CONSISTENT ESTIMATION BASED ON EXTREMES

By RICHARD F. GREEN

Technical Report No. 25

Department of Statistics
University of California
Riverside, California 92502

September, 1975

Summary

In this paper sufficient conditions are given for the existence of consistent estimators of parameters or functions of parameters based on extremes. The parametric families considered are stochastically ordered in the tail by the parameter or function of parameters in question.

When a consistent estimator of a parameter based on the maximum exists this estimator will remain consistent even if the sample size is known only to within bounded relative error. Consistent estimators based on maxima will also be consistent under monotone (decreasing) random truncation.

Many familiar families of distributions admit consistent estimators based on extremes.

This paper was prepared with the partial support of NIH Research Grant No. GM-10525, National Institutes of Health, Public Health Service.

AMS 1970 subject classifications. Primary 62F10; Secondary 62E15.

Key words and phrases. Consistent estimation, extremes, stochastically ordered in the tail, monotone random truncation.

1. Introduction

In his classic paper on the distribution of extreme values Gnedenko (1943) defined a law of large numbers and also relative stability for maxima. He gave necessary and sufficient conditions for both properties to hold. From Gnedenko's definitions it is easy to see that a location family of distributions each satisfying the law of large numbers for maxima admits a consistent estimator of the location parameter based on the maxima. Normal distributions with a known, fixed variance are such a family. Similarly, a scale family of distributions each satisfying the condition that the sequence of successive maxima is relatively stable admits a consistent estimator of the scale parameter based on the maxima. Gamma distributions with known, fixed location and shape parameters are such a family.

This paper gives conditions on the distribution functions similar to those given by Gnedenko for the law of large numbers for maxima and for relative stability of maxima which are sufficient for the existence of consistent estimators of parameters or functions of parameters for a very broad class of families of distributions.

2. Conditions for Consistent Estimation Based on Extremes

Let $\mathfrak{F}$ be a family of distributions parameterized by a set of vectors Θ .

Assume that all members of $\mathfrak{F}$ have infinite tails, that is, $F_\theta(x) < 1$ for finite x. Here $\theta \in \Theta$, $F_\theta \in \mathfrak{F}$.

Let ϕ be a real-valued function on Θ. Assume the range of ϕ is open.

DEFINITION 1. The family of distributions $\mathcal{F}$ will be said to be stochastically ordered in the tail with respect to ϕ if for any $\phi_1 < \phi_2$ in the range of ϕ there exists an x_0 such that for any $\theta_1, \theta_2 \in \Theta$ with $\phi(\theta_1) \leq \phi_1, \phi(\theta_2) \geq \phi_2$ we have

$$F_{\theta_1}(x) \geq F_{\theta_2}(x) \quad \text{for all} \quad x \geq x_0. \tag{1}$$

THEOREM 1. Under the assumptions above and if $\mathcal{F}$ is a family of distributions stochastically ordered in the tail with respect to ϕ then there exists a consistent estimate of ϕ based on the maxima if for any $\phi_1 < \phi_2$ contained in the range of ϕ

$$\frac{P_{\theta_1}(X \geq x)}{P_{\theta_2}(X > x)} \longrightarrow 0 \quad \text{as} \quad x \longrightarrow \infty \tag{2}$$

for all θ_1, θ_2 such that $\phi(\theta_1) = \phi_1, \phi(\theta_2) = \phi_2$.

PROOF. Assume that (1), (2) and the other conditions on $\mathcal{F}$ hold. Let $X_1, X_2, \ldots, X_n$ be independent random variables each having distribution F_{θ_0}, $\theta_0 \in \Theta$. Write $Y_n = \max_{i=1,\ldots,n} X_i$, and $\phi_0 = \phi(\theta_0)$.

An estimator $\hat{\phi}_n$ of ϕ_0 based on Y_n will be constructed and it will be shown that $\hat{\phi}_n \longrightarrow \phi_0$ as $n \longrightarrow \infty$.

For any $\gamma > 0, \varepsilon > 0$ we want to show that there exists an integer n_0 such that $P(|\hat{\phi}_n - \phi_0| < \gamma) \geq 1 - \varepsilon$ for all $n \geq n_0$.

