. An interesting discussion about the neglect of prediction analysis can be found in the preface of Aitchison and Dunsmore's [4] book and this issue is also discussed in Geisser [5]. Antle and Rademaker [6] provided a method of obtaining a prediction bound for the largest observation from a future sample of the Type I extreme value distribution, based on the maximum likelihood estimates of the parameters. They used Monte Carlo simulations to obtain the prediction intervals. Using the well-known relationship between the Weibull distribution and the Type I extreme value distribution one can use their method to construct an upper prediction limit for the largest among a set of future Weibull observations. However this method is valid only for complete samples and limited to constructing an upper prediction limit for the largest among a set of future observations. Lawless [7] proposed a method for constructing prediction intervals for the smallest ordered observation among a set of *k* future observations based on a Type II censored sample of past observations. These results are based on the conditional distribution of the maximum likelihood estimates given a set of ancillary statistics. This procedure is exact, but it requires numerical integration, for each new sample obtained, to determine the prediction bounds. Mee and Kushary [8] provided a simulation based procedure for constructing prediction intervals for Weibull populations for Type II censored case. This procedure is based on maximum likelihood estimation and requires an iterative process to determine the percentile points.

Statistical prediction can be applied in many domains such as engineering, industry, business, and medicine. In each of these domains, it can be used for planning purposes (predict the total medical cost of a population, predict a future number of insurance claims), for process monitoring (predict the number of nuclear scrams in a power plant), or for decision making (software debugging, determination of a maintenance policy).

Although confidence intervals for a mean are often relatively easy to compute and interpret, Christoffersen [9] argues that interval prediction is a better tool than interval parameter estimation for economic planning. For example, a central bank governor would be more interested in forecasting the actual inflation rate over the next six months than estimating its mean inflation in order to carry out a

Manuscript received October 31, 2012. 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.

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

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

Maris Purgailis is with the Cybernetics Department, University of Latvia, Riga LV-1050, Latvia (e-mail: marispur@lanet.lv).

Uldis Rozevskis is with the Informatics Department, University of Latvia, Riga LV-1050, Latvia (e-mail: uldis.rozevskis@lu.lv).

monetary policy. A production manager planning to purchase inventory needs to predict sales in order to decide how much to order. A prediction interval might serve as a control bound for assessing product quality on an assembly line. Data points that are beyond the control bounds are referred to as being “out of control” and could indicate that remedial action needs to be taken. Patel [10] also states that prediction intervals could provide guidelines for establishing warranty limits for the future performance of a product.

This paper deals with frequentist prediction. Frequentist probability interpretations of the methods considered are clear, and analogous to the interpretation of confidence intervals. Bayesian methods are not considered further. We note, however, that, although subjective Bayesian prediction has a clear personal probability interpretation, it is not generally clear how this should be applied to non-personal prediction or decisions. Objective Bayesian methods, on the other hand, do not have clear probability interpretations in finite samples. Prediction methodology based on likelihood has also been proposed. We similarly do not consider these methods in this paper, but see Bjornstad [11] for a review.

The purpose of this paper is to provide a unified treatment of frequentist prediction intervals and predictive distributions.

Consider the following examples of practical problems which often require the computation of prediction bounds and prediction intervals for future values of random quantities: (i) a consumer purchasing a refrigerator would like to have a lower bound for the failure time of the unit to be purchased (with less interest in distribution of the population of units purchased by other consumers); (ii) financial managers in manufacturing companies need upper prediction bounds on future warranty costs; (iii) when planning life tests, engineers may need to predict the number of failures that will occur by the end of the test to predict the amount of time that it will be take for a specified number of units to fail. Some applications require a two-sided prediction interval that will, with a specified high degree of confidence, contain the future random variable of interest. In many applications, however, interest is focused on either an upper prediction bound or a lower prediction bound (e.g., the maximum warranty cost is more important than the minimum, and the time of the early failures in a product population is more important than the last ones). Conceptually, it is useful to distinguish between ‘new-sample’ prediction and ‘within-sample’ prediction. For new-sample prediction, data from a past sample are used to make predictions on a future unit or sample of units from the same process or population. For example, based on previous (possibly censored) life test data, one could be interested in predicting the time to failure of a new unit, time until r failures in a future sample of m units, or number of failures by time $t^{\bullet}$ in a future sample of m units. For within-sample prediction, the problem is to predict future events in a sample or process based on early data from that sample or process. If, for example, m units are followed until $t_{\bullet}$ and there are k observable failures, $X_1 < X_2 < \cdots < X_k$, one could be interested in predicting the time of the next failure, $X_{(k+1)}$; time until r additional failures, $X_{(k+r)}$; number of additional failures in a future interval $(t_{\bullet}, t^{\bullet})$. In general, to predict a future realization of a random quantity one needs the following:

1) *A statistical model to describe the population or process of interest*. This model usually consists of a distribution depending on a vector of parameters θ. In this paper, attention is restricted to location-scale families of distributions which are invariant under location and/or scale changes. In particular, the case may be considered where a previously available complete or type II censored sample is from a continuous distribution with cdf $F((x\text{-}\mu)/\sigma)$, where $F(\cdot)$ is known but the location (μ) and/or scale (σ) parameters are unknown. For such family of distributions the decision problem remains invariant under a group of transformations (a subgroup of the full affine group) which takes μ (the location parameter) and σ (the scale) into $c\mu + b$ and $c\sigma$, respectively, where b lies in the range of μ, $c > 0$. This group acts transitively on the parameter space.

2) *Information on the values of components of the parametric vector θ*. It is assumed that only the functional form of the distribution is specified, but some or all of its parameters are unspecified. In such cases ancillary statistics and pivotal quantities, whose distribution does not depend on the unknown parameters, are used.

The technique used here for constructing prediction intervals (or bounds) emphasizes pivotal quantities relevant for obtaining ancillary statistics. It represents a simple procedure that can be utilized by non-statisticians, and which provides easily computable explicit expressions for both prediction bounds and prediction intervals. The technique is a special case of the method of invariant embedding of sample statistics into a performance index (see, e.g., Nechval et al. [12-24]) applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space.

