BOUNDARY CROSSING PROBABILITIES FOR THE KOLMOGOROV-SMIRNOV STATISTICS[1]

By Richard M. Stanley

University of Arizona

That the Kolmogorov-Smirnov statistics obey iterated logarithm laws is well known. For the purpose of developing nonparametric tests with power one it has become of interest to find accurate upper bounds for the probability that a sequence of Kolmogorov-Smirnov statistics ever exceeds a given boundary sequence. This paper is concerned with finding such probability bounds for a wide class of boundary sequences.

1. Introduction. In their development of nonparametric tests with power one, Darling and Robbins [3] found upper bounds on the probabilities of boundary crossings by Kolmogorov-Smirnov statistics for a class of boundary sequences. In this paper bounds are given for wider classes of sequences. In particular these classes contain sequences which are $O[(n^{-1} \log \log n)^{\frac{1}{2}}]$. For sequences in which the bounds of [3] apply, the bounds of this paper are generally smaller.

The class of sequences for which probability bounds are found for the two-sample statistic contains all upper class sequences satisfying simple monotonicity conditions, and a criterion is given for the upper and lower classes under these conditions. This corresponds to a result of Chung [1] for the one-sample statistic. Unfortunately, the boundary sequences for which we are able to give bounds is not the entire upper class in the one sample case.

2. Statement of results. Let $X_1, X_2, \cdots$, and $Y_1, Y_2, \cdots$ be independent sequences of independent, identically distributed (i.i.d.) random variables. Assume, $P(X_1 \leq t) = P(Y_1 \leq t) = F(t)$ for all real t. Define the empirical distribution functions

$$F_X{}^n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq t\}},$$

and define $F_Y{}^n$ similarly. We will be concerned with the processes

$$U_n = n \sup_t (F_X{}^n(t) - F_Y{}^n(t)), \quad \text{and}$$
$$V_n = n \sup_t (F_X{}^n(t) - F(t)).$$

Let Φ_1 be the collection of positive real sequences $\phi = \{\phi_n : n = 1, 2, \cdots\}$ for which $n^{\frac{1}{2}}\phi_n$ is monotone non-decreasing, and let $\Phi_2 = \{\phi \in \Phi_1 : \phi_n \text{ is monotone non-decreasing and } n^{-\frac{1}{2}}\phi_n \text{ is monotone non-increasing}\}$.

664

THEOREM 1. *Suppose* $\{\phi_m, \phi_{m+1}, \cdots\} \in \Phi_1$. *If* $m = n_1 < n_2 < \cdots$ *is any integer sequence then*

$$p(m, \phi) = P\{U_n \geqq n^{\frac{1}{2}}\phi_n \quad \text{for some} \quad n \geqq m\}$$
$$\leqq 2 \sum_{i=1}^{\infty} \exp(-n_{i+1}\phi_{n_i}^2/n_i) \,.$$

Taking the subsequence defined by $n_1 = m$ and $n_i = \min\{n: \sum_{j=n_{i-1}}^{n-1} \phi_j^2/j \geqq \frac{1}{2}\}$, some simple calculations yield the following corollary.

COROLLARY. *Suppose* $\{\phi_m, \phi_{m+1}, \cdots\} \in \Phi_2$. *Then*

$$(1) \qquad p(m, \phi) \leqq 4 \exp(1 + 2\phi_m^2/m) \sum_{n=m}^{\infty} n^{-1}\phi_n^2 \exp(-\phi_n^2) \,.$$

Considering the result of Chung [1], it is not unexpected that convergence of (1) is a criterion for upper class sequences in Φ_2.

THEOREM 2. *Suppose* $\{\phi_m, \phi_{m+1}, \cdots\} \in \Phi_2$ *for some* m *and* $\sum_{n=m}^{\infty} n^{-1}\phi_n^2 \times$ $\exp(-\phi_n^2) = \infty$. *Suppose also that* F *is continuous. Then* $P\{U_n > n^{\frac{1}{2}}\phi_n$ *for infinitely many* $n\} = 1$.

My best result for the one–sample statistic is:

THEOREM 3. *Suppose* $\{\phi_m, \phi_{m+1}, \cdots\} \in \Phi_2$ *with* $\sup_{n \geqq m} n^{-\frac{1}{2}}\phi_n^3 = b < \infty$. *Then for any integer sequence* $m = n_1 < n_2 < \cdots$

$$q(m, \phi) = P\{V_n \geqq n^{\frac{1}{2}}\phi_n \quad \text{for some} \quad n \geqq m\}$$
$$\leqq \sum_{i=1}^{\infty} \left\{ \left[1 + \frac{4\phi_{n_i}n_i^{\frac{1}{2}}}{n_{i+1} - 1}(1 + (2\pi n_{i+1})^{\frac{1}{2}})\right] \exp\left[-\left(\frac{4n_i\phi_{n_i}^2}{n_{i+1} - 1}\right)\right] \right.$$
$$\left. + \frac{16n_i\phi_{n_i}^2}{(n_{i+1} - 1)} \exp\left[-\frac{2n_i\phi_{n_i}^2}{n_{i+1} - 1} + \frac{16}{3}b \exp\left(\frac{8n_i^{\frac{1}{2}}\phi_{n_i}}{n_{i+1} - 1}\right)\right] \right\} \,.$$

