Optimal Prediction-Intervals for the Exponential Distribution, Based on Generalized Order Statistics

Mohammad Zayed Raqab

***Abstract*—This paper proposes optimal prediction intervals for the future generalized order statistic (GOS) based on the first r GOS (ordinary order statistics, usual record values, k-record values) from an exponential population. A conditional argument is considered for obtaining an optimal prediction interval for future GOS. A numerical example illustrates this technique.**

***Index Terms*—1-record value, exponential distribution, generalized order statistics, K-record value, ordinary order statistic, prediction interval.**

ACRONYMS[1]

OS	order statistic
GOS	generalized OS
HCD	highest conditional density
PLF	predictive likelihood function
MLE	maximum likelihood estimate
PMLE	predictive MLE
MLP	ML predictor
MSPE	mean square prediction error
pdf	probability density function
Cdf	cumulative distribution function
r.v.	random variable
i.i.d.	s-independent and identically distributed

Notation

n	sample size
(m, k)	indexes (positive real numbers) for classifying the model type
$X_{j:j}$	statistic j of the random sample $\{X_1, \ldots, X_j\}$
$X(r, n; m, k)$	GOS r of a sample of size n based on an (m, k) model
$\mathbf{X}$	$(X(1, n; m, k), \ldots, X(r, n; m, k))$: the first r GOS
$(L(\mathbf{X}), U(\mathbf{X}))$	(lower, upper) prediction limits
γ_r	$k + (n - r) \cdot (m + 1)$
C_{r-1}	$\prod_{i=1}^{r}, \gamma_i$
$g_m(x)$	$(1 - (1 - x)^{m+1})/(m + 1)$, for $m \neq -1$; $-\log(1 - x)$, for $m = -1$
$T_{r,m}$	$(m + 1) \cdot \sum_{i=1}^{r-1} X(i, n; m, k) + \gamma_r \cdot X(r, n; m, k)$
$\bar{F}(\cdot)$	$1 - F(\cdot)$
$Q(r, s)$	$\log(\gamma_{r+1})/(\gamma_s)$
$d_1(r, s)$	$\sum_{j=r+1}^{s}(1/\gamma_j)$
$d_2(r, s)$	$\sum_{j=r+1}^{s}(1/\gamma_j^2)$
$q_{r,s}$	$\prod_{j=r+1}^{s} \gamma_j$
$R_m(x, y)$	$(g_m(1 - \exp(-y))/g_m(1 - \exp(-x)))$
$\Gamma(a)$	gamma function
$G(\alpha, \beta)$	gamma distribution with shape, scale parameters α, β
$\mathrm{IG}(t, c)$	$(1/\Gamma(t)) \cdot \int_c^\infty x^{t-1} \cdot \exp(-x)dx$, incomplete gamma function.

Manuscript received April 22, 1998; revised May 24, 2000.
The author is with the Dept. of Mathematics, Univ. of Jordan, Amman 11942 Jordan (e-mail: MRaqab@uaeu.ac.ae).
Publisher Item Identifier S 0018-9529(01)06809-9.

[1]The singular & plural of an acronym are always spelled the same.

I. INTRODUCTION

THE EXPONENTIAL distribution is prominent in life-testing experiments and reliability problems. The use of order statistics in this connection arises quite reasonably. These statistics lead to system-reliability and quality-assurance variables and thus are important to reliability engineers and manufacturers. The ordinary rth OS in a sample of size n represents the life-length of a $(n - r + 1)$-out-of-n:G system made up of n i.i.d. components. All n components begin working simultaneously, and the system is good iff at least $(n - r + 1)$ components are good.

A flexible and more adequate model for a $(n - r + 1)$-out-of-n:G system in practical situations must consider a certain s-dependence structure. The failure of any component in the system can influence the remaining components. That is, the life distribution of the remaining components in the system can change after any component-failure. After failure r at time t, the remaining components possess a different life-distribution; it is truncated on the left at t to ensure realizations arranged in ascending order of magnitude. A triangular scheme of failure times is used, where line r contains $n - r + 1$ failures with Cdf $F_r, 1 \leq r \leq n$. Thus the system is a sequential $(n - r + 1)$-out-of-n:G system for some $r, 1 \leq r \leq n$; the rth sequential OS describes the life-length of the underlying system. See [9], [10] for further details.