In this paper, for the within-sample prediction situation, we obtain optimal prediction intervals for future order statistics under parametric uncertainty of the underlying distributions.

II. Problem Statement of Within – Sample Prediction Intervals for Future Order Statistics

A. Piecewise-Linear Loss Function

Consider a situation described by one of a location-scale family of density functions, indexed by the vector parameter $\theta = (\mu,\sigma)$, 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 maximum likelihood estimator $(\widehat{\mu},\widehat{\sigma})$ (or sufficient statistic) for (μ,σ) with density function of the form

$$p_{\theta}(\widehat{\mu},\widehat{\sigma}) = \frac{1}{\sigma^2} f\left(\frac{\widehat{\mu}-\mu}{\sigma}, \frac{\widehat{\sigma}}{\sigma}\right). \qquad (1)$$

We are thus assuming that for the family of density functions an induced invariance holds under the group $\mathscr{G}$ of transformations:

$$\hat{\mu} \to a\hat{\mu} + b, \;\; \hat{\sigma} \to a\hat{\sigma} \;\; (a > 0). \tag{2}$$

The family of density functions satisfying these conditions is, of course, the limited one of normal, negative exponential and gamma (with known index) density functions. The structure of the problem is, however, more clearly seen within the general framework.

We shall consider the interval prediction problem for the lth order statistic X_l, $k<l\leq m$, in the same sample of size m for the situation where the first k observations $X_1 < X_2 < \cdots < X_k$, $1\leq k<m$, have been observed. Suppose that we assert that an interval (d_1,d_2) contains X_l. If, as is usually the case, the purpose of this interval statement is to convey useful information we incur penalties if d_1 lies above X_l or if d_2 falls below X_l. Suppose that these penalties are $c_1(d_1- X_l)$ and $c_2(X_l-d_2)$, losses proportional to the amounts by which X_l escapes the interval. Since c_1 and c_2 may be different the possibility of differential losses associated with the interval overshooting and undershooting the true μ is allowed. In addition to these losses there will be a cost attaching to the length of interval used. For example, it will be more difficult and more expensive to design or plan when the interval (d_1,d_2) is wide. Suppose that the cost associated with the interval is proportional to its length, say $c(d_2-d_1)$. In the specification of the loss function, σ is clearly a 'nuisance parameter' and no alteration to the basic decision problem is caused by multiplying all loss factors by $1/\sigma$. Thus we are led to investigate the piecewise-linear loss function

$$r(\theta,d_1,d_2) = \begin{cases} \dfrac{c_1(d_1 - X_l)}{\sigma} + \dfrac{c(d_2 - d_1)}{\sigma} & (X_l < d_1), \\ \dfrac{c(d_2 - d_1)}{\sigma} & (d_1 \leq X_l \leq d_2), \\ \dfrac{c(d_2 - d_1)}{\sigma} + \dfrac{c_2(X_l - d_2)}{\sigma} & (X_l > d_2). \end{cases} \tag{3}$$

The decision problem specified by the informative experiment probability distribution function $F((x-\mu)/\sigma)$ and the loss function (3) is invariant under the group of transformations, which takes μ (the location parameter) and σ (the scale) into $c\mu + b$ and $c\sigma$, respectively, where b lies in the range of μ, $c > 0$. This group acts transitively on the parameter space. Thus, the problem is to find the best invariant interval predictor of X_l,

$$(d_1^*, d_2^*) = \arg \min_{(d_1,d_2)\in\mathscr{D}} R(\theta,d_1,d_2), \tag{4}$$

where $\mathscr{D}$ is a set of invariant interval predictors of X_l, $R(\theta,d_1,d_2)=E_\theta\{r(\theta, d_1,d_2)\}$ is a risk function.

B. Transformation of the Loss Function

It follows from (3) that the invariant loss function, $r(\theta, d_1,d_2)$, can be transformed as follows:

$$r(\theta,d_1,d_2) = \ddot{r}(V_1,V_2,\eta_1,\eta_2), \tag{5}$$

where

$$\ddot{r}(V_1,V_2,\eta_1,\eta_2)$$

$$= \begin{cases} c_1(-V_1 + \eta_1 V_2) + c(\eta_2 - \eta_1)V_2 & (V_1 < \eta_1 V_2), \\ c(\eta_2 - \eta_1)V_2 & (\eta_1 V_2 \leq V_1 \leq \eta_2 V_2), \\ c_2(V_1 - \eta_2 V_2) + c(\eta_2 - \eta_1)V_2 & (V_1 > \eta_2 V_2), \end{cases} \tag{6}$$

(V_1,V_2) is a vector of the pivotal quantities,

$$V_1 = (X_l - X_k)/\sigma, \;\; V_2 = \hat{\sigma}/\sigma \;\; (\text{or } V_2^\circ = (X_k - \hat{\mu})/\sigma \;\;), \tag{7}$$

(η_1,η_2) is a vector of decisions (decision function),

$$\eta_1 = (d_1 - X_k)/\hat{\sigma}, \;\; \eta_2 = (d_2 - X_k)/\hat{\sigma}$$

$$(\text{or } \eta_2 = (d_2 - X_k)/(X_k - \hat{\mu})). \tag{8}$$

C. Risk Function

It follows from (4)-(6) that the risk associated with (d_1,d_2) and θ can be expressed as

$$R(\theta,d_1,d_2) = E_\theta\{r(\theta,d_1,d_2)\} = E\{\ddot{r}(V_1,V_2,\eta_1,\eta_2)\}$$

$$= c_1 \int_0^\infty \int_0^{\eta_1 v_2} (-v_1 + \eta_1 v_2) f(v_1,v_2) dv_1 dv_2$$

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

$$+ c(\eta_2 - \eta_1) \int_0^\infty \int_0^\infty v_2 f(v_1,v_2) dv_1 dv_2 = R(\eta_1,\eta_2), \tag{9}$$

which is constant on orbits when an invariant predictor (decision rule) (d_1,d_2) is used, where $f(v_1,v_2)$ represents the probability density function of the pivotal quantities V_1 and V_2. The fact that the risk is independent of θ means that the decision function (η_1,η_2) which minimizes $R(\eta_1,\eta_2)$ is uniformly best invariant.

