# Optimization of Prediction Intervals for Order Statistics Based on Censored Data

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

***Abstract*—Prediction intervals for order statistics are widely used for reliability problems and other related problems. The determination of these intervals has been extensively investigated. But the optimality property of these intervals has not been fully explored. In this paper, in order to discuss this problem, a risk function is introduced to compare prediction intervals. In particular, new-sample prediction based on a previous sample (i.e., when for predicting the future observation in a new sample there are available the data only from a previous sample), and within-sample prediction based on the early observed data from a current experiment (i.e., when for predicting the future observation in a sample there are available the early observed data only from that sample). We restrict attention to families of distributions invariant under location and/or scale changes. 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. An illustrative example is given.**

***Index Terms*—Order statistic, prediction interval, risk function, optimization**

## I. Introduction

PREDICTION of an unobserved random variable is a fundamental problem in statistics. Patel [1] provides an extensive survey of literature on this topic. In the areas of reliability and life-testing, lifetime data are often modeled via the Exponential and the Weibull in order to make predictions about future observations. Prediction intervals are constructed to have a reasonably high probability of containing a specified number of such future observations. These limits may be helpful in establishing warranty policy, determining maintenance schedules, etc. For a very readable discussion of prediction limits and related intervals, see Hahn and Meeker [2].

Many authors have reported their efforts for constructing prediction limits for the Weibull and for the related extreme value distributions (see Patel [1]). Mann and Saunders [3] proposed prediction limits for the Weibull which make use of only two or three order statistics (see also Mann [4]). Antle and Rademaker [5] used simulation to produce a table of factors to use with ML estimates to obtain prediction limits. Lawless [6] proposed prediction limits based on a conditional confidence approach; his limits require both determination of the ML estimates and numerical integration. Engelhardt and Bain [7], [8] and Fertig, Meyer and Mann [9] have proposed various approximate prediction limits for the Weibull. Mee and Kushary [10] 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. Bhaumik and Gibbons [11] and Krishnamoorthy et al. [12] proposed approximate methods for constructing upper prediction limits for a gamma distribution.

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

Manuscript received March 06, 2011. 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: marispur@lanet.lv).

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

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


ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

future events in a sample or process based on early data from that sample or process. If, for example, $n$ 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 $l$ additional failures, $X_{(k+l)}$; 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 $\boldsymbol{\theta}$. In this paper, attention is restricted to 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 $F((x-\mu)/\sigma)$, where $F(\cdot)$ is known but both the location ($\mu$) and scale ($\sigma$) 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 $\mu$ (the location parameter) and $\sigma$ (the scale) into $c\mu + b$ and $c\sigma$, respectively, where $b$ lies in the range of $\mu$, $c > 0$. This group acts transitively on the parameter space.

2) *Information on the values of components of the parametric vector* $\boldsymbol{\theta}$. 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. [13]-[20]) applicable whenever the statistical problem is invariant under a group of transformations, which acts transitively on the parameter space.

## II. Within-Sample Prediction Problem

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. For example, if $n$ units are followed until $t_k$ and there are $k$ observed failures, $t_1, \ldots, t_k$, one could be interested in predicting the time of the next failure $t_{k+1}$; time until $l$ additional failures, $t_{k+l}$; number of additional failures in a future interval.

### A. Location-Scale Family of Density Functions

Consider a situation described by a location-scale family of density functions, indexed by the vector parameter $\boldsymbol{\theta}=(\mu,\sigma)$, where $\mu$ and $\sigma$ (>0) are respectively parameters of location and scale. For this family, invariant under the group $G$ of positive linear transformations: $x \to ax+b$ with $a>0$, we shall assume that there is obtainable (from some informative experiment) the first $k$ order statistics $X_1<X_2< \cdots <X_k$ from a random sample of size $n$ with cumulative distribution function

$$F(x \mid \mu,\sigma) \equiv F\left(\frac{x-\mu}{\sigma}\right),$$

$$(-\infty)\mu < x < \infty, \ -\infty < \mu < \infty, \ \sigma > 0. \tag{1}$$

If $Y$ is an independent future observation from the same sample of size $n$, then $W=(Y - X_k)/S_k$ (or $W=(Y - X_k)/X_k$) is an invariant statistic, the distribution of which does not depend on $(\mu,\sigma)$; $S_k$ is a sufficient statistic (or a maximum likelihood estimator $\hat{\sigma}_k$) for $\sigma$ based on $\mathbf{X}=(X_1, X_2,\ldots, X_k)$.