The assumption $b < \infty$ is not a restrictive one; when $b = \infty$ this result may be applied to $\phi_n \wedge \alpha n^{\frac{1}{6}}$ in place of ϕ_n. However, smaller bounds may be available from [3] in such cases.

Finally, choosing $n_1 = m$ and $n_1 = \min\{n: \sum_{j=n_{i-1}}^{n-1} j^{-1}\phi_j^2 > \frac{1}{4}\}$, we can eliminate the subsequence.

COROLLARY. *Let* ϕ *be as in Theorem* 3. *Then*

$$(2) \qquad q(m, \phi) \leqq c(m, \Phi) \sum_{n=m}^{\infty} n^{-1}\phi_n^4 \exp(-2\phi_n^2) \,,$$

where

$$c(m, \phi) = 4e\{[1 + 4\phi_m m^{-\frac{1}{2}}(1 + (2\pi(m + 1))^{\frac{1}{2}}] \exp(\tfrac{1}{2} - 2\phi_m^2)$$
$$+ 16 \exp[16b \exp(8m^{-\frac{1}{2}}\phi_m)/3]\} \,.$$

Comparing this with Theorem 2 of [1], we see that there exist sequences $\phi \in \Phi_2$ for which $q(m, \phi) < 1$ but for which the series (2) does not converge. For example, if $\phi_n = (\frac{1}{2}\log_2 n + \alpha \log_3 n)^{\frac{1}{2}}$ then $q(m, \phi) < 1$ for m large if $\alpha > 1$, but the series in (2) converges only if $\alpha > \frac{3}{2}$.

3. Proof of Theorem 1. Let $k > 0$ be an integer. We will use a bound due to Darling and Robbins [3]:

$$(3) \qquad\qquad P\{U_n \geqq k\} \leqq \exp[-k^2/(n + 1)] .$$

Let $d_n(t) = n(F_X{}^n(t) - F_Y{}^n(t))$ and

$$\tau_n = \inf\{t : d_n(t) = U_n\} \vee (\min\{Y_1, \cdots, Y_n\} - 1) .$$

For any $N > n$, conditional on $X_1, \cdots, X_n, Y_1, \cdots, Y_n, d_N(\tau_n) - d_n(\tau_n)$ is symmetrically distributed. Hence, since $U_N \geqq d_N(\tau_n)$ and $d_n(\tau_n) = U_n$,

$$(4) \qquad\qquad P\{U_N \geqq U_n \mid X_1, \cdots, X_n, Y_1, \cdots, Y_n\} \geqq \tfrac{1}{2} .$$

From (4) we can use a reflection principle to obtain

$$(5) \qquad P\{U_n \geqq x \text{ for some } n \leqq m\} \leqq 2P\{U_m \geqq x\} \quad (m = 1, 2, \cdots, x > 0) .$$

Equations (3), (4), and (5) now give us

$$p(m, \phi) \leqq \sum_{i=1}^\infty P\{U_n \geqq n_i{}^{\frac{1}{2}} \phi_{n_i} \text{ for some } n_i \leqq n < n_{i+1}\}$$
$$\leqq 2 \sum_{i=1}^\infty P\{U_{n_{i+1}-1} \geqq n_i{}^{\frac{1}{2}} \phi_{n_i}\} \leqq 2 \sum_{i=1}^\infty \exp\{-n_i \phi_{n_i}^2/n_{i+1}\} ,$$

completing the proof.

Given the choice of $\{n_i\}$ outlined, the corollary follows from this if we observe

$$q(m, \phi) \leqq 2 \sum_{i=1}^\infty \sum_{n=n_i}^{n_{i+1}-1} 2n^{-1} \phi_n{}^2 \exp(-\phi_n{}^2 + \phi_n{}^2 - n_i \phi_{n_i}^2/n_{i+1}) ,$$

and

$$\phi_n{}^2 - n_i \phi_{n_i}^2/n_{i+1} \leqq \phi_{n_{i+1}}^2 \left(1 - \left(\frac{n_i}{n_{i+1}}\right)^2\right)$$

$$\leqq 2\phi_{n_{i+1}}^2 (n_{i+1} - n_i)/n_{i+1} \leqq 2 \sum_{n=n_i}^{n_{i+1}-1} \phi_n{}^2/n \leqq 1 + 2\phi_m{}^2/m .$$

4. A lower bound. We give here a lower bound corresponding to (3). Following Chung's example, I have omitted the proof of Theorem 2, but it will now follow by applying the methods of Feller [5].