A related model is the model of record values defined [4] as a model for successive extremes in a sequence of i.i.d. r.v. If several voltages of an equipment are considered, only the voltages greater than the previous one can be recorded. These recorded voltages are the upper record value sequence. In general, one records the voltage X_m of the mth value if $X_m > \max[X_1, \ldots, X_{m-1}]$, $m > 1$. Let $X_{i,j}$ be the voltage of an equipment on day j of location i. To study the local maximum voltages of $X_{i,j}$ at each location, the local maxima are the upper record values.

Let a technical system be subject to shocks, e.g., voltage peaks. If the shocks are viewed as realizations of an i.i.d. sequence, then the model of record values is adequate. Consider a situation in which the values of successive peak voltages and the occurrence of these peaks are of special interest. The waiting times between each two shocks can be modeled by record differences.

In the model of 1-record values, the interest is in the first largest value at each occurrence. If not the record values themselves, but second or third largest values are of special interest, then the model of k-record values [7] is adequate. Let $k = 2$; the sequence of 2-record values is obtained as follows:

1) Determine the 2nd largest failure time based on the first 2 failures.
2) Observe the 3rd failure time, and answer the question: What is the 2nd largest value of the new set of failure times? If the 2nd largest failure time of this set is greater than the previous, then record it; else keep the old set.
3) Record the failure times until a complete set of 2-record values is obtained. Let $T_{k,1} = k$ (k is a positive integer) and $n \geq 2$, then

$$T_{k,n} = \min\left[j : j > T_{k,n-1}, X_{j-k+1:j} > X_{T_{k,n-1}-k+1:T_{k,n-1}}\right].$$

Then the r.v. $Y_{k,n} = X_{T_{k,n}} - k + 1 : T_{k,n}, n \geq 1$ are the sequence of k-record values and the $T_{k,n}$ are the record times. Ordinary record values are obtained for $k = 1$. In reliability theory, the nth k-record value is the failure time of a k-out-of-$T_{k,n}$:G system.

The concept of GOS is introduced [10] as a unified approach to ordinary OS, record values, and k-record values. Let $X(1, n; m, k), \ldots, X(n, n; m, k)$ be GOS from an absolutely continuous Cdf $F(x)$ and pdf $f(x)$. Their joint pdf is [10: pp 50 – 51]

$$\begin{aligned} f_{1,\ldots,n}(x_1, \ldots, x_n) =& C_{n-1} \cdot \left[\prod_{i=1}^{n-1} f(x_i) \cdot [\bar{F}(x_i)]^m\right] \\ & \cdot f(x_n) \cdot \left[\bar{F}(x_n)\right]^{k-1}, \\ -\infty <& x_1 > \cdots < x_n \infty. \end{aligned} \tag{1.1}$$

- If $m = 0$ and $k = 1$, then $X(r, n; m, k)$ reduces to ordinary order statistic, r, and (1.1) is the joint pdf of the n OS, $X_{1:n} \leq \cdots \leq X_{n:n}$ [2], [5].
- If $m = -1$ and $k = 1$, then (1.1) is the joint pdf of the first n upper record values [1], [14].
- If $m = -1$ and $k \neq 1$, then (1.1) becomes the joint pdf of k-record values [7].

In this context, the model of GOS based on F can always be interpreted as sequential OS.

The prediction problems of the life-time models associated with the exponential distribution are very important, and have been studied in [3], [6], [8], [11], [13]. This paper considers the prediction problem: On the basis of the first r GOS from an exponential distribution, predict the GOS s, $X(s, n; m, k)$, $1 \leq r < s$. We use the idea of conditioning to obtain an optimal prediction interval for $X(s, n; m, k)$. Let an interval $(L(\mathbf{X}), U(\mathbf{x}))$ be a $(1 - \alpha)$ optimal prediction interval for $X(s, n; m, k)$ if

$$P_\theta(L(\mathbf{X}) < X(s, n; m, k) < U(\mathbf{X})) = 1 - \alpha,$$

for all θ, and its length is minimized. An HCD prediction interval is such that the conditional pdf of $X(s, n; m, k)$, given $\mathbf{X}$ for every point inside the interval, is greater than that for every point outside of it (see, in Bayes context, [15]).

