Theory of Stochastic Processes
Vol. 15 (31), no. 2, 2009, pp. 156–161

S. Y. NOVAK

LOWER BOUNDS TO THE ACCURACY OF SAMPLE MAXIMUM ESTIMATION

We derive lower bounds for sup-norm losses of estimators of the distribution function of a sample maximum, and show that their consistent estimation in a general situation is impossible.

1. Introduction

The asymptotic behavior of the sample maximum

$$M_n = \max_{1\le i\le n} X_i\,,$$

where $X_1, ..., X_n$ is a sample of independent copies of a random variable X with the distribution function (d.f.) F plays the central role in extreme value theory (see Embrechts et al. [4], Galambos [6], Leadbetter et al. [8], among others). The important question is whether the distribution function, F^n, of the sample maximum can be consistently estimated from a sample of independent and identically distributed random variables (r.v.s).

An arbitrary d.f. $\hat F_n(\cdot) \equiv \hat F_n(\cdot, X_1, ..., X_n)$ is called an estimator of F^n. An estimator $\hat F_n$ is consistent if $\|\hat F_n - F^n\| \xrightarrow[p]{} 0$ as $n \to \infty$, where $\|\cdot\|$ is the sup–norm. An estimator $\hat F_n$ is called L_1–consistent if $\mathbb{E}_F\|\hat F_n - F^n\| \to 0$ as $n \to \infty$.

The celebrated Glivenko–Cantelli theorem states that the empirical distribution function F_n is a consistent estimator of the unknown distribution function F. This could make one expect that F_n^n approximates F^n. In fact, F_n^n would be a poor approximation to F^n. Indeed, note that $\sup_{0\le p\le 1} |(1-p)^n - e^{-np}| \to 0$ as $n \to \infty$ by Taylor's formula. Therefore, $\|F^n - e^{-n(1-F)}\| \to 0$ and $\|F_n^n - e^{-n(1-F_n)}\| \to 0$ as $n \to \infty$ (cf. a remark in O'Brien [13]). Thus, F_n^n approximates F^n only if $n(1 - F_n(x)) = \sum_{i=1}^n \mathbb{I}\{X_i \ge x\}$ approximates $n\mathbb{P}(X \ge x)$ for all x. However,

$$1 - F_n(x) - \mathbb{P}(X \ge x) = \zeta_n(x)/\sqrt{n}\,,$$

where $\zeta_n(x) \Rightarrow \mathcal{N}(0; F(x)(1 - F(x)))$ by the central limit theorem. Hence $\|F_n^n - F^n\|$ does not converge to 0 in probability.

According to Beirlant & Devroye [3], for any estimator $\{\hat F_n\}$ of F^n there exists a d.f. F on $[2;\infty)$ such that

$$\limsup_{n\to\infty} \mathbb{E}_F\|\hat F_n - F^n\| \ge 1/2e^3 \tag{1}$$

(i.e., there exists an infinite sequence $\{k_n\}$ of natural numbers such that $\mathbb{E}_F\|\hat F_{k_n} - F^{k_n}\| \ge 1/2e^3$). Thus, there are no L_1-consistent estimators of F^n.

Note that one can derive a lower bound for $\limsup_{n\to\infty} \mathbb{P}_F(\|\hat F_n - F^n\| \ge c)$ where $c > 0$, from (1): by the Paley–Zygmund inequality [14],

$$\mathbb{P}(X \ge \theta\mathbb{E}X) \ge (1-\theta)^2(\mathbb{E}X)^2/\mathbb{E}X^2 \qquad (\forall\theta \in [0;1]) \tag{2}$$

2000 *Mathematics Subject Classification.* Primary 62G32.
Key words and phrases. Sample maximum, consistent estimation, lower bounds.
Supported by the EPSRC grant EP/H006613.

if $X \geq 0$ and $\mathbb{E}X < \infty$. As $\|\hat{F}_n - F^n\| \leq 1$, (1), (2), and Chebyshev's inequality yield

$$\limsup_{n\to\infty} \mathbb{P}_F(\|\hat{F}_n - F^n\| \geq \theta/2e^3) \geq (1-\theta)^2/4e^6.$$

Choosing, for instance, $\theta = 1/3$, we get

$$\mathbb{P}_F(\|\hat{F}_{n'} - F^{n'}\| \geq 1/6e^3) \geq 1/9e^6 \tag{3}$$

for an infinite sequence $\{n'\}$ of natural numbers.