Define. $\underline{\phi}_n = \inf\{\phi : \phi = \phi(\theta),\ P_\theta(X \geq Y_n) \geq 1/n\}$

$\overline{\phi}_n = \sup\{\phi : \phi = \phi(\theta),\ P_\theta(X > Y_n) \leq 1/n\}$

$\hat{\phi}_n = \underline{\phi}_n/2 + \overline{\phi}_n/2$, unless $\underline{\phi}_n = -\infty$, $\overline{\phi}_n = +\infty$,

in which case $\hat{\phi}_n = 0$.

Notice that $\underline{\phi}_n \le \bar{\phi}_n$.

Choose values $\phi_1, \phi_2, \phi_3, \phi_4$ in the range of ϕ such that

$$\phi_0 - \varepsilon < \phi_1 < \phi_2 < \phi_0 < \phi_3 < \phi_4 < \phi_0 + \varepsilon$$

We will show that there exists an integer n_0 such that for all $n \ge n_0$

1. $P(\phi_1 \le \underline{\phi}_n) > 1 - \varepsilon/2$, and

2. $P(\bar{\phi}_n \le \phi_4) > 1 - \varepsilon/2$.

Part 1. Let $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4 \in \Theta$ such that $\phi(\theta_i) = \phi_i$ for $i = 0,\ldots,4$.

Define. $Y_n^* = \inf\{y_n : P_{\theta_2}(X \ge y_n) < 1/n\}$.

Notice that $P_{\theta_2}(X \ge Y_n^*) \ge 1/n$.

Consider $P_{\theta_0}(Y_n \ge Y_n^*)$.

Pick an x_0 such that

$$\frac{P_{\theta_2}(X \ge x)}{P_{\theta_0}(X > x)} < \delta_1 \quad \text{for all} \quad x \ge x_0$$

where $\delta_1 = 1/\ln(2/\varepsilon)$.

Choose n such that $Y_n^* \ge x_0$. Then

$$\frac{P_{\theta_2}(X \ge Y_n^*)}{P_{\theta_0}(X > Y_n^*)} < \delta_1, \quad \text{or}$$

$$P_{\theta_0}(X \ge Y_n^*) > P_{\theta_2}(X \ge Y_n^*)/\delta_1 \ge 1/n\delta_1,$$

$$P_{\theta_0}(X \le Y_n^*) < 1 - 1/n\delta_1. \quad \text{Therefore}$$

$$P_{\theta_0}(Y_n \leq Y_n^*) < (1 - 1/n\delta_1)^n < \exp(-1/\delta_1), \quad \text{and}$$

$$P_{\theta_0}(Y_n > Y_n^*) > 1 - \exp(-1/\delta_1) = 1 - \varepsilon/2. \tag{3}$$

Now show that

$$\{Y_n > Y_n^*\} \subset \{\phi_1 \leq \underline{\phi}_n\}. \tag{4}$$

Consider any point contained in the left-hand set in (4), that is, $Y_n = y > Y_n^*$, and any $\theta_* \in \Theta$ such that $\phi(\theta_*) = \phi_* < \phi_1$, and show that

$$P_{\theta_*}(X \geq y) < 1/n.$$

We know that

$$P_{\theta_2}(X \geq y) < 1/n \tag{5}$$

by definition of Y_n^* and the fact that $y > Y_n^*$.

Since $\mathcal{F}$ is stochastically ordered in the tail we know that there exists an x_1 such that

$$F_{\theta_*}(x) \geq F_{\theta_2}(x) \tag{6}$$

for all $x \geq x_1$ and for all θ_* such that $\phi(\theta_*) \leq \phi_1$.

Let n be large enough so that $Y_n^* \geq x_1$.

Then, from (5) and (6) we have

$$P_{\theta_*}(X \geq y) < 1/n.$$

But this is true for all θ_* such that $\phi(\theta_*) = \phi_*$, $\phi_* < \phi_1$ and so by definition of $\underline{\phi}_n$ we have

$$\phi_1 \leq \underline{\phi}_n.$$

Thus

$$\{Y_n > Y_n^*\} \subset \{\phi_1 \le \underline{\phi}_n\} \quad \text{and, using (3),}$$

$$P(\phi_1 \le \underline{\phi}_n) > 1 - \varepsilon/2.$$

Part 2.