### B. Piecewise-Linear Loss Function

We shall consider the interval prediction problem for the $r$th order statistic $X_r$, $k<r\leq n$, in the same sample of size $n$ for the situation where the first $k$ observations $X_1 < X_2 < \cdots < X_k$, $1\leq k<n$, have been observed. Suppose that we assert that an interval $\mathbf{d}=(d_1,d_2)$ contains $X_r$. 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 $X_r$ or if $d_2$ falls below $X_r$. Suppose that these penalties are $c_1(d_1- X_r)$ and $c_2(X_r-d_2)$, losses proportional to the amounts by which $X_r$ 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 $\mu$ 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 $\mathbf{d}=(d_1,d_2)$ is wide. Suppose that the cost associated with the interval is proportional to its length, say $c(d_2-d_1)$. In the specification of the loss function, $\sigma$ 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

$$r(\boldsymbol{\theta},\mathbf{d}) = \begin{cases} \dfrac{c_1(d_1 - X_r)}{\sigma} + \dfrac{c(d_2-d_1)}{\sigma} & (X_r < d_1), \\ \dfrac{c(d_2-d_1)}{\sigma} & (d_1 \leq X_r \leq d_2), \\ \dfrac{c(d_2-d_1)}{\sigma} + \dfrac{c_2(X_r - d_2)}{\sigma} & (X_r > d_2). \end{cases} \tag{2}$$

The decision problem specified by the informative experiment density function (1) and the loss function (2) is invariant under the group $G$ of transformations. Thus, the problem is to find the best invariant interval predictor of $X_r$,

$$\mathbf{d}^* = \arg\min_{\mathbf{d}\in\mathscr{D}} R(\boldsymbol{\theta},\mathbf{d}), \tag{3}$$

where $\mathscr{D}$ is a set of invariant interval predictors of $X_r$, $R(\boldsymbol{\theta},\mathbf{d})=E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d})\}$ is a risk function.

### C. Transformation of the Loss Function

It follows from (2) that the invariant loss function, $r(\boldsymbol{\theta},\mathbf{d})$, can be transformed as follows:

$$r(\boldsymbol{\theta},\mathbf{d}) = \ddot{r}(\mathbf{V},\boldsymbol{\eta}), \tag{4}$$

where

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$\ddot{r}(\mathbf{V},\boldsymbol{\eta}) = \begin{cases} c_1(-V_1+\eta_1 V_2)+c(\eta_2-\eta_1)V_2 & (V_1<\eta_1 V_2), \\ c(\eta_2-\eta_1)V_2 & (\eta_1 V_2 \le V_1 \le \eta_2 V_2), \\ c_2(V_1-\eta_2 V_2)+c(\eta_2-\eta_1)V_2 & (V_1>\eta_2 V_2), \end{cases} \quad (5)$$

$$\mathbf{V}=(V_1,V_2), \quad V_1=(X_r-X_k)/\sigma, \quad V_2=S_k/\sigma;$$

$$\boldsymbol{\eta}=(\eta_1,\eta_2), \quad \eta_1=(d_1-X_k)/S_k, \quad \eta_2=(d_2-X_k)/S_k. \quad (6)$$

*D. Risk Function*

It follows from (5) that the risk associated with $\mathbf{d}$ and $\boldsymbol{\theta}$ can be expressed as

$$R(\boldsymbol{\theta},\mathbf{d}) = E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d})\} = E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$$

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

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

$$+c(\eta_2-\eta_1)\int_0^{\infty}\int_0^{\infty}v_2 f(v_1,v_2)dv_1dv_2, \quad (7)$$

which is constant on orbits when an invariant predictor (decision rule) $\mathbf{d}$ is used, where $f(v_1,v_2)$ is defined by the joint probability density of the first $k$ observations $X_1 < X_2 < \cdots < X_k$ and $X_r$,

$$f(x_1,x_2,...,x_k,x_r\,|\,\mu,\sigma) = \frac{n!}{(r-k-1)!(n-r)!}$$

$$\times[F(x_r\,|\,\mu,\sigma)-F(x_k\,|\,\mu,\sigma)]^{r-k-1}[1-F(x_r\,|\,\mu,\sigma)]^{n-r}$$

$$\times\prod_{i=1}^{k} f(x_i\,|\,\mu,\sigma)f(x_r\,|\,\mu,\sigma). \quad (8)$$

*E. Risk Minimization and Invariant Prediction Rules*

The following theorem gives the central result in this section.