A sharper bound is valid for $\sup_F \mathbb{P}_F(\|\hat{F}_n - F_i^n\| \geq 1/4)$ [12]:

$$\sup_F \mathbb{P}_F\Big(\|\hat{F}_n - F^n\| \geq 1/4\Big) \geq 1/4 \qquad (n \geq 1) \tag{4}$$

for any estimator $\{\hat{F}_n\}$ of the distribution function of the sample maximum.

In this paper, we strengthen (1) and (3). We establish also a sharp result for scale-invariant estimators.

2. Results

Theorem 1. *For any estimator $\{\hat{F}_n\}$ of the distribution function of the sample maximum, there exists a d.f. F such that*

$$\limsup_{n\to\infty} \mathbb{P}_F\Big(\|\hat{F}_n - F^n\| \geq 1/9\Big) \geq 1/3. \tag{5}$$

Chebyshev's inequality and (5) imply that there exists a d.f. F such that

$$\limsup_{n\to\infty} \mathbb{E}_F\|\hat{F}_n - F^n\| \geq 1/27.$$

An estimator $\tilde{a}_n(\cdot)$ is called scale-invariant if

$$\tilde{a}_n(x, x_1, ..., x_n) = \tilde{a}_n(cx, cx_1, ..., cx_n)$$

for all $x, x_1, ..., x_n, c > 0$. Examples of scale-invariant estimators of F^n include F_n^n, where F_n is the empirical distribution function, and the "blocks" estimator $\tilde{F}_n = \sum_{i=1}^{[n/r]} \mathbb{I}\{M_{i,r} < x\}/[n/r]$, where $M_{i,r} = \max\{X_{(i-1)r+1}, ..., X_{ir}\}$ $(1 \leq r \leq n)$.

Theorem 2. *For any scale-invariant estimator $\{\tilde{F}_n\}$ of the distribution function of the sample maximum,*

$$\mathbb{P}_{F_0}\Big(\|\tilde{F}_n - F_0^n\| \geq 1/4\Big) \geq 1/4 \qquad (n \geq 1), \tag{6}$$

where F_0 is the uniform d.f. on $[0;1]$.

Theorems 1 and 2 indicate that the consistent estimation of the distribution function of the sample maximum is possible only under certain assumptions on the unknown distribution. One approach is to use the fact that there are three types of limit laws for F^n [6, 7, 8, 4] if

$$\lim_{n\to\infty} \mathbb{P}(X > x)/\mathbb{P}(X \geq x) = 1. \tag{7}$$

If (7) holds, then there exist sequences $\{b_n\}$ and $\{c_n\}$ such that $(M_n - c_n)/b_n$ converges weakly as $n \to \infty$ [4, 8]. Under this assumption, a consistent estimator of the limiting d.f. for $(M_n - c_n)/b_n$ can be suggested (see, e.g., Athreya & Fukuchi [1]); a similar result for the case of a stationary sequence $\{X_i, i \geq 1\}$ is given in [2].

Our conjecture is that the estimation accuracy can be arbitrarily poor if (7) is the only restriction on the class of possible distributions: for an arbitrary estimator $\{\hat{F}_n\}$ of the distribution function of the sample maximum and any sequence $\varepsilon_n \downarrow 0$, there exists a d.f. F obeying (7) such that $\limsup_{n\to\infty} \mathbb{E}_F\|\hat{F}_n - F^n\| \geq \varepsilon_n$.

3. Proofs

Our approach incorporates some ideas from [3] and [12]. Suppose we want to estimate a functional a_F of the unknown distribution function F. We assume that a_F is an element of a normed space of real-valued functions defined on $\mathbb{R}$ or on an interval in $\mathbb{R}$. Examples include $a_F = F$, where $F(x) = \mathbb{P}(X < x)$ is a distribution function, $a_F = f$, where $f = F'$ is a density, $a_{F_\theta} = \theta$, where F_θ is an element of a parametric family of distributions $\{F_\theta, \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, etc..

Denote, by $\mathcal{A}$, the class of "scale-preserving" functionals: $a \in \mathcal{A}$ if $a_{F_1}(x) = a_F(cx)$ $(\forall x)$, where $F_1(\cdot) = F(c\cdot)$, $c > 0$. Among the elements of $\mathcal{A}$ are the tail index and the tail constant of a heavy-tailed distribution and the distribution function of the sample maximum [11].

Let $\hat{a}_n$ be an arbitrary estimator of a_F. If a_F belongs to a specified class of functions, then $\hat{a}_F$ is presumed to be an element of the same class. When we use a subscript to denote a particular distribution (say, $\mathbb{P}_*$), symbols F and f with the same subscript (say, F_* and f_*) denote the corresponding distribution function and its density.