D. Risk Minimization and Invariant Prediction Rules

The following theorem gives the central result in this paper.

Theorem 1. Suppose that (U_1,U_2) is a random vector having density function

$$u_2 f(u_1,u_2) \left[\int_0^\infty \int_0^\infty u_2 f(u_1,u_2) du_1 du_2 \right]^{-1} \;\; (u_1, u_2 > 0), \tag{10}$$

where f is defined by $f(v_1,v_2)$, and let Q be the probability distribution function of U_1/U_2.

(i) If $c/c_1+c/c_2<1$ then the optimal invariant linear-loss interval predictor of X_l based on the first k ordered observations (order statistics) in a sample of size m is $(X_k+\eta_1\hat{\sigma}, X_k+\eta_2\hat{\sigma})$, where

$$Q(\eta_1) = c/c_1, \quad Q(\eta_2) = 1 - c/c_2. \tag{11}$$

(ii) If $c/c_1+c/c_2 \geq 1$ then the optimal invariant linear-loss interval predictor of X_l based on the first k ordered observations in a sample of size m degenerates into a point predictor $X_k + \eta_\bullet \hat{\sigma}$, where

$$Q(\eta_\bullet) = c_2/(c_1 + c_2). \tag{12}$$

Proof. From (9),

$$\frac{\partial R(\eta_1,\eta_2)}{\partial \eta_1}$$

$$= c_1 \int_0^\infty \int_0^{\eta_1 v_2} v_2 f(v_1,v_2) dv_1 dv_2 - c \int_0^\infty \int_0^\infty v_2 f(v_1,v_2) dv_1 dv_2$$

$$= \int_0^\infty \int_0^\infty v_2 f(v_1,v_2) dv_1 dv_2 [c_1 Q(\eta_1) - c], \tag{13}$$

and

$$\frac{\partial R(\eta_1,\eta_2)}{\partial \eta_2}$$

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

$$= \int_0^\infty \int_0^\infty v_2 f(v_1,v_2) dv_1 dv_2 [-c_2(1 - Q(\eta_2)) + c], \tag{14}$$

where

$$Q(\eta) = \int_0^\eta q(w) dw, \tag{15}$$

$$q(w) = \frac{\int_0^\infty v_2^2 f(wv_2,v_2) dv_2}{\int_0^\infty \int_0^\infty v_2 f(v_1,v_2) dv_1 dv_2}, \tag{16}$$

$$W = V_1/V_2. \tag{17}$$

Now $\partial R(\eta_1,\eta_2)/\partial \eta_1 = \partial R(\eta_1,\eta_2)/\partial \eta_2 = 0$ if and only if (11) hold. We thus obtain one stationary value for $R(\eta_1,\eta_2)$ provided (11) has a solution with $\eta_1 < \eta_2$ and this is so if $1-c/c_2 > c/c_1$. It is easily confirmed that this (η_1,η_2) gives the minimum value of $R(\eta_1,\eta_2)$. Thus (i) is established.

If $c/c_1+c/c_2 \geq 1$ then the minimum of $R(\eta_1,\eta_2)$ in the region $\eta_2 \geq \eta_1$ occurs where $\eta_1 = \eta_2 = \eta_\bullet$, $\eta_\bullet$ being determined by setting

$$\partial R(\eta_\bullet,\eta_\bullet)/\partial \eta_\bullet = 0 \tag{18}$$

and this reduces to

$$c_1 Q(\eta_\bullet) - c_2[1 - Q(\eta_\bullet)] = 0, \tag{19}$$

which establishes (ii). □

Corollary 1.1. The minimum risk of the optimal invariant predictor of X_l is given by

$$R(\theta, d_1^*, d_2^*) = E_\theta\{r(\theta, d_1^*, d_2^*)\} = E\{\ddot{r}(V_1,V_2,\eta_1,\eta_2)\} = R(\eta_1^*, \eta_2^*)$$

$$= -c_1 \int_0^\infty \int_0^{\eta_1 v_2} v_1 f(v_1,v_2) dv_1 dv_2 + c_2 \int_0^\infty \int_{\eta_2 v_2}^\infty v_1 f(v_1,v_2) dv_1 dv_2 \tag{20}$$

for case (i) with (η_1,η_2) as given by (11) and for case (ii) with $\eta_1 = \eta_2 = \eta_\bullet$ as given by (12).

Proof. These results are immediate from (9) when use is made of $\partial R(\eta_1,\eta_2)/\partial \eta_1 = \partial R(\eta_1,\eta_2)/\partial \eta_2 = 0$ in case (i) and $\partial R(\eta_\bullet,\eta_\bullet)/\partial \eta_\bullet = 0$ in case (ii). □

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

III. Equivalent Confidence Coefficient

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

Theorem 2. Suppose that (V_1,V_2) is a random vector having density function $f(v_1,v_2)$ $(v_1,v_2>0)$ and let H be the distribution function of $W=V_1/V_2$, i.e., the probability density function of W is given by

$$h(w) = \int_0^\infty v_2 f(wv_2,v_2) dv_2. \tag{21}$$

Then the confidence coefficient based on the first k ordered observations in a sample of size m and associated with the optimum prediction interval (d_1,d_2), where $d_1 = X_k + \eta_1 \hat{\sigma}$, $d_2 = X_k + \eta_2 \hat{\sigma}$, is

$$\Pr\{(d_1,d_2): d_1 < X_l < d_2 \mid \mu,\sigma\}$$

$$= H[Q^{-1}(1-c/c_2)] - H[Q^{-1}(c/c_1)]. \tag{22}$$

Proof. The confidence coefficient for (d_1,d_2) corresponding to (μ,σ) is given by

$$\Pr\{(X_k,\hat{\sigma}): X_k + \eta_1\hat{\sigma} < X_l < X_k + \eta_2\hat{\sigma} \mid \mu,\sigma\}$$

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

$$= H(\eta_2) - H(\eta_1) = H[Q^{-1}(1-c/c_2)] - H[Q^{-1}(c/c_1)]. \tag{23}$$

This is independent of (μ,σ). □

The way in which (23) varies with c, c_1 and c_2, and the fact that c_1 and c_2 are the factors of proportionality associated with losses from overshooting and undershooting relative to loss involved in increasing the length of interval, provides an interesting interpretation of confidence interval prediction.