*Theorem 1* (*Optimal Predictor of $X_r$ Based on* $\mathbf{X}$). Suppose that $(u_1, u_2)$ is a random vector having density function

$$u_2 f(u_1,u_2)\left[\int_0^{\infty}\int_0^{\infty}u_2 f(u_1,u_2)du_1du_2\right]^{-1} \quad (u_1,u_2>0), \quad (9)$$

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 $X_r$ based on $\mathbf{X}$ is $\mathbf{d}^*=(X_k+\eta_1 S_k, X_k+\eta_2 S_k)$, where

$$Q(\eta_1)=c/c_1, \quad Q(\eta_2)=1-c/c_2. \quad (10)$$

(ii) If $c/c_1+c/c_2\ge 1$ then the optimal invariant linear-loss interval predictor of $X_r$ based on $\mathbf{X}$ degenerates into a point predictor $X_k+\eta_\bullet S_k$, where

$$Q(\eta_\bullet)=c_2/(c_1+c_2). \quad (11)$$

*Proof.* From (7)

$$\frac{\partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}}{\partial\eta_1}$$

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

$$= \int_0^{\infty}\int_0^{\infty}v_2 f(v_1,v_2)dv_1dv_2[c_1Q(\eta_1)-c], \quad (12)$$

and

$$\frac{\partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}}{\partial\eta_2} = \int_0^{\infty}\int_0^{\infty}v_2 f(v_1,v_2)dv_1dv_2[-c_2(1-Q(\eta_2))+c], \quad (13)$$

where

$$Q(\eta)=\int_0^{\eta}q(w)dw, \quad (14)$$

$$q(w)=\frac{\int_0^{\infty}v_2^2 f(wv_2,v_2)dv_2}{\int_0^{\infty}\int_0^{\infty}v_2 f(v_1,v_2)dv_1dv_2}, \quad (15)$$

$$W=V_1/V_2. \quad (16)$$

Now $\partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}/\partial\eta_1 = \partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}/\partial\eta_2 = 0$ if and only if (10) hold. Thus, $E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$ provided (10) 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 $\boldsymbol{\eta}=(\eta_1,\eta_2)$ gives the minimum value of $E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$. Thus (i) is established.

If $c/c_1+c/c_2\ge 1$ then the minimum of $E\{\ddot{r}(\mathrm{v},\eta)\}$ in the region $\eta_2\ge\eta_1$ occurs where $\eta_1=\eta_2=\eta_\bullet$, $\eta_\bullet$ being determined by setting

$$\partial E\{\ddot{r}(V,(\eta_\bullet,\eta_\bullet))\}/\partial\eta_\bullet=0 \quad (17)$$

and this reduces to

$$c_1Q(\eta_\bullet)-c_2[1-Q(\eta_\bullet)]=0, \quad (18)$$

which establishes (ii). □

*Corollary 1.1* (*Minimum Risk of the Optimal Invariant Predictor of $X_r$ Based on* $\mathbf{X}$). The minimum risk is given by

$$R(\boldsymbol{\theta},\mathbf{d}^*) = E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d}^*)\} = E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$$

$$= -c_1\int_0^{\infty}\int_0^{\eta_1 v_2}v_1 f(v_1,v_2)dv_1dv_2 + c_2\int_0^{\infty}\int_{\eta_2 v_2}^{\infty}v_1 f(v_1,v_2)dv_1dv_2 \quad (19)$$

for case (i) with $\boldsymbol{\eta}=(\eta_1,\eta_2)$ as given by (10) and for case (ii) with $\eta_1=\eta_2=\eta_\bullet$ as given by (11).

*Proof.* These results are immediate from (7) when use is made of $\partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}/\partial\eta_1 = \partial E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}/\partial\eta_2 = 0$ in case (i) and $\partial E\{\ddot{r}(\mathbf{V},(\eta_\bullet,\eta_\bullet))\}/\partial\eta_\bullet=0$ in case (ii). □

The underlying reason why $c/c_1+c/c_2$ acts as a separator of interval and point prediction is that for $c/c_1+c/c_2\ge 1$ every

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

interval predictor is inadmissible, there existing some point predictor with uniformly smaller risk.

*Theorem 2* (*Optimal Invariant Predictor of $X_r$ Based on $X_k$*). Suppose that $\mu$=0 and

$$\mathbf{V}=(V_1,V_2),\quad V_1=(X_r-X_k)/\sigma,\quad V_2=X_k/\sigma;$$

$$\boldsymbol{\eta}=(\eta_1,\eta_2),\quad \eta_1=(d_1-X_k)/X_k,\quad \eta_2=(d_2-X_k)/X_k. \tag{20}$$