II. Estimation and Prediction for $X(s, n; m, k)$

The exponential pdf and Cdf are, respectively,

$$\begin{aligned} f(x; \theta) =& \frac{1}{\theta} \cdot \exp\left[-\frac{x}{\theta}\right], \quad x > 0, \ \theta > 0; \\ F(x; \theta) =& 1 - \exp\left[-\frac{x}{\theta}\right], \quad x > 0, \ \theta > 0. \end{aligned} \tag{2.1}$$

From (1.1), the joint pdf of $X(1, n; m, k), \ldots, X(r, n; m, k)$, $1 < r \leq n$, is

$$\begin{aligned} f_{1,\ldots,r}(x_1, \ldots, x_r) =& C_{r-1} \cdot \left[\prod_{i=1}^{r-1} f(x_i) \cdot [\bar{F}(x_i)]^m\right] \\ & \cdot f(x_r) \cdot [\bar{F}(x_r)]^{\gamma_r - 1}, \\ -\infty <& x_1 > \cdots x_r < \infty. \end{aligned} \tag{2.2}$$

Similarly, the joint pdf of $X(r, n; m, k)$ and $X(s, n; m, k), 1 < r < s$ (2.3), and the marginal pdf of $X(r, n; m, k)$ (2.4) are

$$\begin{aligned} f_{r,s}(x_r, x_s) =& \frac{C_{s-1}}{\Gamma(r) \cdot \Gamma(s-r)} \cdot g_m^{r-1}(F(x_r)) \cdot [\bar{F}(x_r)]^m \\ & \cdot [g_m(F(x_s)) - g_m(F(x_r))]^{s-r-1} \\ & \cdot [\bar{F}(x_s)]^{\gamma_s - 1} \cdot f(x_r) \cdot f(x_s), \\ -\infty <& x_r < x_s < \infty; \end{aligned} \tag{2.3}$$

$$\begin{aligned} f_r(x_r) =& \frac{C_{r-1}}{\Gamma(r)} \cdot g_m^{r-1}(F(x_r)) \cdot [\bar{F}(x_r)]^{\gamma_r - 1} \cdot f(x_r), \\ -\infty <& x_r < \infty. \end{aligned} \tag{2.4}$$

From (2.1), (2.2), [8], the PLF of $X(s, n; m, k)$ and θ are

$$\begin{aligned} L(x_s, \theta; \mathbf{x}) =& \frac{C_{s-1}}{\Gamma(s-r) \cdot \theta^{r+1}} \\ & \cdot \exp\left[-\frac{1}{\theta} \cdot [T_{r,m} + \gamma_s \cdot x_s - \gamma_{r+1} \cdot x_r]\right] \\ & \cdot \left[g_m\left(1 - \exp\left(\frac{-x_s}{\theta}\right)\right)\right. \\ & \left. - g_m\left(1 - \exp\left(\frac{-x_r}{\theta}\right)\right)\right]^{s-r-1}, \\ 0 <& x_1 > \cdots < x_r < x_s < \infty. \end{aligned} \tag{2.5}$$

Take the logarithm of the PLF in (2.5), and differentiate with respect to $X(s, n; m, k)$ and θ, respectively, the results are the MLP and PMLE

$$\begin{aligned} & \hat{X}(s, n; m, k) \\ & = \begin{cases} X(r, n; m, k) + \frac{\hat{\theta}(1)}{m+1} \cdot Q(r, s), & \text{for } m \neq -1, \\ \frac{s \cdot X(r,n;m,k)}{r+1}, & \text{for } m = -1; \\ & \hat{\theta}(1) = \frac{T_{r,m}}{r+1}. \end{cases} \end{aligned} \tag{2.6}$$

Then

$$\begin{aligned}&\text{MSPE}[\hat{X}(s,n;m,k)]\\&=\theta^2\cdot\Bigg[d_2(r,s)+d_1^2(r,s)\\&\qquad+\frac{r\cdot Q(r,s)\cdot[Q(r,s)-2d_1(r,s)]}{(r+1)\cdot(m+1)^2}\Bigg],\\&m\neq -1.\end{aligned}$$