IV. Finding Joint Distributions of Pivotal Quantities Coming from the Underlying Distributions

In this section, the technique of finding the joint distributions of the pivotal quantities V_1, V_2 is given. These joint distributions are required to construct the optimal prediction intervals for future order statistics coming from the underlying distributions under parametric uncertainty.

A. *Mathematical Preliminaries*

Theorem 3. Let $X_1 \le ... \le X_k$ be the first k ordered observations (order statistics) in a sample of size m from a continuous distribution with some probability density function $f_\theta(x)$ and distribution function $F_\theta(x)$, where θ is a parameter (in general, vector). Then the joint probability density function of $X_1 \le ... \le X_k$ and the lth order statistics X_l $(1 \le k < l \le m)$ is given by

$$f_\theta(x_1,...,x_k,x_l) = f_\theta(x_1,...,x_k)g_\theta(x_l \mid x_k), \tag{24}$$

where

$$f_\theta(x_1,...,x_k) = \frac{m!}{(m-k)!}\prod_{i=1}^{k} f_\theta(x_i)[1-F_\theta(x_k)]^{m-k}, \tag{25}$$

$$g_\theta(x_l \mid x_k) = \frac{(m-k)!}{(l-k-1)!(m-l)!}\left[\frac{F_\theta(x_l)-F_\theta(x_k)}{1-F_\theta(x_k)}\right]^{l-k-1}$$

$$\times\left[1-\frac{F_\theta(x_l)-F_\theta(x_k)}{1-F_\theta(x_k)}\right]^{m-l}\frac{f_\theta(x_l)}{1-F_\theta(x_k)}$$

$$= \frac{(m-k)!}{(l-k-1)!(m-l)!}\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}(-1)^j$$

$$\times\left[\frac{1-F_\theta(x_l)}{1-F_\theta(x_k)}\right]^{m-l+j}\frac{f_\theta(x_l)}{1-F_\theta(x_k)}$$

$$= \frac{(m-k)!}{(l-k-1)!(m-l)!}\sum_{j=0}^{m-l}\binom{m-l}{j}(-1)^j$$

$$\times\left[\frac{F_\theta(x_l)-F_\theta(x_k)}{1-F_\theta(x_k)}\right]^{l-k-1+j}\frac{f_\theta(x_l)}{1-F_\theta(x_k)} \tag{26}$$

represents the conditional probability density function of X_l given $X_k=x_k$.

Proof. The joint density of $X_1 \le ... \le X_k$ and X_l is given by

$$f_\theta(x_1,...,x_k,x_l) = \frac{m!}{(l-k-1)!(m-l)!}\prod_{i=1}^{k} f_\theta(x_i)$$

$$\times[F_\theta(x_l)-F_\theta(x_k)]^{l-k-1} f_\theta(x_l)[1-F_\theta(x_l)]^{m-l}$$

$$= f_\theta(x_1,...,x_k)g_\theta(x_l \mid x_k). \tag{27}$$

It follows from (27) that

$$f_\theta(x_l \mid x_1,...,x_k) = \frac{f_\theta(x_1,...,x_k,x_l)}{f_\theta(x_1,...,x_k)} = g_\theta(x_l \mid x_k), \tag{28}$$

i.e., the conditional distribution of X_l, given $X_i = x_i$ for all $i = 1,\ldots, k$, is the same as the conditional distribution of $X_i = x_i$, given only $X_k = x_k$, which is given by (26). This ends the proof. □

Corollary 3.1. The conditional probability distribution function of X_l given $X_k=x_k$ is

$$P_\theta\{X_l \le x_l \mid X_k = x_k\} = 1-\frac{(m-k)!}{(l-k-1)!(m-l)!}$$

$$\times\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}\frac{(-1)^j}{m-l+1+j}\left[\frac{1-F_\theta(x_l)}{1-F_\theta(x_k)}\right]^{m-l+1+j}$$

$$= \frac{(m-k)!}{(l-k-1)!(m-l)!}\sum_{j=0}^{m-l}\binom{m-l}{j}\frac{(-1)^j}{l-k+j}$$

$$\times\left[\frac{F_\theta(x_l)-F_\theta(x_k)}{1-F_\theta(x_k)}\right]^{l-k+j}. \tag{29}$$

B. *Two-Parameter Exponential Distribution*

Let us assume that the random variable X follows the two-parameter exponential distribution with the probability density function

$$f_\theta(x) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), \quad x > \mu, \quad \sigma > 0, \tag{30}$$

and the probability distribution function

$$F_\theta(x) = 1-\exp\left(-\frac{x-\mu}{\sigma}\right), \tag{31}$$

where $\theta = (\mu,\sigma)$, μ is the location parameter, and σ is the scale parameter $(\sigma > 0)$.

Theorem 4. Let $X_1 \le ... \le X_k$ be the first k ordered observations (order statistics) in a sample of size m from the two-parameter exponential distribution (30). Then the conditional probability density function of the lth order statistic X_l ($1 \le k < l \le m$) given $X_k = x_k$ is

$$g_\sigma(x_l \mid x_k) = \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j} (-1)^j$$

$$\times \frac{1}{\sigma} \exp\left(-\frac{(m-l+1+j)(x_l - x_k)}{\sigma} \right)$$

$$= \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{m-l} \binom{m-l}{j} (-1)^j$$

$$\times \frac{1}{\sigma} \left[1 - \exp\left(-\frac{x_l - x_k}{\sigma} \right) \right]^{l-k-1+j} \exp\left(-\frac{x_l - x_k}{\sigma} \right), \quad (32)$$

and the conditional probability distribution function of the lth order statistic X_l given $X_k = x_k$ is

$$P_\sigma\{X_l \le x_l \mid X_k = x_k\}$$

$$= 1 - \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j}$$

$$\times \frac{(-1)^j}{m-l+1+j} \exp\left(-\frac{(m-l+1+j)(x_l - x_k)}{\sigma} \right)$$

$$= \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{m-l} \binom{m-l}{j} \frac{(-1)^j}{l-k+j}$$

$$\times \left[1 - \exp\left(-\frac{x_l - x_k}{\sigma} \right) \right]^{l-k+j}. \quad (33)$$