Given two distribution functions F_0 and F_1, we put $\mathbb{P}_i = \mathbb{P}_{F_i}$,

$$s_n = \|a_{F_0} - a_{F_1}\|\,, \tag{8}$$

and let

$$d_n = \sup_{A \in \mathcal{B}} |\mathbb{P}_0\left((X_1, ..., X_n) \in A\right) - \mathbb{P}_1\left((X_1, ..., X_n) \in A\right)|$$

denote the total variation distance between $\mathbb{P}_0^n$ and $\mathbb{P}_1^n$ (the supremum is taken over the class $\mathcal{B}$ of Borel sets).

Lemma 3. *If $a \in \mathcal{A}$, then for any scale-invariant estimator $\{\tilde{a}_n\}$, d.f. F_0 and $c > 0$,*

$$\mathbb{P}_0(\|\tilde{a}_n - a_{F_0}\| \geq s_n/2) \geq (1 - d_n)/2 \qquad (n \geq 1)\,, \tag{9}$$

where s_n is given by (8), and $F_1(\cdot) = F_0(c\,\cdot)$.

According to the well-known property of the total variation distance,

$$1 - d_n \geq (1 - d_1)^n\,,$$

Chebyshev's inequality and (9) imply that

$$\mathbb{E}_{F_0}\|\hat{a}_n - a_{F_0}\| \geq s_n(1 - d_1)^n/4. \tag{10}$$

The right-hand side of (10) hints that, in a sense, it is "optimal" to choose $(\mathbb{P}_0, \mathbb{P}_1)$ obeying $s_n \sim (1 - d_1)^n$.

Proof of Lemma 3. Let F_0 and F_1 be two d.f.s. Note that

$$\mathbb{P}_0(\|\hat{a}_n - a_{F_0}\| \geq s_n/2) + \mathbb{P}_1(\|\hat{a}_n - a_{F_1}\| \geq s_n/2) \geq 1 - d_n \tag{11}$$

for any $\hat{a}_n$ and $n \geq 1$.

Indeed, suppose that the supremum $\|a_{F_0} - a_{F_1}\|$ is achieved at a point x_0 and $s_n = a_{F_0}(x_0) - a_{F_1}(x_0)$. Since $\|\hat{a}_n - a_{F_0}\| \geq a_{F_0}(x_0) - \hat{a}_n(x_0)$, we have

$$\begin{aligned}&\mathbb{P}_0\left(\|\hat{a}_n - a_{F_0}\| \geq u + s_n\right) = \mathbb{P}_0\left(\|\hat{a}_n - a_{F_0}\| \geq u + a_{F_0}(x_0) - a_{F_1}(x_0)\right)\\ &\geq \mathbb{P}_0\left(a_{F_1}(x_0) - \hat{a}_n(x_0) \geq u\right) \geq \mathbb{P}_1\left(-u \geq \|\hat{a}_n - a_{F_1}\|\right) - d_n\,.\end{aligned}$$

Choosing $u = -s_n/2$, we get

$$\mathbb{P}_0\left(\|\hat{a}_n - a_{F_0}\| \geq s_n/2\right) + \mathbb{P}_1\left(\|\hat{a}_n - a_{F_1}\| > s_n/2\right) \geq 1 - d_n\,.$$

A similar argument applies if $s_n = a_{F_1}(x_0) - a_{F_0}(x_0)$.

In a general situation, for every $\varepsilon > 0$ there exists $x_0 = x_0(\varepsilon)$ such that $s_n \leq a_{F_0}(x_0) - a_{F_1}(x_0) + \varepsilon$ or $s_n \leq a_{F_1}(x_0) - a_{F_0}(x_0) + \varepsilon$. We can repeat our argument with u replaced by $u + \varepsilon$. Since the events

$$A_\varepsilon = \{\|\hat{a}_n - a_{F_0}\| \geq u + \varepsilon\} \text{ and } B_\varepsilon = \{\|\hat{a}_n - a_{F_1}\| > -u - \varepsilon\}$$

are monotone in ε, we have

$$\mathbb{P}_0(A_\varepsilon) \to \mathbb{P}_0(A_0)\,,\ \mathbb{P}_1(B_\varepsilon) \to \mathbb{P}_1(B_0)$$

as $\varepsilon \to 0$. Hence, (11) holds.