Define. $Y_n^{**} = \sup\{y_n : P_{\theta_3}(X > y_n) > 1/n\}$.

Notice that $P_{\theta_3}(X > Y_n^{**}) \le 1/n$.

Consider $P_{\theta_0}(Y_n \le Y_n^{**})$.

Pick an x_2 such that

$$\frac{P_{\theta_0}(X \ge x)}{P_{\theta_3}(X > x)} < \delta_2 \quad \text{for all} \quad x \ge x_2$$

where $\delta_2 = \varepsilon/2$.

Pick n large enough so that $Y_n^{**} \ge x_2$. Then

$$\frac{P_{\theta_0}(X \ge Y_n^{**})}{P_{\theta_3}(X > Y_n^{**})} < \delta_2.$$

$$P_{\theta_0}(X \ge Y_n^{**}) < \delta_2 P_{\theta_3}(X > Y_n^{**}) \le \delta_2 P_{\theta_3}(X \ge Y_n^{**})$$

$< \delta_2/n$. Then

$$P_{\theta_0}(X < Y_n^{**}) > 1 - \delta_2/n, \quad \text{and}$$

$$P_{\theta_0}(Y_n < Y_n^{**}) > (1 - \delta_2/n)^n > 1 - \delta_2 = 1 - \varepsilon/2. \tag{7}$$

Now show that

$$\{Y_n < Y_n^{**}\} \subset \{\bar{\phi}_n \le \phi_4\}. \tag{8}$$

Consider any $Y_n = y < Y_n^{**}$ and any θ^* such that $\phi(\theta^*) = \phi^* > \phi_4$ and show that

$$P_{\theta_0}(X > y) > 1/n.$$

Pick some small $\Delta > 0$ and

Define. $y_\Delta = \max(y, Y_n^{**} - \Delta)$.

By definition of Y_n^{**} we know that

$$P_{\theta_3}(X > y_\Delta) > 1/n, \text{ since } y_\Delta < Y_n^{**}.$$

Since $\mathcal{F}$ is stochastically ordered in the tail we know that there exists an x_3 such that

$$F_{\theta_3}(x) \geq F_{\theta^*}(x) \quad \text{for all} \quad x \geq x_3 \quad \text{and for all}$$

θ^* such that $\phi(\theta^*) \geq \phi_3$.

Let n be large enough so that $Y_n^{**} - \Delta \geq x_3$. Then

$$P_{\theta^*}(X > y) \geq P_{\theta^*}(X > y_\Delta) \geq P_{\theta_3}(X > y_\Delta) > 1/n,$$

since $y_\Delta < Y_n^{**}$.

But this is true for all θ^* such that $\phi(\theta^*) > \phi_4$ and so by definition of $\bar{\phi}_n$ we have

$$\bar{\phi}_n < \phi_4.$$

Thus

$$\{Y_n < Y_n^{**}\} \subset \{\bar{\phi}_n \leq \phi_4\} \quad \text{and, using (7)}$$

$$P(\bar{\phi}_n \leq \phi_4) > 1 - \varepsilon/2.$$

Finally, for n large enough

$$P(\phi_0 - \gamma < \phi_1 \leq \underline{\phi}_n \leq \bar{\phi}_n \leq \phi_4 < \phi_0 + \gamma) > 1 - \varepsilon$$

and so

$$P(|\hat{\phi}_n - \phi_0| < \gamma/2) > 1 - \varepsilon \quad \text{for } n \text{ sufficiently large.} \quad \square$$

Comment. The condition (2) in Theorem 1 is sufficient but it is not necessary. This is shown in the following:

Example 1. Let F_θ be a location family of geometric distributions defined by

$$P_\theta(X = k + \theta) = (1/2)^k, \ k = 1,2,3,\ldots; \ \theta \in (0,1).$$

Clearly $Y_n(\text{mod } 1)$, where Y_n is the maximum of n independent observations from distribution F_θ, is consistent for θ, but the condition (2) fails to hold here.