Proof. It follows from (26) and (29), respectively. □

Corollary 4.1. It follows from (32) and (33) that the probability density function of the pivotal quantity $V_1 = (X_l - X_k)/\sigma$ is given by

$$f_1(v_1) = \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j} (-1)^j$$

$$\times \exp(-(m-l+1+j)v_1)$$

$$= \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{m-l} \binom{m-l}{j} (-1)^j$$

$$\times [1 - \exp(-v_1)]^{l-k-1+j} \exp(-v_1), \quad (34)$$

and the probability distribution function of the pivotal quantity V_1 is given by

$$F_1(v_1) = 1 - \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j}$$

$$\times \frac{(-1)^j}{m-l+1+j} \exp(-(m-l+1+j)v_1)$$

$$= \frac{1}{\mathrm{B}(l-k,(m-l+1)} \sum_{j=0}^{m-l} \binom{m-l}{j} \frac{(-1)^j}{l-k+j}$$

$$\times [1 - \exp(-v_1)]^{l-k+j}. \quad (35)$$

It is known that the statistic $(\hat{\mu}, \hat{\sigma})$, where

$$\hat{\mu} = X_1 \quad (36)$$

and

$$\hat{\sigma} = \frac{\sum_{i=1}^{k} (x_i - x_1) + (m-k)(x_k - x_1)}{k-1}, \quad (37)$$

is sufficient for $\theta = (\mu, \sigma)$; also $\hat{\mu}$ and $\hat{\sigma}$ are independently distributed, with

$$p_\theta(\hat{\mu}, \hat{\sigma}) = p_\theta(\hat{\mu}) p_\sigma(\hat{\sigma}), \quad (38)$$

where

$$p_\theta(\hat{\mu}) = \frac{m}{\sigma} \exp\left[-\frac{m(\hat{\mu} - \mu)}{\sigma} \right], \quad \hat{\mu} > \mu, \quad (39)$$

and

$$p_\sigma(\hat{\sigma}) = \left(\frac{k-1}{\sigma} \right)^{k-1} \frac{\hat{\sigma}^{k-2}}{\Gamma(k-1)} \exp\left(-\frac{(k-1)\hat{\sigma}}{\sigma} \right), \quad \hat{\sigma} > 0. \quad (40)$$

It follows from (40) that that the probability density function of the pivotal quantity $V_2 = \hat{\sigma}/\sigma$ is given by

$$f_2(v_2) = (k-1)^{k-1} \frac{v_2^{k-2}}{\Gamma(k-1)} \exp[(k-1)v_2], \quad v_2 > 0. \quad (41)$$

Thus, the joint probability density function of the pivotal quantities V_1, V_2 is given by

$$f(v_1, v_2) = f_1(v_1) f_2(v_2), \quad (42)$$

which is required to construct the optimal prediction intervals for future order statistics coming from the two-parameter exponential distribution (30).

Theorem 5. Let $X_1 \le \ldots \le X_k$ be the first k ordered observations (order statistics) in a sample of size m from the exponential distribution (30). Then the probability density function of the pivotal quantity $V_2^\circ = (X_k - X_1)/\sigma$ is given by

$$f_2^{\circ}(v_2^{\circ})=\frac{1}{\mathrm{B}(k-1,(m-k+1)}\sum_{j=0}^{k-2}\binom{k-2}{j}(-1)^j$$

$$\times\exp(-(m-k+1+j)v_2^{\circ})$$

$$=\frac{1}{\mathrm{B}(k-1,(m-k+1)}\sum_{j=0}^{m-k}\binom{m-k}{j}(-1)^j$$

$$\times[1-\exp(-v_2^{\circ})]^{k-2+j}\exp(-v_2^{\circ}),\quad v_2^{\circ}>0. \tag{43}$$

Proof. It follows from (32) that the conditional probability density function of the kth order statistic X_k ($1 \le k < l \le m$) given $X_1 = x_1$ is

$$g_{\sigma}(x_k \mid x_1)=\frac{1}{\mathrm{B}(k-1,(m-k+1)}\sum_{j=0}^{k-2}\binom{k-2}{j}(-1)^j$$

$$\times\frac{1}{\sigma}\exp\left(-\frac{(m-k+1+j)(x_k-x_1)}{\sigma}\right)$$

$$=\frac{1}{\mathrm{B}(k-1,(m-k+1)}\sum_{j=0}^{m-k}\binom{m-k}{j}(-1)^j$$

$$\times\frac{1}{\sigma}\left[1-\exp\left(-\frac{x_k-x_1}{\sigma}\right)\right]^{k-2+j}\exp\left(-\frac{x_k-x_1}{\sigma}\right), \tag{44}$$

It is clear that (43) follows from (44). This ends the proof. □

In this case, the joint probability density function of the pivotal quantities V_1, V_2° is given by

$$f^{\circ}(v_1,v_2^{\circ})=f_1(v_1)f_2^{\circ}(v_2^{\circ}) \tag{45}$$

and may be considered as the alternative to (42). It will be noted that (42) and (45) are not necessarily alternative.

C. Gumbel Distribution

Let us assume that the random variable X follows the Gumbel distribution with the probability density function

$$f_{\theta}(x)=\frac{1}{\sigma}\exp\left(\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)\quad(-\infty<x<\infty),$$

$$-\infty<x<\infty,\quad \sigma>0, \tag{46}$$

and the probability distribution function

$$\Pr\{X\le x\}=1-\exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]. \tag{47}$$

where $\theta = (\mu,\sigma)$, μ is the location parameter, and σ is the scale parameter ($\sigma > 0$).