Let us assume that $(u_1, u_2)$ is a random vector having density function

$$u_2 f_0(u_1,u_2)\left[\int_0^\infty\int_0^\infty u_2 f_0(u_1,u_2)du_1du_2\right]^{-1}\quad (u_1,u_2>0), \tag{21}$$

where $f_0$ is defined by $f_0(v_1,v_2)$, and let $Q_0$ 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 $X_r$ based on $X_k$ is $\mathbf{d}^*=((1+\eta_1)X_k, (1+\eta_2)X_k)$, where

$$Q_0(\eta_1)=c/c_1,\quad Q_0(\eta_2)=1-c/c_2. \tag{22}$$

(ii) If $c/c_1+c/c_2\geq 1$ then the optimal invariant linear-loss interval predictor of $X_r$ based on $X_k$ degenerates into a point predictor $(1+\eta_\bullet)X_k$, where

$$Q_0(\eta_\bullet)=c_2/(c_1+c_2). \tag{23}$$

*Proof.* For the proof we refer to Theorem 1. □

*Corollary 2.1* (*Minimum Risk of the Optimal Invariant Predictor of $X_r$ Based on $X_k$*). The minimum risk is given by

$$R(\boldsymbol{\theta},\mathbf{d}^*)=E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d}^*)\}=E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$$

$$=-c_1\int_0^\infty\int_0^{\eta_1 v_2} v_1 f_0(v_1,v_2)dv_1dv_2+c_2\int_0^\infty\int_{\eta_2 v_2}^\infty v_1 f_0(v_1,v_2)dv_1dv_2 \tag{24}$$

for case (i) with $\boldsymbol{\eta}=(\eta_1,\eta_2)$ as given by (22) and for case (ii) with $\eta_1=\eta_2=\eta_\bullet$ as given by (23).

*Proof.* For the proof we refer to Corollary 1.1. □

## III. Equivalent Confidence Coefficient

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

*Theorem 3* (*Equivalent Confidence Coefficient for* $\mathbf{d}^*$ *Based on* $\mathbf{X}$). Suppose that $\mathbf{V}=(V_1,V_2)$ is a random vector having density function $f(v_1,v_2)$ $(v_1,v_2>0)$ where $f$ is defined by (8) and let $H$ be the distribution function of $W=V_1/V_2$, i.e., the probability density function of $W$ is given by

$$h(w)=\int_0^\infty v_2 f(wv_2,v_2)dv_2. \tag{25}$$

Then the confidence coefficient associated with the optimum prediction interval $\mathbf{d}^*=(d_1,d_2)$, where $d_1=X_k+\eta_1S_k$, $d_2=X_k+\eta_2S_k$, is

$$\Pr\{\mathbf{d}^*: d_1<X_r<d_2\mid\mu,\sigma\}$$

$$=H[Q^{-1}(1-c/c_2)]-H[Q^{-1}(c/c_1)]. \tag{26}$$

*Proof.* The confidence coefficient for $\mathbf{d}^*$ corresponding to $(\mu,\sigma)$ is given by

$$\Pr\{(X_k,S_k): X_k+\eta_1S_k<X_r<X_k+\eta_2S_k\mid\mu,\sigma\}$$

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

$$=H(\eta_2)-H(\eta_1)=H[Q^{-1}(1-c/c_2)]-H[Q^{-1}(c/c_1)]. \tag{27}$$

This is independent of $(\mu,\sigma)$. □

*Theorem 4* (*Equivalent Confidence Coefficient for* $\mathbf{d}^*$ *Based on* $X_k$). Suppose that $\mathbf{V}=(V_1,V_2)$ is a random vector having density function $f_0(v_1,v_2)$ ($v_1$ real, $v_2>0$), where $f_0$ is defined by

$$f(x_k,x_r\mid\mu,\sigma)=\frac{1}{\mathrm{B}(k,r-k)\mathrm{B}(r,n-r+1)}$$

$$\times[F(x_k\mid\mu,\sigma)]^{r-1}[F(x_r\mid\mu,\sigma)-F(x_k\mid\mu,\sigma)]^{r-k-1}$$

$$\times[1-F(x_r\mid\mu,\sigma)]^{n-r}f(x_k\mid\mu,\sigma)f(x_r\mid\mu,\sigma), \tag{28}$$

and let $H_0$ be the distribution function of $W=V_1/V_2$, i.e., the probability density function of $W$ is given by

$$h_0(w)=\int_0^\infty v_2 f_0(wv_2,v_2)dv_2. \tag{29}$$

Then the confidence coefficient associated with the optimum prediction interval $\mathbf{d}^*=(d_1,d_2)$, where $d_1=(1+\eta_1)X_k$, $d_2=(1+\eta_2)X_k$, is

$$\Pr\{\mathbf{d}^*: d_1<X_r<d_2\mid\mu,\sigma\}$$

$$=H_0[Q_0^{-1}(1-c/c_2)]-H_0[Q_0^{-1}(c/c_1)]. \tag{30}$$

*Proof.* For the proof we refer to Theorem 3. □

The way in which (26) (or (30)) 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. New-Sample Prediction Problem

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 item, time until $l$ failures in a future sample of $m$ units, or number of failures by time $t_\bullet$ in a future sample of $m$ units.