Comment. The condition (2) is equivalent to

$$\frac{P_{\theta_1}(X > x)}{P_{\theta_2}(X > x)} \longrightarrow 0 \text{ as } x \longrightarrow \infty \tag{8}$$

for all θ_1, θ_2 such that $\phi(\theta_1) = \phi_1$, $\phi(\theta_2) = \phi_2$ in the case that the distributions considered are continuous. Condition (8) is not sufficient in the case of discrete distributions, however. This is shown in the following:

Example 2. For $\theta \in (0,1)$, define F_θ by

$$P_\theta(X = -1) = 1 - \exp(-2\theta),$$

$$P_\theta(X = k) = \exp(-(k+2\theta)(k+1)) - \exp(-(k+1+2\theta)(k+2))$$

for $k = 0,1,2,\ldots$

This family satisfies condition (8) and the conditions of Theorem 1 except (2) but it does not admit a consistent estimator of θ based on the maxima.

The families of distributions which satisfy the conditions of Theorem 1 include normal distributions with fixed, known mean; normal distributions with the means lying in some known, finite interval; gamma distributions with fixed, known location and scale; and Poisson distributions.

While maxima are often more interesting than minima Theorem 1 obviously may be applied to minima as well as maxima.

3. Sample Size Unknown and Monotone Random Truncation

Procedures based on extremes have the marked disadvantage that except in a few unusual cases they are inefficient, usually very inefficient. The distributions of extremes change very slowly with sample size. As a consequence, estimation based on extremes may be little affected by errors in sample size or by systematically missing observations. This is made precise in the following two theorems.

DEFINITION 2. Let n_k be the value of the sample size used in the estimation when k is the true sample size. Then the sample size will be said to be known within bounded relative error if there exists a finite $\beta > 1$ such that

$$1/\beta \le n_k/k \le \beta \quad \text{for all} \quad k.$$

THEOREM 2. If a consistent estimator of a parameter based on the maxima exists then this estimator will remain consistent if the sample size is only known within bounded relative error.

The proof of this theorem and the next will rely on the following:

LEMMA 1. Let $\{P_k\}$ and $\{Q_k\}$ be sequences of probability measures on the measurable space $\{X,\mathcal{A}\}$ and let $\{A_k\}$ be a sequence of events in $\mathcal{A}$. Assume that P_k is absolutely continuous with respect to Q_k for all k and let $f_k(x)$ be the Radon-Nikodym derivative of P_k with respect to Q_k. Further, assume that $Q_k(A_k) \longrightarrow 1$ as $k \longrightarrow \infty$. Then $P_k(A_k) \longrightarrow 1$ as $k \longrightarrow \infty$ if for each $\varepsilon > 0$ there exists a finite number b, a sequence of sets $\{C_k\}$, $C_k \in \mathcal{A}$ and an integer k_0 such that $\{x : f_k(x) \geq b\} \subset C_k$ and $P_k(C_k) < \varepsilon$ for all $k \geq k_0$.

PROOF OF LEMMA 1. Assume that

$$Q_k(A_k) \longrightarrow 1 \text{ as } k \longrightarrow \infty.$$

Pick an $\varepsilon > 0$ and let b, $\{C_k\}$ and k_0 satisfy the conditions of the lemma. Then

$$P_k(\bar{A}_k) < \varepsilon + bQ_k(\bar{A}_k) \quad \text{for} \quad k \geq k_0.$$

But $Q_k(\bar{A}_k) \longrightarrow 0$ as $k \longrightarrow \infty$ by assumption. Since ε is arbitrary we have

$$P_k(A_k) \longrightarrow 1 \text{ as } k \longrightarrow \infty. \quad \square$$

Now we return to the proof of the theorem.

PROOF OF THEOREM 2. Let the value we desire to estimate be ϕ_0 and let $\hat{\phi}_n \longrightarrow \phi_0$ as $n \longrightarrow \infty$ where $\hat{\phi}_n = \hat{\phi}_n(n, Y_n)$, where Y_n represents the maximum of n i.i.d. r.v.'s.

Consider two cases: (1) $n_k < k$, and (2) $n_k > k$.

Case 1. $n_k < k$.

Here P_k is the distribution of Y_{n_k}, the maximum of n_k independent variables having distribution F, and Q_k is the distribution of Y_k. $f_k(x) = dP_k(x)/dQ_k(x)$.