Theorem 6. Let $X_1 \le \ldots \le X_k$ be the first k ordered observations (order statistics) in a sample of size m from the Gumbel distribution (46). Then the conditional probability distribution function of the lth order statistic X_l ($1 \le k < l \le m$) given $X_k = x_k$ is given by

$$G_{\theta}(x_l \mid x_k)=P_{\theta}\{X_l\le x_l \mid X_k=x_k\}=1-\frac{(m-k)!}{(l-k-1)!(m-l)!}$$

$$\times\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}\frac{(-1)^j}{m-l+1+j}\left[\exp\left(-\left(e^{\frac{x_l-\mu}{\sigma}}-e^{\frac{x_k-\mu}{\sigma}}\right)\right)\right]^{m-l+1+j}. \tag{48}$$

Proof. The proof follows from (29) and (47). □

Corollary 6.1. The conditional probability density function of X_l given $X_k=x_k$ is

$$g_{\theta}\{x_l \mid x_k\}=\frac{(m-k)!}{(l-k-1)!(m-l)!}$$

$$\times\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}(-1)^j\left[\exp\left(-\left(e^{\frac{x_l-\mu}{\sigma}}-e^{\frac{x_k-\mu}{\sigma}}\right)\right)\right]^{m-l+1+j}\frac{1}{\sigma}e^{\frac{x_l-\mu}{\sigma}}. \tag{49}$$

Theorem 7. The probability distribution function of the pivotal quantity $V_1=(X_l-X_k)/\sigma$ is given by

$$F_1(v_1)=P\{V_1\le v_1\}$$

$$=1-\frac{m!}{(l-k-1)!(m-l)!(k-1)!}\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}\frac{(-1)^j}{m-l+1+j}$$

$$\times\sum_{r=0}^{k-1}\binom{k-1}{r}(-1)^r\frac{1}{(\exp(v_1)-1)(m-l+1+j)+(m-k+1+r)}$$

$$=1-\frac{m!}{(l-k-1)!(m-l)!}\sum_{j=0}^{l-k-1}\binom{l-k-1}{j}\frac{(-1)^j}{m-l+1+j}$$

$$\times\left(\prod_{r=0}^{k-1}[(\exp(v_1)-1)(m-l+1+j)+(m-k+1+r)]\right)^{-1}. \tag{50}$$

Proof. We reduce (48) to

$$P_{\theta}\left\{\frac{X_l-X_k}{\sigma}\le\frac{x_l-x_k}{\sigma}\middle| X_k=x_k\right\}$$

$$= P_\theta\{V_1 \le v_1 \mid X_k = x_k\} = G_\theta(v_1 \mid x_k) = 1 - \frac{(m-k)!}{(l-k-1)!(m-l)!}$$

$$\times \sum_{j=0}^{l-k-1} \binom{l-k-1}{j} \frac{(-1)^j}{m-l+1+j} \left[\exp\left(-e^{\frac{x_k-\mu}{\sigma}} \left(e^{\frac{x_l-x_k}{\sigma}} - 1 \right) \right) \right]^{m-l+1+j}$$

$$= 1 - \frac{(m-k)!}{(l-k-1)!(m-l)!} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j} \frac{(-1)^j}{m-l+1+j}$$

$$\times \left[\exp\left(-\exp\left(\frac{x_k-\mu}{\sigma} \right) (\exp(v_1)-1) \right) \right]^{m-l+1+j}, \quad (51)$$

where $V_1 = (X_l - X_k)/\sigma$ is the pivotal quantity. Using the known distribution of the kth order statistic X_k, we eliminate the location parameter μ from the problem as

$$F_1(v_1) = P\{V_1 \le v_1\} = \int_0^\infty G_\theta(v_1 \mid x_k) g_\theta(x_k) dx_k, \quad (52)$$

where

$$g_\theta(x_k) = \frac{m!}{(k-1)!(m-k)!} F_\theta^{k-1}(x_k)[1 - F_\theta(x_k)]^{m-k} f_\theta(x_k)$$

$$= \frac{m!}{(k-1)!(m-k)!} \left[1 - \exp\left[-\exp\left(\frac{x_k-\mu}{\sigma} \right) \right] \right]^{k-1}$$

$$\times \left(\exp\left[-\exp\left(\frac{x_k-\mu}{\sigma} \right) \right] \right)^{m-k} \frac{1}{\sigma} \exp\left(\frac{x_k-\mu}{\sigma} \right) \exp\left(-\exp\left(\frac{x_k-\mu}{\sigma} \right) \right),$$

$$x_k \in (-\infty, \infty), \quad (53)$$

represents the probability density function of the kth order statistic X_k coming from the Gumbel distribution (46). This ends the proof. □

Corollary 7.1. The probability density function of the pivotal quantity $V_1 = (X_l - X_k)/\sigma$ is given by

$$f_1(v_1) = \frac{m!}{(l-k-1)!(m-l)!(k-1)!} \sum_{j=0}^{l-k-1} \binom{l-k-1}{j} (-1)^j$$

$$\sum_{r=0}^{k-1} \binom{k-1}{r} (-1)^r \frac{\exp(v_1)}{[(\exp(v_1)-1)(m-l+1+j) + (m-k+1+r)]^2}. \quad (54)$$

Theorem 8. Let $X_1 \le \ldots \le X_k$ be the first k ordered observations from a sample of size m, which follow the Gumbel distribution (46). Then the joint probability density function of the pivotal quantities

$$S_1 = \frac{\hat{\mu} - \mu}{\sigma}, \quad V_2 = \frac{\hat{\sigma}}{\sigma}, \quad (55)$$

conditional on fixed

$$\mathbf{z}^{(k)} = (z_1, \ldots, z_k), \quad (56)$$

where

$$Z_i = \frac{X_i - \hat{\mu}}{\hat{\sigma}}, \quad i = 1, \ldots, k, \quad (57)$$

are ancillary statistics, any k–2 of which form a functionally independent set, $\hat{\mu}$ and $\hat{\sigma}$ are the maximum likelihood estimates for μ and σ based on the first k ordered observations ($X_1 \le \ldots \le X_k$) from a sample of size m from the Gumbel distribution (46), which can be found from solution of

$$\hat{\mu} = \hat{\sigma} \ln\left(\left[\sum_{i=1}^k e^{x_i/\hat{\sigma}} + (m-k) e^{x_k/\hat{\sigma}} \right] \Big/ k \right), \quad (58)$$