### A. Location-Scale Family of Density Functions

Consider a situation described by a location-scale family of density functions, indexed by the vector parameter $\boldsymbol{\theta}=(\mu,\sigma)$, where $\mu$ and $\sigma$ (>0) are respectively parameters of location and scale. For this family, invariant under the group

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

of positive linear transformations: $x \to ax+b$ with $a>0$, we shall assume that there is obtainable from some informative experiment (the first $k$ order statistics $X_1<X_2< \cdots <X_k$ from a random sample of size $n$) a sufficient statistic $(M_k,S_k)$ (or a maximum likelihood estimator $(\widehat{\mu}_k, \widehat{\sigma}_k)$) for $(\mu,\sigma)$ based on $\mathbf{X}=(X_1, X_2,\ldots, X_k)$ with density function

$$p(m_k, s_k \mid \mu,\sigma) = \sigma^{-2} p_0[(m_k-\mu)/\sigma, s_k/\sigma]$$

$$-\infty < m_k < \infty,\ \ 0 < s_k < \infty,\ \ -\infty < \mu > \infty,\ \ \sigma > 0. \quad (31)$$

We are thus assuming that for the family of density functions an induced invariance holds under the group $G$ of transformations: $m_k \to am_k+b$, $s_k \to as_k$ or $\widehat{\mu}_k \to a\widehat{\mu}_k+b$, $\widehat{\sigma}_k \to a\widehat{\sigma}_k$ $(a>0)$. 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 structure of the problem is, however, more clearly seen within the general framework.

Let $Y$ be an independent future observation from a new sample. If $Y$ is invariantly predictable then $W=(Y-M_k)/S_k$ (or $W=(Y-\widehat{\mu}_k)/\widehat{\sigma}_k$) is a maximal invariant pivotal, conditional on $\mathbf{X}$.

### B. *Piecewise-Linear Loss Function*

We shall consider the interval prediction problem for the $s$th order statistic $Y_s$, $1 \le s \le m$, in a future sample of size $m$ for the situation where the first $k$ observations $X_1 < X_2 < \cdots < X_k$, $1 \le k < n$, from a past sample of size $n$ have been observed. Suppose that we assert that an interval $\mathbf{d}=(d_1,d_2)$ contains $Y_s$. 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 $Y_s$ or if $d_2$ falls below $Y_s$. Suppose that these penalties are $c_1(d_1 - Y_s)$ and $c_2(Y_s - d_2)$, losses proportional to the amounts by which $Y_s$ 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 $\mu$ 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 $\mathbf{d}=(d_1,d_2)$ is wide. Suppose that the cost associated with the interval is proportional to its length, say $c(d_2-d_1)$. In the specification of the loss function, $\sigma$ 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

$$r(\boldsymbol{\theta},\mathbf{d}) = \begin{cases} \dfrac{c_1(d_1-Y_s)}{\sigma} + \dfrac{c(d_2-d_1)}{\sigma} & (Y_s < d_1), \\ \dfrac{c(d_2-d_1)}{\sigma} & (d_1 \le Y_s \le d_2), \\ \dfrac{c(d_2-d_1)}{\sigma} + \dfrac{c_2(Y_s-d_2)}{\sigma} & (Y_s > d_2). \end{cases} \quad (32)$$

The decision problem specified by the informative experiment density function (31) and the loss function (32) is invariant under the group $G$ of transformations. Thus, the problem is to find the optimal interval predictor of $Y_s$,

$$\mathbf{d}^* = \arg\min_{\mathbf{d}\in\mathcal{D}} R(\boldsymbol{\theta},\mathbf{d}), \quad (33)$$

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

### C. *Transformation of the Loss Function*

It follows from (2) that the invariant loss function, $r(\boldsymbol{\theta},\mathbf{d})$, can be transformed as follows:

$$r(\boldsymbol{\theta},\mathbf{d}) = \ddot{r}(\mathbf{V},\boldsymbol{\eta}), \quad (34)$$

where

$$\ddot{r}(\mathbf{V},\boldsymbol{\eta}) = \begin{cases} c_1(-V_1+\eta_1 V_2) + c(\eta_2 - \eta_1)V_2 & (V_1 < \eta_1 V_2), \\ c(\eta_2-\eta_1)V_2 & (\eta_1 V_2 \le V_1 \le \eta_2 V_2), \\ c_2(V_1-\eta_2 V_2) + c(\eta_2-\eta_1)V_2 & (V_1 > \eta_2 V_2), \end{cases} \quad (35)$$

$$\mathbf{V}=(V_1,V_2),\quad V_1=(Y_s-M_k)/\sigma,\quad V_2=S_k/\sigma;$$

$$\boldsymbol{\eta}=(\eta_1,\eta_2),\quad \eta_1=(d_1-M_k)/S_k,\quad \eta_2=(d_2-M_k)/S_k. \quad (36)$$