If $\tilde{a}_n$ is a scale-invariant estimator, $a \in \mathcal{A}$ and $F_1(x) = F_0(cx)$ $(\forall x)$, then

$$\begin{aligned} & \mathbb{P}_1\left(\|\tilde{a}_n - a_{F_1}\| \geq s_n/2\right) \\ = & \int \ldots \int 1\!\!1\{\sup_x |\tilde{a}_n(x, x_1, ..., x_n) - a_{F_0}(cx)| \geq s_n/2\} dF_0(cx_1)...dF_0(cx_n) \\ = & \int \ldots \int 1\!\!1\{\sup_y |\tilde{a}_n(y/c, y_1/c, ..., y_n/c) - a_{F_0}(y)| \geq s_n/2\} dF_0(y_1)...dF_0(y_n) \\ = & \mathbb{P}_0\left(\|\tilde{a}_n - a_{F_0}\| \geq s_n/2\right)\,. \end{aligned}$$

This and (11) yield (9). The proof is complete. □

Proof of Theorem 1. Let $X_t = 2W + t_W$, where $t \in [0;1]$, $\mathbb{P}(W = m) = 2^{-m}$ $(m \geq 1)$ and $t_k \in \{0;1\}$ is the k^{th} element of the binary expansion $t = \sum_{k \geq 1} t_k 2^{-k}$ [3]. Denote by $\mathbb{P}_t$ and F_t the distribution and the d.f. of X_t. Distribution function F_t in (5) is a member of the parametric family $\{F_t\}_{t \in [0;1]}$. By construction,

$$\begin{aligned} &\mathbb{P}_t(2m) = 2^{-m} \qquad (t \in A_m), \quad \mathbb{P}_t(2m) = 0 \qquad (t \in B_m), \\ &\mathbb{P}_t(2m+1) = 0 \qquad (t \in A_m), \quad \mathbb{P}_t(2m+1) = 2^{-m} \quad (t \in B_m), \end{aligned}$$

where

$$\begin{aligned} A_m &= [0; 1/2^m) \cup \ldots \cup [(2^m - 2)/2^m; (2^m - 1)/2^m), \\ B_m &= [1/2^m; 2/2^m) \cup [(2^m - 1)/2^m; 1]. \end{aligned}$$

Let $k \equiv k(n) = [k_n]$ and $c \equiv c(n) = ((1 - 2^{-k})^n - (1 - 2^{-k+1})^n)/2$, where

$$k_n = -\log_2(1 - (2/3)^{1/n}) = \log_2(n/(\ln 1.5)) + o(1)\,.$$

Then

$$\begin{aligned} & \int_0^1 \mathbb{P}_t(|\hat{F}_n(2k+1) - F_t^n(2k+1)| \geq c)\, dt \\ = & \int_0^1 \sum_{m_1, \ldots, m_n} 1\!\!1\{|\hat{F}_n(2k+1) - F_t^n(2k+1)| \geq c\} \mathbb{P}_t(m_1)...\mathbb{P}_t(m_n)\, dt \\ = & \sum_{m_1, \ldots, m_n} \left(\int_{A_k} dt + \int_{B_k} dt \right) 1\!\!1\{|\hat{F}_n(2k+1) - F_t^n(2k+1)| \geq c\} \mathbb{P}_t(m_1)...\mathbb{P}_t(m_n). \end{aligned}$$

If $t \in A_m$, then

$$\mathbb{P}_t(l) = \mathbb{P}_{t+2^{-m}}(l) \qquad (l \notin \{2m; 2m+1\}).$$

One can check also that

$$F_t(2k+1) = 1 - 2^{-k} \quad (t \in A_k)\,, \quad F_t(2k+1) = 1 - 2^{-k+1} \quad (t \in B_k).$$

Denote $\bar m = (m_1, ...m_n), D_n = \{\bar m : m_i \notin \{2k; 2k+1\}\ (\forall i \le n)\}$. Then

$$\int_0^1 \mathbb{P}_t(\|\hat F_n - F_t^n\| \ge c)\,dt \ge \int_{A_k} dt \sum_{\bar m \in D_n} \mathbb{P}_t(m_1)...\mathbb{P}_t(m_n)$$
$$\left(\mathbb{I}\{|\hat F_n(2k+1) - (1-2^{-k})^n| \ge c\} + \mathbb{I}\{|\hat F_n(2k+1) - (1-2^{-k+1})^n| \ge c\}\right).$$

Using the triangle inequality, we derive

$$\int_0^1 \mathbb{P}_t(\|\hat F_n - F_t^n\| \ge c)\,dt \ge \int_{A_k} \sum_{\bar m \in D_n} \mathbb{P}_t(m_1)...\mathbb{P}_t(m_n)\,dt$$
$$= \int_{A_k} \mathbb{P}_t((X_1, ..., X_n) \in D_n) dt = (1-2^{-k})^n \int_{A_k} dt = (1-2^{-k})^n/2 = 1/3\,. \tag{12}$$