and

$$\hat{\sigma} = \left(\sum_{i=1}^k x_i e^{x_i/\hat{\sigma}} + (m-k) x_k e^{x_k/\hat{\sigma}} \right)$$

$$\times \left(\sum_{i=1}^k e^{x_i/\hat{\sigma}} + (m-k) e^{x_k/\hat{\sigma}} \right)^{-1} - \frac{1}{k} \sum_{i=1}^k x_i, \quad (59)$$

is given by

$$f(s_1, v_2 \mid \mathbf{z}^{(k)}) = \vartheta^\bullet(\mathbf{z}^{(k)}) v_2^{k-2} \exp\left(v_2 \sum_{i=1}^k z_i \right)$$

$$\times e^{ks_1} \exp\left(-e^{s_1} \left[\sum_{i=1}^k \exp(z_i v_2) + (m-k) \exp(z_k v_2) \right] \right)$$

$$= f(v_2 \mid \mathbf{z}^{(k)}) f(s_1 \mid v_2, \mathbf{z}^{(k)}), \quad s_1 \in (-\infty, \infty), \quad v_2 \in (0, \infty), \quad (60)$$

where

$$\vartheta^\bullet(\mathbf{z}^{(k)}) = \left(\Gamma(k) \int_0^\infty v_2^{k-2} \exp\left(v_2 \sum_{i=1}^k z_i \right) \right.$$

$$\left. \times \left[\sum_{i=1}^k \exp(z_i v_2) + (m-k) \exp(z_k v_2) \right]^{-k} dv_2 \right)^{-1} \quad (61)$$

is the normalizing constant,

$$f(v_2 \mid \mathbf{z}^{(k)}) = \vartheta(\mathbf{z}^{(k)}) v_2^{k-2} \exp\left(v_2 \sum_{i=1}^k z_i \right)$$

$$\times\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]^{-k}, \quad v_2\in(0,\infty), \quad (62)$$

$$\vartheta(\mathbf{z}^{(k)})=\left(\int_0^\infty v_2^{k-2}\exp\left(v_2\sum_{i=1}^{k}z_i\right)\right.$$

$$\left.\times\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]^{-k}dv_2\right)^{-1}, \quad (63)$$

$$f(s_1\mid v_2,\mathbf{z}^{(k)})=\frac{1}{\Gamma(k)}\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]^{k}$$

$$\times e^{ks_1}\exp\left(-e^{s_1}\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]\right),$$

$$s_1\in(-\infty,\infty). \quad (64)$$

Proof. The joint density of $X_1 \le ... \le X_k$ is given by

$$f_\theta(x_1,...,x_k)=\frac{m!}{(m-k)!}\prod_{i=1}^{k}\frac{1}{\sigma}\exp\left(\frac{x_i-\mu}{\sigma}-\exp\left(\frac{x_i-\mu}{\sigma}\right)\right)$$

$$\times\exp\left(-(m-k)\exp\left(\frac{x_k-\mu}{\sigma}\right)\right). \quad (65)$$

Using the invariant embedding technique [12-24], we reduce (65) to

$$f_\theta(x_1,...,x_k)\,d\widehat{\mu}\,d\widehat{\sigma}$$

$$=\frac{m!}{(m-k)!}\frac{1}{\widehat{\sigma}^{k-2}}\Gamma(k)[\vartheta(\mathbf{z}^{(k)})]^{-1}$$

$$\times\vartheta(\mathbf{z}^{(k)})v_2^{k-2}\exp\left(v_2\sum_{i=1}^{k}z_i\right)$$

$$\times\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]^{-k}dv_2$$

$$\times\frac{1}{\Gamma(k)}\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]^{k}$$

$$\times e^{ks_1}\exp\left(-e^{s_1}\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]\right)ds_1$$

$$=\frac{m!}{(m-k)!}\frac{1}{\widehat{\sigma}^{k-2}}\Gamma(k)[\vartheta(\mathbf{z}^{(k)})]^{-1}$$

$$\times f_2(v_2\mid\mathbf{z}^{(k)})dv_2\, f_1(s_1\mid\mathbf{z}^{(k)})ds_1. \quad (66)$$

It follows from (66) that the probability element of the joint density of S_1, V_2, conditional on fixed $\mathbf{z}^{(k)}=(z_1,...,z_k)$, is

$$f(s_1,v_2\mid\mathbf{z}^{(k)})ds_1dv_2$$

$$=\vartheta^{\bullet}(\mathbf{z}^{(k)})v_2^{k-2}\exp\left(v_2\sum_{i=1}^{k}z_i\right)e^{ks_1}$$

$$\times\exp\left(-e^{s_1}\left[\sum_{i=1}^{k}\exp(z_i v_2)+(m-k)\exp(z_k v_2)\right]\right)ds_1dv_2,$$

$$s_1\in(-\infty,\infty), \quad v_2\in(0,\infty). \quad (67)$$

This ends the proof. □

Thus, the joint probability density function of the pivotal quantities V_1, V_2 is given by

$$f(v_1,v_2\mid\mathbf{z}^{(k)})=f_1(v_1)f_2(v_2\mid\mathbf{z}^{(k)}), \quad (68)$$

which is required to construct the optimal prediction intervals for future order statistics coming from the Gumbel distribution (46).

V. Conclusions and Directions for Future Research

In many statistical decision problems it is reasonable co confine attention to rules that are invariant with respect to a certain group of transformations. If a given decision problem admits a sufficient statistic, it is well known that the class of invariant rules based on the sufficient statistic is essentially complete in the class of all invariant rules under some assumptions. This result may be used to show that if there exists an optimal invariant rule among invariant rules based on sufficient statistic, it is optimal among all invariant rules. In this paper, we consider statistical prediction problems which are invariant with respect to a certain group of transformations and construct the optimal invariant interval predictors. The method used is that of the invariant embedding of sample statistics in a loss function in order to form pivotal quantities which allow one 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.

More work is needed, however, to obtain optimal prediction intervals for future order statistics 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.