Notice that $f_k(x)$ is monotone non-increasing in x. Pick an $\varepsilon > 0$.

Define. $x_0 = \sup\{x : F(x)^{n_k} \leq \varepsilon/2\}$.

a) Assume first that x_0 is a continuity point of F. For $x \geq x_0$

$$f_k(x) \leq n_k F(x_0)^{n_k}/kF(x_0)^k = n_k F(x_0)^{(n_k-k)/k}$$

$$\leq (n_k/k)(2/\varepsilon)^{\beta-1} < (2/\varepsilon)^{\beta-1} \quad \text{finite since}$$

$F(x_0)^{n_k} \geq \varepsilon/2$ and $1 < k/n_k \leq \beta$. Also

$$P_k\{x : f_k(x) > (2/\varepsilon)^{\beta-1}\} \leq F(x_0)^{n_k} \leq \varepsilon/2.$$

b) Assume now that x_0 is a discontinuity point.

For $x > x_0$ $f_k(x)$ will be bounded by $(2/\varepsilon)^{\beta-1}$ by the same argument as in a).

Consider

$$f_k(x_0) = \frac{F(x_0)^{n_k} - F(x_0-)^{n_k}}{F(x_0)^k - F(x_0-)^k}$$

$$< F(x_0)^{n_k-k} \leq (2/\varepsilon)^{\beta-1} \quad \text{finite.}$$

Again

$$P_k\{x : f_k(x) > (2/\varepsilon)^{\beta-1}\} \leq F(x_0)^{n_k} \leq \varepsilon/2.$$

Case 2. $n_k > k$.

$$f_k(x) \leq n_k/k \leq \beta \quad \text{for all } x.$$

In either case the conditions of Lemma 1 hold. Existence of a consistent estimate $\hat{\phi}_n$ of ϕ_0 implies that $\hat{\phi}_n(Y_n)$ is eventually in any neighborhood of ϕ_0 with arbitrarily high probability less than one. For any neighborhood of ϕ_0, let $\{A_k\}$ be a sequence of sets such that $\hat{\phi}_n(Y_n)$ is in the neighborhood of ϕ_0 if $Y_n \in A_n$. Thus, for some such sequence $\{A_k\}$

$$Q_k(A_k) \longrightarrow 1 \text{ as } k \longrightarrow \infty$$

and this implies that

$$P_k(A_k) \longrightarrow 1 \text{ as } k \longrightarrow \infty.$$

Therefore $\hat{\phi}_n(Y_n)$ is consistent for ϕ_0 even if n is known only within bounded relative error. □

DEFINITION 3. A distribution F will be considered to be subject to monotone random truncation if there exists a monotone function $p(x)$ which will be the probability that a variable $X = x$ from F will be observable. The truncation will be said to be decreasing if $p(x)$ is non-decreasing.

If Y is the observable variable then Y will have the distribution G: $G(y) = P(Y \leq y) = \int_{-\infty}^{y} p(x)dF(x)/c$ where $c = \int p(x)dF(x)$.

Assume that $c > 0$ since if it isn't nothing is observable.

THEOREM 3. If there exists a consistent estimator $\hat{\phi}_n(Y_n)$ of the value ϕ_0 based on the maxima then this estimator will also be consistent under monotone (decreasing) random truncation.

PROOF. Let P_k be the distribution of the maximum of k independent variables each having distribution G, Q_k the distribution of the maximum of k independent variables each having distribution F. P_k is absolutely continuous with respect to Q_k for all k.

$$f_k(x) = dP_k(x)/dQ_k(x).$$

For x a continuity point of F

$$f_k(x) = (G(x)/F(x))^{k-1}(p(x)/c) \leq 1/c$$

finite for all x.

For x a discontinuity point of F

$$f_k(x) = \frac{G(x)^k - G(x-)^k}{F(x)^k - F(x-)^k} \leq 1/c$$

finite for all x.

The conditions of Lemma 1 are thus met and the consistency of $\hat{\phi}_n(Y_n)$, where the variables for which Y_n is the maximum are drawn from G, follows from the same argument with which the proof of Theorem 2 was concluded. □

Acknowledgment

My thanks to Professor D. V. Gokhale for some helpful suggestions.

References

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Ann. of Math. 44, 423-453.