### D. *Risk Function*

It follows from (35) that the risk associated with $\mathbf{d}$ and $\boldsymbol{\theta}$ can be expressed as

$$R(\boldsymbol{\theta},\mathbf{d}) = E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d})\} = E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$$

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

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

$$+ c(\eta_2-\eta_1)\int_0^{\infty}\int_{-\infty}^{\infty} v_2 f(v_1,v_2)\,dv_1 dv_2, \quad (37)$$

which is constant on orbits when an invariant predictor (decision rule) $\mathbf{d}$ is used, where $f(v_1,v_2)$ is defined by the joint probability density of the first $k$ observations $X_1 < X_2 < \cdots < X_k$ from the past random sample of size $n$ and the $s$th order statistic $Y_s$ in the future sample of size $m$,

$$f(x_1,x_2,\ldots,x_k,y_s \mid \mu,\sigma) = \frac{n!}{(n-k)!}\frac{m!}{(s-1)!(m-s)!}$$

$$\times \prod_{i=1}^{k} f(x_i \mid \mu,\sigma)[1-F(x_k \mid \mu,\sigma)]^{n-k}$$

$$\times [F(y_s \mid \mu,\sigma)]^{s-1}[1-F(y_s \mid \mu,\sigma)]^{m-s} f(y_s \mid \mu,\sigma). \quad (38)$$

### E. *Risk Minimization and Invariant Prediction Rules*

The following theorem gives the central result in this section.

*Theorem 5* (*Optimal Invariant Predictor of $Y_s$ Based on* $\mathbf{X}$). Suppose that $(u_1, u_2)$ is a random vector having density function

$$u_2 f(u_1,u_2)\left[\int_0^\infty \int_{-\infty}^\infty u_2 f(u_1,u_2)du_1du_2\right]^{-1} \quad (u_1 \text{ real},\ u_2>0), \tag{39}$$

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 $Y_s$ based on $\mathbf{X}$ is $\mathbf{d}^*=(M_k+\eta_1 S_k, M_k+\eta_2 S_k)$, where

$$Q(\eta_1)=c/c_1,\quad Q(\eta_2)=1-c/c_2. \tag{40}$$

(ii) If $c/c_1+c/c_2\geq 1$ then the optimal invariant linear-loss interval predictor of $Y_s$ based on $\mathbf{X}$ degenerates into a point predictor $M_k+\eta_\bullet S_k$, where

$$Q(\eta_\bullet)=c_2/(c_1+c_2). \tag{41}$$

*Proof*. For the proof we refer to Theorem 1. □

*Corollary 5.1* (*Minimum Risk of the Optimal Invariant Predictor of $Y_s$ Based on* **X**). The minimum risk is given by

$$R(\boldsymbol{\theta},\mathbf{d}^*)=E_{\boldsymbol{\theta}}\{r(\boldsymbol{\theta},\mathbf{d}^*)\}=E\{\ddot{r}(\mathbf{V},\boldsymbol{\eta})\}$$

$$=-c_1\int_0^\infty \int_{-\infty}^{\eta_1 v_2} v_1 f(v_1,v_2)dv_1dv_2+c_2\int_0^\infty \int_{\eta_2 v_2}^\infty v_1 f(v_1,v_2)dv_1dv_2 \tag{42}$$

for case (i) with $\boldsymbol{\eta}=(\eta_1,\eta_2)$ as given by (40) and for case (ii) with $\eta_1=\eta_2=\eta_\bullet$ as given by (41).

*Proof*. For the proof we refer to Corollary 1.1. □

*Theorem 6* (*Equivalent Confidence Coefficient for* $\mathbf{d}^*$ *Based on* **X**). Suppose that $\mathbf{V}=(V_1,V_2)$ is a random vector having density function $f(v_1,v_2)$ ($v_1$ real, $v_2>0$) where $f$ is defined by (38) and let $H$ be the distribution function of $W=V_1/V_2$, i.e., the probability density function of $W$ is given by

$$h(w)=\int_0^\infty v_2 f(wv_2,v_2)dv_2. \tag{43}$$

Then the confidence coefficient associated with the optimum prediction interval $\mathbf{d}^*=(d_1,d_2)$, where $d_1=M_k+\eta_1 S_k$, $d_2=M_k+\eta_2 S_k$, is

$$\Pr\{\mathbf{d}^*: d_1<X_r<d_2 \mid \mu,\sigma\}$$

$$=H[Q^{-1}(1-c/c_2)]-H[Q^{-1}(c/c_1)]. \tag{44}$$

*Proof*. For the proof we refer to Theorem 3. □

## V. Example

### A. Within-Sample Prediction of $X_r$ Based on X

*Exponential Distribution*. Let $X_1< X_2< \cdots <X_n$ be order statistics of size $n$ from the exponential distribution with the density

$$f(x\mid\sigma)=\frac{1}{\sigma}\exp\left(-\frac{x}{\sigma}\right),\quad x>0,\quad \sigma>0. \tag{45}$$