Note that

$$(1-2^{-k})^n - (1-2^{-k+1})^n \ge (1-2^{-k})^n - (1-2^{-k})^{2n} = 2/9\,.$$

Hence $c \ge 1/9$, and (12) yields

$$\int_0^1 \mathbb{P}_t(\|\hat F_n - F_t^n\| \ge 1/9)\,dt \ge 1/3 \qquad (n \ge 1).$$

By Fatou's lemma,

$$\int_0^1 \limsup_{n\to\infty} \mathbb{P}_t(\|\hat F_n - F_t^n\| \ge 1/9)\,dt \ge$$
$$\limsup_{n\to\infty} \int_0^1 \mathbb{P}_t(\|\hat F_n - F_t^n\| \ge 1/9)\,dt \ge 1/3.$$

Hence there exists a d.f. F such that (5) holds. The proof is complete. □

Proof of Theorem 2. Let $\{\tilde F_n\}$ be a scale-invariant estimator. We will construct two d.f.s F_0 and $F_1 \equiv F_{1,n}$ such that both s_n and d_n are bounded away from zero. One of those two d.f.s, F_0, can be chosen almost arbitrarily, while $F_1 \equiv F_{1,n}$ is a modification of F_0 such that d_1 and $\|F_0 - F_1\|$ decay like $1/n$ as $n \to \infty$.

We put

$$F_0(x) = x \ \text{ as } \ 0 \le x \le 1\,, \quad F_1(x) = x2^{-1/n} \ \text{ as } \ 0 \le x \le 2^{1/n}\,. \tag{13}$$

With $a_{P_i} = F_i^n$, we have

$$d_1 = 1 - 2^{-1/n}\,,\ s_n = 1/2\,. \tag{14}$$

Note that $s_n = (1-d_1)^n$. Combining (14) and (9), we derive (6). The proof is complete. □

Acknowledgements. The author is grateful to the referee for careful reading and helpful remarks.

References

1. K.B. Athreya and J. Fukuchi (1994) *Bootstrapping extremes of i.i.d. random variables.* In: Proc. Conf. Extreme Value Theory Appl. (J. Galambos, J. Lechner, Simiu E., eds.), v. 3, NIST 866.
2. K.B. Athreya, J. Fukuchi, S.N. Lahiri (1999) *On the bootstrap and the moving block bootstrap for the maximum of a stationary process.* — J. Statist. Plan. Infer., v. 76, 1–17.
3. J. Beirlant and L. Devroye (1999) *On the impossibility of estimating densities in the extreme tail.* — Statist. Probab. Letters, v. 43, No 1, 57–64.
4. P. Embrechts, C. Klüppelberg, and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer.

5. W. Feller (1971) *An introduction to probability theory and its applications.* — New York: Wiley.
6. J. Galambos (1987) *The asymptotic theory of extreme order statistics.* — Melbourne: Krieger.
7. B.V. Gnedenko (1943) *Sur la distribution du terme maximum d'une série aléatoire.* — Ann. Math., v. 44, 423–453.
8. M.R. Leadbetter, G. Lindgren and H. Rootzén (1983) *Extremes and related properties of random sequences and processes.* — New York: Springer.
9. S.Y. Novak (1999) *Generalized kernel density estimator.* — Theory Probab. Appl., v. 44, No 3, 634–645.
10. S.Y. Novak (2002) *Inference on heavy tails from dependent data.* — Siberian Adv. Math., v. 12, No 2, 73–96.
11. S.Y. Novak (2009) *Advances in Extreme Value Theory with Applications to Finance.* In: "New Business and Finance Research Developments", 199–251, New York: Nova Science.
12. S.Y. Novak (2009) *On impossibility of consistent estimation of the distribution function of a sample maximum.* — Statistics (to appear).
13. G.L. O'Brien (1987) *Extreme values for stationary and Markov sequences.* — Ann. Probab., v. 15, No 1, 281–291.
14. R.E. Paley and A. Zygmund (1932) *A note on analytic functions in the unit circle.* — *Proc. Camb. Phil. Soc., v. 28, 266–272.*
15. J. Pfanzagl (2000) *On local uniformity for estimators and confidence limits.* — J. Statist. Plann. Inference, v. 84, 27–53.

Middlesex University, MUBS, The Burroughs, London NW44BT, UK