If $m=-1$, then

a) GOS $i, X(i,n;m,k)$ is the k-record i value.
b) Equation (2.1), (2.4) imply that $X(i,n;m,k)$ behaves like $G(i,\theta/k)$. Thus

$$\text{MSPE}(\hat{X}(s,n;m,k))=\frac{\theta^2\cdot s\cdot[s-(2k-1)\cdot r+1]}{(r+1)\cdot k^2},\quad \text{for } m=-1.$$

The MLP of $X(s,n;m,k)$ and the PMLE of θ for the ordinary order statistics $(m=0,k=1)$ and upper record values $(m=-1,k=1)$ are included in (2.6). From (2.1) and (2.2), the MLE of θ, based on the first r GOS, is $\hat{\theta}(2)=T_{r,m}/r$.

The PLF in (2.5) becomes

$$\begin{aligned}&L(x_s,\theta;\mathbf{x})\\&=\frac{1}{\theta^{r+1}}\cdot f\left(\frac{x_1}{\theta},\ldots,\frac{x_r}{\theta},\frac{x_s}{\theta}\right),\quad f_{1,\ldots,r,s}(x_1,\ldots,x_r,x_s)\\&=\frac{C_{s-1}}{\Gamma(s-r)}\cdot\exp[-T_{r,m}-\gamma_s\cdot x_s+\gamma_{r+1}\cdot x_r]\\&\quad\cdot[g_m(1-\exp(-x_s))-g_m(1-\exp(-x_r))]^{s-r-1},\\&0<x_1<\cdots<x_r<x_s<\infty.\end{aligned}$$

Thus the PLF $L(x_s,\theta;\mathbf{x})$ is a scale family; and it can be decomposed into

$$\begin{aligned}L_1(x_s,\theta|\mathbf{x})&=\frac{1}{\theta}\cdot f\left(\frac{x_s}{\theta}|\frac{x_r}{\theta}\right),\\L_2(\mathbf{x})&=\frac{1}{\theta^r}\cdot f_{1,\ldots,r}\left(\frac{x_1}{\theta},\ldots,\frac{x_r}{\theta}\right);\\f_{s|r}(x_s|x_r)&\equiv\frac{q_{r,s}}{\Gamma(s-r)}\cdot\exp(-\gamma_s\cdot x_s+\gamma_{r+1}\cdot x_r)\\&\quad\cdot[g_m(1-\exp(-x_r))]^{s-r-1}\\&\quad\cdot[R_m(x_r,x_s)-1]^{s-r-1},\\0&<x_r<x_s\infty.\end{aligned}\tag{2.7}$$

Because θ is unknown, replace it by the MLE of θ: $\hat{\theta}(2)$. This leads to the approximate conditional pdf $\hat{f}(x_s|x_r)$; thus the pivotal statistic is

$$V=\frac{X(s,n;m,k)-X(r,n;m,k)}{T_{r,m}}.$$

Adopt an s-expected length as a criterion; then the result [16] can be used to get an optimal prediction-interval among the scale-invariant prediction intervals. Thus the optimal prediction interval is in the form of

$$[X(r,n;m,k)+a\cdot T_{r,m}, X(r,n;m,k)+b\cdot T_{r,m}].$$

The constants a, b are determined through the pivot statistic V such that

$$\Pr\{a<V<b\}=1-\alpha,\ \text{and}\ h(a)=h(b);\ h(\cdot)=\text{pdf}[V].$$

By unimodality, it suffices to derive an optimal prediction interval for $X(s,n;m,k)$, $[L(\mathbf{x}),U(\mathbf{x})]$, such that

$$\int_{L(\mathbf{x})}^{U(\mathbf{x})}\hat{f}_{s|r}(x_s|x_r)dx_s=1-\alpha,\tag{2.8}$$

$$\hat{f}_{s|r}(L(\mathbf{x})|x_r)=\hat{f}_{s|r}(U(\mathbf{x})|x_r).\tag{2.9}$$

This is an HCD prediction interval. For $m=-1$, the approximate conditional pdf in (2.8) and (2.9) can be simplified to

$$\begin{aligned}\hat{f}(x_s|x_r)=&\frac{1}{\Gamma(s-r)}\cdot\left(\frac{r}{x_r}\right)^{s-r}\cdot(x_s-x_r)^{s-r-1}\\&\cdot\exp\left[-r\cdot\frac{x_s-x_r}{x_r}\right].\end{aligned}$$

