A note on asymptotic behaviour of some estimators based on records

ARUP BOSE* SREELA GANGOPADHYAY‡

November 7, 2008

Abstract

Asymptotic properties of estimators are not known in any generality for records based estimators. In this note we study asymptotic properties of BLUE in a few specific situations. In particular we consider (i) One-parameter Uniform distribution (ii) Two-parameter Uniform distribution and (iii) Two-parameter Extreme Value distribution. We establish the almost sure consistency (or lack of it) for these estimates and also derive the distributional convergence of their appropriate functionals.

Keywords: Upper records, best linear unbiased estimators, convergence in distribution, almost sure convergence, extreme value distribution.

AMS (2000) Subject Classification: Primary 62F12 60F05 60F15: Secondary 60G70.

---

*Stat Math Unit, Kolkata, Indian Statistical Institute

†Stat Math Unit, Kolkata, Indian Statistical Institute

‡The research of the second author is supported by the Department of Science and Technology, Government of India, under scheme no. SR/WOS-A/MS-02/2006.

1 Results

Suppose data on a parametric family is available in the form of records. Statistical estimation of parameters in such models has attracted recent attention, see for example [1] and [4]. Nevertheless, we seem to be far away from any reasonable general theory on statistical inference based on records. In particular, asymptotic properties of estimators such as the maximum likelihood estimator (MLE), best linear estimator (BLUE) are not known. For the BLUE based on record values, for some specific distributions such as the *two parameter exponential distribution*, *one parameter Weibull distribution*, nice expressions for these estimators and closed form formula for their means and variances, are known. The asymptotic properties in these cases are also known, by direct applications of results on records.

In this note we study asymptotic properties of BLUE in a few other specific situations. In particular we consider (i) One-parameter Uniform distribution (ii) Two-parameter Uniform distribution and (iii) Two-parameter Extreme Value distribution. We establish the almost sure consistency properties of these estimates and also derive the distributional convergence of their appropriate functionals.

Let $\{X_0, X_1, \ldots\}$ be a sequence of i.i.d. observations which has common cdf F. Let $L_0 \equiv 0$. After defining L_n, define L_{n+1} inductively as follows:

$$L_{n+1} = \inf\{j > L_n : X_j > X_{L_n}\}.$$

Then $\{X_{L_n}\}$ denoted by $\{R_n\}$ is the sequence of *upper records* of cdf F. We shall assume that given n, the data available is the first n upper records $\{R_0, R_1, \ldots R_n\}$.

1.1 One parameter Uniform Distribution.

Consider the one parameter uniform distribution $U(0,\sigma)$, so that $\{X_i\}$ are i.i.d with density $f(x) = \sigma^{-1},\ 0 < x < \sigma$. Then the BLUE of σ based on the first n upper records $\{R_i\ 1 \leq i \leq n\}$ is given by (see Arnold et al [3]),

$$\sigma_n^* = \frac{2}{3^{n+1}-1}\{2 \times 3^n R_n - 3^{n-1}R_{n-1} - 3^{n-2} - \cdots R_0\} \tag{1}$$

and the variance of σ_n^* is given by

$$Var(\sigma_n^*) = \frac{2\sigma^2}{3(3^{n+1}-1)}. \tag{2}$$

Since $E(\sigma_n^*) = \sigma$ and $\sum_{n=0}^{\infty} Var(\sigma_n^*) < \infty$, it immediately follows that σ_n^* is strongly consistent, that is, $\sigma_n^* \to \sigma$ almost surely. The following result gives the asymptotic distribution of σ_n^*.

Proposition 1 $\frac{\log|\sigma_n^* - \sigma| + (n+1)}{\sqrt{n+1}} \xrightarrow{\mathcal{D}} N(0,1)$.

Proof. It is well known that $\{R_n\}$ of $U(0,\sigma)$ has the following representation:

$$\{R_n\}_{n\geq 0} \stackrel{\mathcal{D}}{=} \{\sigma(1 - \prod_{i=0}^{n} U_i)\}_{n\geq 0}. \tag{3}$$

where U_i's are i.i.d. $\sim U(0,1)$. Now using the expression (1) for σ_n^* and representation (3) for $\{R_n\}_{n\geq 0}$ a simple algebra gives:

$$\frac{3^{n+1}-1}{4.3^n}(\sigma_n^* - \sigma) \stackrel{\mathcal{D}}{=} \sigma \prod_{i=0}^{n} U_i[-1 + \frac{1}{2.3}U_n^{-1} + \cdots + \frac{1}{2.3^n}\prod_{i=1}^{n} U_i^{-1} + \frac{1}{2.3^n}\prod_{i=0}^{n} U_i^{-1}]. \tag{4}$$