We shall consider the prediction problem of $X_r$ for the situation where the first $k$ observations $X_1< X_2< \cdots <X_k$, $1\leq k < r \leq n$, have been observed. Let $G$ be the group of transformations $x_i=ax_i$ ($i$=1, …, $k$, $r$, $n$, $a$>0) We are now concerned with optimization of the prediction interval for $X_r$ under the loss function (2).

Let $\mathbf{X}=(X_1, X_2,\dots, X_k)$ and $X_r > X_k$ for $r \leq n$. Then the joint probability density function of **X** and $X_r$ is given by

$$f(x_1,x_2,\dots,x_k,x_r\mid\sigma)=\frac{n!}{(r-k-1)!(n-r)!}$$

$$\times[F(x_r\mid\sigma)-F(x_k\mid\sigma)]^{r-k-1}[1-F(x_r\mid\sigma)]^{n-r}$$

$$\times\prod_{i=1}^k f(x_i\mid\sigma)f(x_r\mid\sigma)$$

$$=\frac{n!}{(r-k-1)!(n-r)!}\frac{1}{\sigma^{k+1}}\exp\left(-\frac{\sum_{i=1}^k x_i+(n-k)x_k}{\sigma}\right)$$

$$\times\left[1-\exp\left(-\frac{x_r-x_k}{\sigma}\right)\right]^{r-k-1}\left[\exp\left(-\frac{x_r-x_k}{\sigma}\right)\right]^{n-r+1}. \tag{46}$$

Let

$$V_1=\frac{X_r-X_k}{\sigma} \tag{47}$$

and

$$V_2=\frac{S_k}{\sigma}=\frac{\sum_{i=1}^k X_i+(n-k)X_k}{\sigma}. \tag{48}$$

Using the invariant embedding technique [13]-[20], we then find in a straightforward manner that the joint density of $V_1$, $V_2$ is

$$f(v_1,v_2)=f_1(v_1)f_2(v_2), \tag{49}$$

where

$$f_1(v_1)=\frac{[1-e^{-v_1}]^{r-k-1}[e^{-v_1}]^{n-r+1}}{\mathrm{B}(r-k,n-r+1)}$$

$$=\frac{1}{\mathrm{B}(r-k,n-r+1)}\sum_{j=0}^{r-k-1}\binom{r-k-1}{j}(-1)^j e^{-v_1(n-r+1+j)},$$

$$v_1>0, \tag{50}$$

and

$$f(v_2)=\frac{1}{\Gamma(k)}v_2^{k-1}e^{-v_2},\quad v_2>0. \tag{51}$$

It follows from (16) and (49) that

$$q(w)=\frac{\int_0^\infty v_2^2 f(wv_2,v_2)dv_2}{\int_0^\infty\int_0^\infty v_2 f(v_1,v_2)dv_1dv_2}=\frac{1}{k}\int_0^\infty v_2^2 f_1(wv_2)f_2(v_2)dv_2$$

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$= \frac{k+1}{\mathrm{B}(r-k, n-r+1)}$$

$$\times \sum_{j=0}^{r-k-1} \binom{r-k-1}{j} (-1)^j \frac{1}{[1+w(n-r+1+j)]^{k+2}}. \quad (52)$$

It follows from (25) and (49) that

$$h(w) = \int_0^\infty v_2 f(wv_2, v_2) dv_2 = \int_0^\infty v_2 f_1(wv_2) f_2(v_2) dv_2$$

$$= \frac{k}{\mathrm{B}(r-k, n-r+1)}$$

$$\times \sum_{j=0}^{r-k-1} \binom{r-k-1}{j} (-1)^j \frac{1}{[1+w(n-r+1+j)]^{k+1}}. \quad (53)$$

If $c/c_1+c/c_2<1$ then the optimal invariant linear-loss interval predictor of $X_r$ based on $\mathbf{X}$ is given by

$$\mathbf{d}^*=(X_k+\eta_1 S_k,\ X_k+\eta_2 S_k), \quad (54)$$

where

$$\eta_1 = \arg\left( \int_0^{\eta_1} q(w)dw = \frac{c}{c_1} \right) \quad (55)$$

and

$$\eta_2 = \arg\left( \int_0^{\eta_1} q(w)dw = 1 - \frac{c}{c_2} \right). \quad (56)$$

The confidence coefficient associated with the optimum prediction interval $\mathbf{d}^*=(d_1,d_2)$, where $d_1=X_k+\eta_1 S_k$, $d_2=X_k+\eta_2 S_k$, is given by

$$\Pr\{\mathbf{d}^* : d_1 < X_r < d_2 \mid \mu,\sigma\} = H[\eta_2] - H[\eta_1] = \int_{\eta_1}^{\eta_2} h(w)dw. \quad (57)$$

## VI. Conclusion

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 a minimax invariant rule among invariant rules based on sufficient statistic, it is minimax 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.