Therefore, the $(1-\alpha)$ HCD prediction interval $[(1+w_1)\cdot X(r,n;-1,k),(1+w_2)\cdot X(r,n;-1,k)]$, must satisfy the 2 equations

$$\text{IG}(s-r,r\cdot w_1)-\text{IG}(s-r,r\cdot w_2)=1-\alpha,\tag{2.10}$$

$$\exp[-r\cdot(w_2-w_1)]=\left(\frac{w_1}{w_2}\right)^{s-r-1},\quad s=r+2,r+3,\ldots.\tag{2.11}$$

For $m\neq-1$, an approximate conditional pdf is

$$\begin{aligned}\hat{f}(x_s|x_r)=&\frac{r\cdot q_{r,s}}{\Gamma(s-r)\cdot(m+1)^{s-r-1}\cdot T_{r,m}}\\&\cdot\exp\left[\frac{-r\cdot\gamma_s\cdot(x_s-x_r)}{T_{r,m}}\right]\\&\cdot\left(1-\exp\left[-\frac{r\cdot(m+1)\cdot(x_s-x_r)}{T_{r,m}}\right]\right)^{s-r-1}.\end{aligned}$$

The $(1-\alpha)$ HCD prediction interval of $X(s,n;m,k)$ is in the form of

$$[X(r,n;m,k)+w_3\cdot T_{r,m}, X(r,n;m,k)+w_4\cdot T_{r,m}];$$

the w_3, w_4 are found by simultaneous solution of

$$\frac{m+1}{B\left(\frac{\gamma_s}{m+1},s-r\right)}\cdot\sum_{j=0}^{s-r-1}\frac{\Psi_3(j,s)-\Psi_4(j,s)}{\gamma_s+(m+1)\cdot j}=1-\alpha,\tag{2.12}$$

$$\begin{aligned}&\frac{\Psi_4(0,s)}{\Psi_3(0,s)}=\left(\frac{1-\exp[-r\cdot(m+1)\cdot w_3]}{1-\exp[-r\cdot(m+1)\cdot w_4]}\right)^{s-r-1};\\&\Psi_l(j,s)\equiv(-1)^j\cdot\binom{s-r-1}{j}\\&\qquad\cdot\exp\left[-r\cdot(\gamma_s+(m+1)\cdot j)\cdot w_l\right],\\&l=3,4,\quad s=r+2,r+3,\ldots.\end{aligned}\tag{2.13}$$

III. Example

Consider the following data which represent failure times, in minutes, for a specific type of electrical insulation in an experiment in which the insulation was subjected to a continuously increasing voltage stress [12]: 21.8, 70.7, 24.4, 138.6, 151.9, 75.3, 12.3, 95.5, 98.1, 43.2, 28.6, 46.9.

An exponential model is appropriate for these data. To illustrate the methods of inference in this paper, consider the 3 models:$(m = 0, k = 1), (m = -1, k = 1), (m = -1, k = 2)$. A numerical method was used to solve (2.10) and (2.11) for '$m = -1$' models, and (2.12) and (2.13) for '$m \neq -1$' models with Mathematica 3.0.

1) Ordered Data Analysis $(m = 0, k = 1)$

   Assume that only the first 8 ordered failure times are available and that the last 4 failure times are censored — since the experimenter failed to observe these failure times. Hence: $n = 12, r = 8, T_{r,m} = 624.4$. The PMLE & MLE of θ are, respectively: $\hat{\theta}(1) = 69.378, \hat{\theta}(2) = 78.050$. From (2.6), the MLP of the last failure time is: $\hat{X}(12, 12, 0, 1) = 171.478$. The interval (97.455, 421.988) provides a 95% HCD prediction interval of $X(12, 12; 0, 1)$.