Taking $\log$ of absolute values of both sides we have

$$\log|\frac{3^{n+1}-1}{4.3^n}|+\log|\sigma_n^*-\sigma| \stackrel{\mathcal{D}}{=} \log\sigma+\sum_{i=0}^{n}\log U_i+\log|-1+\frac{1}{2.3}U_0^{-1}+\cdots+\frac{1}{2.3^n}\prod_{i=0}^{n-1}U_i^{-1}+\frac{1}{2.3^n}\prod_{i=0}^{n}U_i| \tag{5}$$

Note that $-\log U_i \sim Exp(1)$ and by strong law of large numbers (SLLN),

$$(\log\prod_{i=0}^{k}U_i^{-1})/(k+1) \to 1 \text{ almost surely.}$$

Therefore, for any $\epsilon > 0$,

$$(e^{1-\epsilon})^{k+1} < \prod_{i=0}^{k}U_i^{-1} < (e^{\epsilon+1})^{k+1} \ a.s. \ \forall \text{ large } k.$$

Choose ϵ such that $e^{1+\epsilon}/3 < 1$. Then it is easy to see that

$$\log|-1+\frac{1}{2.3}U_0^{-1}+\cdots+\frac{1}{2.3^n}\prod_{i=0}^{n-1}U_i^{-1}+\frac{1}{2.3^n}\prod_{i=0}^{n}U_i^{-1}| \to X, \text{ almost surely where, } X \text{ is finite.}$$

Since by CLT $\frac{\sum_{i=0}^{n}\log U_i+(n+1)}{\sqrt{n+1}} \stackrel{\mathcal{D}}{\to} N(0,1)$, from (5) we have

$$\frac{\log|\sigma_n^*-\sigma|+(n+1)}{\sqrt{n+1}} \stackrel{\mathcal{D}}{\to} N(0,1).$$

Remark 1: Note that the convergence of σ_n^* to σ is extremely fast. The above result may be used to compute asymptotically correct confidence interval of σ_n^*.

1.2 Two Parameter Uniform Distribution

Suppose that $\{X_i\}$ are i.i.d with density $f(x) = \sigma^{-1},\ \mu < x < \mu+\sigma$. Then the BLUEs of μ and σ based on upper record values $\{R_i, 1 \le i \le n\}$ are given respectively by (see Arnold et. al. [3])

$$\mu_n^* = \frac{2}{3^n-1}\{3^nR_0+R_1+3R_2+\cdots 3^{n-2}R_{n-1}-2\times 3^{n-1}R_n\}$$

and

$$\sigma_n^* = \frac{2}{3^n-1}\{4.3^{n-1}R_n-2.3^{n-2}-\cdots-2.3R_2-2R_1-(3^n+1)R_0\}.$$

Further,

$$Var(\mu_n^*) = \sigma^2\frac{3^{n+1}-1}{9(3^n-1)} \text{ and } Var(\sigma_n^*) = \sigma^2\frac{3^{n+1}+5}{9(3^n-1)}.$$

From developments given earlier note that

$$\{R_n\}_{n\ge 0} \stackrel{\mathcal{D}}{=} \{\sigma(1-\prod_{i=0}^{n}U_i)+\mu\}_{n\ge 0}$$

where U_i's are i.i.d. $\sim uniform(0,1)$. We now have the following proposition.

Proposition 2 *i)* $\frac{\log|\mu_n^* - 2R_0 + \mu + \sigma| + (n+1)}{\sqrt{n+1}} \xrightarrow{\mathcal{D}} N(0,1)$,
ii) $\frac{\log|\sigma_n^* + 2R_0 - 2\mu - 2\sigma| + (n+1)}{\sqrt{n+1}} \xrightarrow{\mathcal{D}} N(0,1)$