## References

[1] J. K. Patel, "Prediction Intervals − A Review," *Communications in Statistics −Theory and Methods*, vol. 13, pp. 2393–2465, 1989.
[2] G. J. Hahn and W. Q. Meeker, *Statistical Intervals: A Guide for Practitioners*. New York: Wiley, 1991.
[3] N. R. Mann and S. C. Saunders, "On Evaluation of Warranty Assurance when Life Has a Weibull Distribution," *Biometrika*, vol. 56, pp. 615–625, 1969.
[4] N. R. Mann, "Warranty Periods Based on Three Ordered Sample Observations from a Weibull Population," *IEEE Transactions on Reliability*, vol. R–19, pp. 167–171, 1970.
[5] 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.
[6] J. F. Lawless, "On the Estimation of Safe Life when the Underlying Life Distribution is Weibull," *Technometrics*, vol. 15, pp. 857–865, 1973.
[7] M. Engelhardt and L. J. Bain, "Prediction Limits and Two-Sample Problems with Complete or Censored Weibull Data," *Technometrics*, vol. 21, pp. 233–237, 1979.
[8] M. Engelhardt and L. J. Bain, "On Prediction Limits for Samples from a Weibull or Extreme-Value Distribution," *Technometrics*, vol. 24, pp. 147–150, 1982.
[9] K. W. Fertig, M. Mayer, and N. R. Mann, "On Constructing Prediction Intervals for Samples from a Weibull or Extreme Value Distribution," *Technometrics*, vol. 22, pp. 567–573, 1980.
[10] R. W. Mee and D. Kushary, "Prediction Limits for the Weibull Distribution Utilizing Simulation," *Computational Statistics & Data Analysis*, vol. 17, pp. 327–336, 1994.
[11] D. K. Bhaumik and R. D. Gibbons, "One-Sided Prediction Intervals for at Least *p* of *m* Observations from a Gamma Population at Each of *r* Locations," *Technometrics*, vol. 48, pp. 112–129, 2006.
[12] K. Krishnamoorthy, T. Mathew, and S. Mukherjee, "Normal Based Methods for a Gamma Distribution: Prediction and Tolerance Intervals and Stress-Strength Reliability," *Technometrics*, vol. 50, pp. 69–78, 2007.
[13] N. A. Nechval and K. N. Nechval, "Invariant Embedding Technique and Its Statistical Applications," in *Conference Volume of Contributed Papers of the 52nd Session of the International Statistical Institute*, Finland, Helsinki: ISI–International Statistical Institute, 1999, pp. 1–2. Available: http://www.stat.fi/isi99/proceedings/arkisto/varasto/nech0902.pdf.
[14] N. A. Nechval and E. K. Vasermanis, *Improved Decisions in Statistics*. Riga: SIA "Izglitibas soli", 2004.
[15] 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.
[16] N. A. Nechval, M. Purgailis, G. Berzins, K. Cikste, J. Krasts, and K. N. Nechval, "Invariant Embedding Technique and Its Applications for Improvement or Optimization of Statistical Decisions," in *Analytical and Stochastic Modeling Techniques and Applications*, K. Al-Begain, D. Fiems, and W. Knottenbelt (Eds.). LNCS, vol. 6148, Berlin, Heidelberg: Springer-Verlag, 2010, pp. 306–320.
[17] N. A. Nechval, M. Purgailis, K. Cikste, G. Berzins, U. Rozevskis, and K. N. Nechval, "Prediction Model Selection and Spare Parts Ordering Policy for Efficient Support of Maintenance and Repair of Equipment," in *Analytical and Stochastic Modeling Techniques and Applications*, K. Al-Begain, D. Fiems, and W. Knottenbelt (Eds.). LNCS, vol. 6148, Berlin, Heidelberg: Springer-Verlag, 2010, pp. 321–338.
[18] 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: World Congress on Engineering 2010*, pp. 1776–1782.
[19] 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: World Congress on Engineering 2010*, pp. 2470–2475.
[20] N. A. Nechval and M. Purgailis, "Improved State Estimation of Stochastic Systems via a New Technique of Invariant Embedding," in *Stochastic Control*, Chris Myers (Ed.). Croatia, India, Publisher: Sciyo, 2010, pp. 167–193.

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)