2) Record Data Analysis $(m = -1, k = 1)$

   A failure time X_j is a record (upper record) if it exceeds all preceding failure times. Therefore, the 1-record values from the above data are: 21.8, 70.7, 138.6, 151.9. After the occurrence of a record used as a shock model, each shock is allowed to influence the magnitude of the subsequent shock. Here, $r = 4, T_{r,m} = 151.9$. The PMLE & MLE of θ are, $\hat{\theta}(1) = 30.380, \hat{\theta}(2) = 37.975$. The MLP and 95% HCD prediction interval of the 6th 1-recorded failure time are, respectively, $\hat{X}(6, 12; -1, 1) = 182.280, (L, U) = (153.509, 332.857)$.

3) k-Record Data Analysis $(m = -1)$

   The second largest failure time, based on the first 2 failures, is 21.8.

   If the value 24.4 is observed, then this value is the second largest failure time, based on the preceding failure times.

   If the value 138.6 is observed, it is the first largest failure time, and the 2nd largest failure time is 70.7.

   Proceeding similarly, the recorded failure times are: 21.8, 24.4, 70.7, 138.6.

   The failure times after 151.9 are not recorded because they have no influence on the previous failure times in the context of this model. The PMLE & MLE of θ, based on the first four 2-record values are, $\hat{\theta}(1) = 55.44, \hat{\theta}(2) = 69.30$.

   The MLP and 95% HCD prediction interval of the 7th 2-recorded failure time are: $\hat{X}(7, 12; -1, 2) = 194.04, (L, U) = (149.116, 360.402)$.

Acknowledgment

I am pleased to thank the Associate Editor (Dr. Kuo) for his helpful and constructive suggestions, which led to a substantially improved exposition of this paper.

References

[1] M. Ahsanullah, *Record Statistics*: Nova Science Publishers, 1995.

[2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*: John Wiley, 1992.

[3] A. M. Awad and M. Z. Raqab, "Prediction intervals for the future record values from exponential distribution: Comparative study," *J. Statistical Computation and Simulation*, vol. 65, pp. 325–340, 2000.

[4] K. N. Chandler, "The distribution and frequency of record values," *J. Royal Statistical Soc*, vol. B14, pp. 220–228, 1952.

[5] H. A. David, *Order Statistics*: John Wiley, 1981.

[6] I. R. Dunsmore, "The Bayesian predictive distribution in life testing models," *Technometrics*, vol. 16, pp. 455–460, 1974.

[7] W. Dziubdziela and B. Kopocinski, "Limiting properties of the kth record values," *Applied Mathematics*, vol. 15, pp. 187–190, 1976.

[8] K. S. Kaminsky and L. S. Rhodin, "Maximum likelihood prediction," *Annals of the Institute of Statistical Mathematics*, vol. 37, pp. 507–517, 1985.

[9] U. Kamps, "A concept of generalized order statistics," *J. Statistical Planning & Inference*, vol. 48, pp. 1–23, 1995.

[10] ——, "A Concept of Generalized Order Statistics," in *Habilitationsschrift*: Aachen Univ. of Technology, 1995.

[11] J. F. Lawless, "Prediction intervals for the two parameter exponential distribution," *Technometrics*, vol. 19, pp. 469–472, 1977.

[12] ——, *Statistical Models & Methods for Lifetime Data*: John Wiley, 1982.

[13] J. Likes, "Prediction of sth ordered observation for the two-parameter exponential distribution," *Technometrics*, vol. 16, pp. 241–244, 1974.

[14] H. N. Nagaraja, "Record values and related statistics – A review," *Communications in Statistics – Theory & Methods*, vol. 17, pp. 2223–2238, 1988.

[15] S. K. Sinha and H. A. Howlader, "Credible and HPD intervals of the parameter and reliability of Rayleigh distribution," *IEEE Trans. Reliability*, vol. R-32, pp. 217–220, 1983.

[16] K. Takeuchi, *Statistical Prediction Theory* Baihükan, Takyö, 1975.

Mohammad Zayed Raqab is an Associate Professor of Statistics at the University of Jordan. He received the B.S. (1981) and M.S. (1985) in Mathematics from the University of Jordan and Ph.D. (1992) in Statistics from the Ohio State University. He received the TWAS Prize to Youngest Scientists in 1998. He is now on leave at United Arab Emirates University. Dr Raqab is a member of ASA and IAPQ Transaction. His research interests include applied probability, reliability, statistical inferences involving order statistics and record values.