Proof. Observe that

$$\begin{aligned} \mu_n^* &= \frac{2}{3^n - 1}\{3^n R_0 + R_1 + 3R_2 + \cdots 3^{n-2}R_{n-1} - 2 \times 3^{n-1}R_n\} \\ \frac{3^n - 1}{2.3^n}(\mu_n^* - 2R_0) &= \frac{1}{3^n}R_1 + \frac{1}{3^{n-1}}R_2 + \cdots + \frac{1}{3^2}R_{n-1} - \frac{2}{3}R_n + \frac{1}{3^n}R_0 \end{aligned}$$

and hence

$$\begin{aligned} & \frac{3^n - 1}{2.3^n}(\mu_n^* - 2R_0 + (\sigma + \mu)) \\ \stackrel{\mathcal{D}}{=} \quad & \frac{2}{3}\sigma\prod_{i=0}^{n} U_i - \frac{1}{3^2}\sigma\prod_{i=0}^{n-1} U_i - \cdots - \frac{1}{3^n}\sigma U_0 U_1 - \frac{1}{3^n}\sigma U_0 \\ = \quad & \frac{2}{3}\sigma\prod_{i=0}^{n} U_i[1 - \frac{1}{2.3}U_n^{-1} - \frac{1}{2.3^2}U_n^{-1}U_{n-1}^{-1} - \cdots - \frac{1}{2.3^{n-1}}\prod_{i=2}^{n} U_i^{-1} - \frac{1}{2.3^{n-1}}\prod_{i=1}^{n} U_i^{-1}]. \end{aligned}$$

Taking log of absolute values of both sides and arguing as in section 1.1 we get

$$\frac{\log|\mu_n^* - 2R_0 + \mu + \sigma| + (n+1)}{\sqrt{n+1}} \xrightarrow{\mathcal{D}} N(0,1).$$

Similar argument shows that

$$\begin{aligned} & \frac{3^n - 1}{2.4.3^{n-1}}[\sigma_n^* + 2R_0 - 2\sigma - 2\mu] \\ \stackrel{\mathcal{D}}{=} \quad & -\sigma\prod_{i=0}^{n} U_i[1 - \frac{1}{2.3}U_n^{-1} - \frac{1}{2.3^2}U_n^{-1}U_{n-1}^{-1} - \cdots - \frac{1}{2.3^{n-1}}\prod_{i=2}^{n-1} - \frac{1}{2.3^{n-1}}\prod_{i=1}^{n} U_i^{-1}]. \end{aligned}$$

Taking log of absolute values of both sides and centering by $(n+1)$ and scaling by $\sqrt{n+1}$, we get (ii) of the proposition.

Remark 2: For large samples, this information can be used to form confidence intervals of $\mu + \sigma$. But from asymptotic distributions of μ_n^* and σ_n^* we do not gain much knowledge about the behaviour of μ and σ separately.

Remark 3. Since

$$\frac{\log|\mu_n^* - 2R_0 + \mu + \sigma| + (n+1)}{\sqrt{n+1}} \stackrel{\mathcal{D}}{=} W + Y$$

and

$$\frac{\log|\sigma_n^* + 2R_0 - 2\mu - 2\sigma| + (n+1)}{\sqrt{n+1}} \stackrel{\mathcal{D}}{=} W + Z$$

where $W \xrightarrow{\mathcal{D}} N(0,1)$ and both Y and Z converge to zero in probability, we immediately conclude that

$$\left(\frac{\log|\mu_n^* - 2R_0 + \mu + \sigma| + (n+1)}{\sqrt{n+1}}, \frac{\log|\sigma_n^* + 2R_0 - 2\mu - 2\sigma| + (n+1)}{\sqrt{n+1}}\right) \xrightarrow{\mathcal{D}} (N, N)$$

where $N \sim N(0,1)$.

Remark 4: Note that $Var(\mu_n^*)$ and $Var(\sigma_n^*)$ do not converge to zero, and μ_n^* and σ_n^* are not consistent. However, for suitable constants C_1 and C_2,

$$E|\mu_n^* - 2R_0 + (\mu+\sigma)|^2 \leq C_1[E|\frac{2}{3}\prod_{i=0}^{n} U_i - \frac{1}{3^2}\prod_{i=0}^{n-1} U_i - \cdots - \frac{1}{3^n}U_0U_1 - \frac{1}{3^n}U_0|^2 \leq C_2\frac{n^2}{3^n}$$

which is summable. As a consequence, almost surely,

$$|\mu_n^* + \mu + \sigma| \to 2R_0 \text{ and } |\sigma_n^* - 2\mu - 2\sigma| \to 2R_0.$$

1.3 Two Parameter Extreme Value Distribution

Here we assume that population distribution is $EV(\mu,\sigma)$ with cdf

$$F(x) = 1 - \exp(-e^{\frac{x-\mu}{\sigma}}).$$

Let $\{R_n\}_{n\geq 0}$ be the records of $EV(\mu,\sigma)$. Then

$$\{R_n\}_{n\geq 0} \stackrel{\mathcal{D}}{=} \{\mu + \sigma\log(\sum_{i=0}^{n} X_i^*)\}_{n\geq 0}$$

where X_i^* are i.i.d. with common distribution $Exp(1)$. The BLUE of μ and σ based on $\{R_i, 1 \leq i \leq n\}$ are given respectively by:

$$\mu_n^* = \frac{\alpha_n}{n}\sum_{i=0}^{n-1} R_i + (1-\alpha_n)R_n \quad \text{and} \quad \sigma_n^* = R_n - \frac{1}{n}\sum_{i=0}^{n-1} R_i$$

where $\alpha_n = -\gamma + \sum_{i=1}^{n}\frac{1}{i}$, γ being the Euler's constant.