Remark. It should be remarked that if we deal, for instance, with within-sample prediction and wish to obtain the best invariant prediction interval (d_1,d_2) for X_l, which has the prescribed confidence coefficient (or level) γ, we have to minimize the risk function $R(\theta, d_1,d_2) = E_\theta\{r(\theta, d_1,d_2)\}$ under constraint

$$\Pr\{d_1 \le X_l \le d_2\} = \gamma. \tag{69}$$

It can be shown that this problem is reduced to the following one:

Minimize

$$R(\eta_1,\eta_2) = c_1\int_0^\infty\int_0^{\eta_1 v_2}(-v_1+\eta_1 v_2)\,f(v_1,v_2)dv_1dv_2$$

$$+\,c_2\int_0^\infty\int_{\eta_2 v_2}^\infty(v_1-\eta_2 v_2)\,f(v_1,v_2)dv_1dv_2$$

$$+\,c(\eta_2-\eta_1)\int_0^\infty\int_0^\infty v_2\,f(v_1,v_2)dv_1dv_2 \tag{70}$$

Subject to

$$\int_0^\infty\int_{\eta_1 v_2}^{\eta_2 v_2} f(v_1,v_2)dv_1dv_2 = \gamma, \tag{71}$$

i.e., the unknown parameter θ is eliminated from the problem.

ACKNOWLEDGMENT

This research 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.

REFERENCES

[1] K. Pearson,"The fundamental problem of practical statistics," *Biometrika*, vol.13, pp. 1–16, 1920.

[2] G. Baker, "The probability that the mean of a second sample will differ from the mean of a first sample by less than a certain multiple of the standard deviation of the first sample," *Annals of Mathematical Statistics*, vol. 6, pp. 197–201, 1935.

[3] S. Wilks, "Statistical prediction with special reference to the problem of tolerance limits," *Annals of Mathematical Statistics*, vol. 13, pp. 400–409, 1942.

[4] J. Aitchison and I. Dunsmore, *Statistical Prediction Analysis*. Cambridge University Press, 1975.

[5] S. Geisser, *Predictive Inference: An Introduction*. Chapman & Hall, 1993.

[6] C.E. Antle and F. Rademaker, "An upper confidence limit on the maximum of *m* future observations from a Type I extreme-value distribution," *Biometrika*, vol. 59, pp. 475–477, 1972.

[7] J.F. Lawless, "On estimation of safe life when the underlying distribution is Weibull," *Technometrics*, vol. 15, pp. 857–865, 1973.

[8] R.W. Mee and D. Kushary, "Prediction limits for the Weibull distribution utilizing simulation," *Computational Statistics & Data Analysis*, vol. 17, pp. 327–336, 1994.

[9] P.F. Christoffersen, "Evaluating interval forecasts," *International Economic Review*, vol. 39, pp. 841-862, 1998.

[10] J.K. Patel, "Prediction intervals − a review," *Communications in Statistics –Theory and Methods*, vol. 13, pp. 2393 −2465, 1989.

[11] J.F. Bjørnstad, "Predictive likelihood: a review (with discussion)," *Statistical Science*, vol. 5, pp. 242-265, 1990.

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

[13] N.A. Nechval, K.N. Nechval, and E.K. Vasermanis, "Effective state estimation of stochastic systems," *Kybernetes (An International Journal of Systems & Cybernetics)*, vol. 32, pp. 666 –678, 2003.

[14] 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, pp. 141–159, 2008.

[15] N.A. Nechval, K.N. Nechval, and M. Purgailis, "Prediction of future values of random quantities based on previously observed data," *Engineering Letters*, vol. 9, pp. 346–359, 2011.

[16] N.A. Nechval, M. Purgailis, K.N. Nechval, and V.F. Strelchonok, "Optimal predictive inferences for future order statistics via a specific loss function," *IAENG International Journal of Applied Mathematics*, vol. 42, pp. 40 –51, 2012.

[17] N.A. Nechval, M. Purgailis, K. Cikste, G. Berzins, and K.N. Nechval, "Optimization of statistical decisions via an invariant embedding technique," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2010*, WCE 2010, 30 June - 2 July, 2010, London, U.K., pp. 1776–1782.

[18] N.A. Nechval, M. Purgailis, K. Cikste, and K.N. Nechval, "Planning inspections of fatigued aircraft structures via damage tolerance approach," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2010*, WCE 2010, 30 June - 2 July, 2010, London, U.K., pp. 2470–2475.

[19] N.A. Nechval, M. Purgailis, K.N. Nechval, and U. Rozevskis, "Optimization of prediction intervals for order statistics based on censored data," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2011*, WCE 2011, 6-8 July, 2011, London, U.K., pp. 63-69.

[20] N.A. Nechval, K.N. Nechval, and M. Purgailis, "Statistical inferences for future outcomes with applications to maintenance and reliability," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2011*, WCE 2011, 6-8 July, 2011, London, U.K., pp. 865-871.

[21] N.A. Nechval, M. Purgailis, K.N. Nechval, and I. Bruna, "Optimal inventory control under parametric uncertainty via cumulative customer demand," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2012*, WCE 2012, 4-6 July, 2012, London, U.K., pp. 6 –11.

[22] N.A. Nechval, M. Purgailis, K.N. Nechval, and I. Bruna, "Optimal prediction intervals for future order statistics from extreme value distributions," in *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2012*, WCE 2012, 4-6 July, 2012, London, U.K., pp. 1340-1345.

[23] N.A. Nechval, K.N. Nechval, and M. Purgailis, "Inspection policies in service of fatigued aircraft structures," in *Electrical Engineering and Applied Computing*, Sio-Iong Ao and Len Gelman (Eds.). LNEE, vol. 90,. Berlin: Springer-Verlag, 2011, pp. 459–472

[24] N.A. Nechval, K.N. Nechval, and M. Purgailis, "Weibull prediction limits for a future number of failures under parametric uncertainty," in *Electrical Engineering and Intelligent Systems*, Sio-Iong Ao and Len Gelman (Eds.). LNEE, vol. 130, Berlin: Springer-Verlag, 2013, pp. 273–284.