Proposition 3 *i)* $\frac{\sqrt{n}(\mu_n^*-\mu)}{\sigma\alpha_n} \stackrel{\mathcal{D}}{\to} N(0,1)$
ii) $\frac{\sqrt{n}(\sigma_n^*-\sigma)}{\sigma} \stackrel{\mathcal{D}}{\to} N(0,1)$
iii) $\mu_n^* \to \sigma\ a.s.$ *and* $\sigma_n^* \to \sigma\ a.s.$

Proof. For the standard Extreme-Value distribution, that is, when $\mu = 0$ and $\sigma = 1$, denote the corresponding nth records by $\{R_i^{st}\}$. Clearly, $R_i = \mu + \sigma R_i^{st}$. To prove the proposition we require following decomposition of $\sum_{i=0}^{n} R_i^{st}$ introduced by Arnold et al ([2]):

$$T_n^{st} = \sum_{i=0}^{n} R_i^{st} \stackrel{\mathcal{D}}{=} -\sum_{i=1}^{n}\tilde{X}_i^* + (n+1)R_n^{st}$$

where $\tilde{X}_i^*$ are i.i.d. $Exp(1)$ r.v.s and independent of R_n^{st}. For the sake of completeness, we give the arguments to establish this:

$$T_n^{st} = \sum_{i=0}^{n} R_i^{st} \stackrel{\mathcal{D}}{=} \sum_{i=0}^{n}\log(\sum_{j=0}^{i} X_j^*) = \log[\prod_{k=1}^{n}(\frac{\sum_{i=0}^{k-1} X_i^*}{\sum_{i=0}^{k} X_i^*})^k][\sum_{i=0}^{n} X_i^*]^{n+1} \tag{6}$$

where X_i^* are i.i.d. $Exp(1)$. Now, $(\sum_{i=0}^{k-1} X_i^*/\sum_{i=0}^{k} X_i^*)$, $k = 1,2,\ldots n$ and $(\sum_{i=0}^{n} X_i^*)$ are independent. Also, for each k,

$$(\sum_{i=0}^{k-1} X_i^*/\sum_{i=0}^{k} X_i^*)^k \sim Uniform(0,1).$$

Using the fact that for $U \sim Uniform(0,1)$, $-\log U \sim Exp(1)$, we get the necessary decomposition. Using this decompsition,

$$
\begin{aligned}
\mu_n^* &= \frac{\alpha_n}{n}\sum_{i=0}^{n-1} R_i + (1-\alpha_n)R_n \\
&= \frac{\alpha_n}{n}\sigma T_n^{st} + \mu + \sigma(1-\alpha_n - \frac{\alpha_n}{n})R_n^{st} \\
&\overset{\mathcal{D}}{=} \frac{\alpha_n}{n}\sigma[-\sum_{i=1}^{n}\tilde{X}_i^* + (n+1)R_n^{st}] + \mu + \sigma(1-\alpha_n - \frac{\alpha_n}{n})R_n^{st}
\end{aligned}
$$

Therefore,

$$
\mu_n^* - \mu \overset{\mathcal{D}}{=} \frac{\alpha_n}{n}\sigma[-\sum_{i=1}^{n}(\tilde{X}_i^* - 1)] + \sigma[R_n^{st} - \alpha_n]. \tag{7}
$$

Dividing both sides by $\sigma\frac{\alpha_n}{n}\sqrt{n}$, we get

$$
\frac{\mu_n^* - \mu}{\sqrt{n}\frac{\alpha_n}{n}\sigma} \overset{\mathcal{D}}{=} \frac{\frac{\alpha_n}{n}\sigma[-\sum_{i=1}^{n}(\tilde{X}_i^* - 1)]}{\frac{\alpha_n}{n}\sigma\sqrt{n}} + \frac{\sigma(R_n^{st} - \alpha_n)}{\frac{\alpha_n}{n}\sigma\sqrt{n}}. \tag{8}
$$

Recall that $\tilde{X_i}^*$'s and R_n^{st} are independent and $\sqrt{n}(R_n^{st} - \alpha_n) \overset{\mathcal{D}}{\to} N(0,1)$ (see Arnold et al [2]). Using CLT for $\sum_{i=0}^{n}\tilde{X_i}^*$ we have

$$
\frac{\sqrt{n}(\mu_n^* - \mu)}{\sigma\alpha_n} \overset{\mathcal{D}}{\to} N(0,1)
$$

which proves (i). Similar calculations for σ_n^* give

$$
\sigma_n^* \overset{\mathcal{D}}{=} \frac{1}{n}\mu + \frac{1}{n}\sigma\sum_{i=1}^{n}\tilde{X}_i^* \tag{9}
$$

and hence

$$
\frac{\sqrt{n}(\sigma_n^* - \sigma)}{\sigma} \overset{\mathcal{D}}{\to} N(0,1)
$$

proving (ii). To prove (iii), recall that the moment generating function of $R_n^{st})$ is given by (see Arnold et al [3], p.32),

$$
E\exp(tR_n^{st}) = \frac{\Gamma(n+1+t)}{\Gamma(n+1)}.
$$

Using this and series representation of polygamma functions we can compute the fourth moments of

$$
\mu_n^* - \mu \overset{\mathcal{D}}{=} \frac{\alpha_n}{n}\sigma[-\sum_{i=1}^{n}(\tilde{X}_i^* - 1)] + \sigma[R_n^{st} - \alpha_n]
$$

and

$$
\sigma_n^* - \sigma \overset{\mathcal{D}}{=} \frac{1}{n}\mu + \frac{1}{n}\sigma\sum_{i=1}^{n}(\tilde{X}_i^* - 1)
$$

which are of the order $\frac{\alpha_n^4}{n^2}$ and $\frac{1}{n^2}$ respectively. We omit the details. Recall that α_n is of the order $\log n$. Hence both $E(\mu_n^* - \mu)^4$ and $E(\sigma_n^* - \sigma)^4$ are summable. Therefore μ_n^* and σ_n^* converge almost surely to μ and σ respectively.

Remark 5: $E(R_n^{st}) = \alpha_n$ and variance of R_n^{st} is $\frac{\pi^2}{6} - \sum_{i=0}^{n} \frac{1}{i^2}$, see Arnold et al [3]. Easy calculation gives,

$$Var(\mu_n^*) = \sigma^2[\frac{\alpha_n^2}{n} + (\frac{\pi^2}{6} - \sum_{i=1}^{n} \frac{1}{i^2})] \text{ and } Var(\sigma_n^*) = \frac{\sigma^2}{n}.$$

These variances are of order *no less* than $\frac{1}{n}$ and hence are not summable. This is why we require fourth order moment arguments in the above proof.

Remark 6: It is interesting to note that

$$\frac{\sqrt{n}(\mu_n^* - \mu)}{\sigma\alpha_n} \stackrel{\mathcal{D}}{=} W + Y \text{ and } \frac{\sqrt{n}(\sigma_n^* - \sigma)}{\sigma} \stackrel{\mathcal{D}}{=} W + Z,$$

where $W \stackrel{\mathcal{D}}{\to} N(0,1)$ and both Y and Z converge to zero in probability. Hence

$$(\frac{\sqrt{n}(\mu_n^* - \mu)}{\sigma\alpha_n}, \frac{\sqrt{n}(\sigma_n^* - \sigma)}{\sigma}) \stackrel{\mathcal{D}}{\to} (N, N)$$

where $N \sim N(0,1)$.

References

[1] Ahsanullah, M. (1986) Estimation of the parameters of the rectangular distribution by record values. *Computational Statistics Quarterly*. **2**:119-125.

[2] Arnold, B. C. and Villaseñor, J. A. (1999). The asymptotic distributions of sums of records. *Extremes*, **1(3)**: 351 - 363.

[3] Arnold, B. C.; Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*. Wiley, New York.

[4] Houchens, R. L. (1984) Record Value Theory and Inference. *Ph.D. Dissertation*, University of California, Riverside. California.

[5] Resnick, S. I. (1964). Limit laws for record values. *Stochastic Processes and Their Applications*, **1**: 67 - 82.

Corresponding author:

Sreela Gangopadhyay
Statistics and Mathematics Unit
Indian Statistical Institute
202 B. T. Road
Kolkata 700108, INDIA
Email: gangopadhyay7@gmail.com