LEMMA. *Suppose F is continuous. Then for any $0 < x \leqq n - 1$, $P\{U_n \geqq x\} \geqq$* $\exp[-(x + 1)^2/(n - x - \tfrac{1}{2})]$.

PROOF. In the derivation of (3) Darling and Robbins [3] show for any integer $k = 0, 1, \cdots, n$,

$$-\log P\{U_n \geqq k\} = \int_0^\infty 2 \exp[-(n + \tfrac{1}{2})t] \frac{\sinh^2(kt/2)}{t \sinh(t/2)} dt .$$

Using $y \leqq \sinh y \leqq ye^y$, valid for $y \geqq 0$, we find

$$-\log P\{U_n \geqq k\} \leqq k^2 \int_0^\infty \exp[-(n - k + \tfrac{1}{2})t] dt = \frac{k^2}{n - k + 1/2} .$$

The lemma now follows from $x < [x + 1] \leqq x + 1 \leqq n$.

5. Proof of Theorem 3. Just as in the proof of Theorem 1, let $d_n'(t) = n(F_X{}^n(t) - F(t))$ and $\tau_n' = \inf\{t : d_n'(t) = V_n\}$. Then $E(V_{n+1} \mid X_1, \cdots, X_n) \geq E(d_{n+1}'(\tau_n') \mid \tau_n') = d(\tau_n') = V_n$ and the V_n process is a submartingale. For $t > 0$, $\exp(tV_n)$ is a convex increasing function of a submartingale and therefore is itself a submartingale (Doob [4], page 295). Applying Kolmogorov's inequality for submartingales ([4], page 314) we may write

$$(6) \qquad q(m, \phi) \leq \sum_{i=1}^{\infty} P\{V_n \geq n_i{}^{\frac{1}{2}} \phi_{n_i} \text{ for some } n_i \leq n < n_{i+1}\}$$
$$\leq \sum_{i=1}^{\infty} E(\exp t_i V_{n_{i+1}-1}) \exp[-t_i n_i{}^{\frac{1}{2}} \phi_{n_i}] ,$$

for arbitrary positive numbers t_i. Caski [2] has given the following bound:

$$(7) \qquad E(\exp tV_n) \leq 1 + tE(V_n) + tn(1 + e^{-t})[\phi(t)]^n ,$$

where

$$\phi(t) = \frac{e^t - 1}{t} \exp\left(\frac{t}{e^t - 1} - 1\right).$$

Darling and Robbins [3] showed that, for integer $k > 0$,

$$P\{V_n \geq k\} \leq 2\sqrt{2} \exp[-k^2/(n + 1)] ,$$

so

$$E(V_n) = \int_0^n P\{V_n \geq y\} \, dy$$
$$(8) \qquad \leq 1 + \int_1^{\infty} 2\sqrt{2} \exp[-(y - 1)^2/(n + 1)] \, dy$$
$$= 1 + (2\pi(n + 1))^{\frac{1}{2}} .$$

For $\phi(t)$, note that $t > 0$ implies $t < e^t - 1$ so

$$\phi(t) \leq \frac{e^t - 1}{t}\left[1 + \left(\frac{t}{e^t - 1} - 1\right) + \tfrac{1}{2}\left(\frac{t}{e^t - 1} - 1\right)^2\right]$$
$$\leq 1 + \frac{1}{2t^2}\left(\sum_{k=2}^{\infty} \frac{t^k}{k!}\right)^2 \leq 1 + \frac{t^2}{8} + \frac{t^3}{12} e^{2t} ,$$

so

$$(9) \qquad [\phi(t)]^n \leq \exp\left(\frac{nt^2}{8} + \frac{nt^3}{12} e^{2t}\right).$$

Combining (6) through (9),

$$q(m, \phi) \leq \sum_{i=1}^{\infty}\left\{1 + t_i(1 + (2\pi n_{i+1})^{\frac{1}{2}}) + (n_{i+1} - 1)t_i{}^2\right.$$
$$\left. \times \exp\left[\frac{(n_{i+1} - 1)t_i{}^2}{8} + \frac{(n_{i+1} - 1)t_i{}^3}{12} e^{2t_i}\right]\right\} \exp(-t_i n_i{}^{\frac{1}{2}} \phi_{n_i}) .$$

The theorem now follows with the choice

$$t_i = 4n_i{}^{\frac{1}{2}} \phi_{n_i}/(n_{i+1} - 1) .$$

RICHARD M. STANLEY

REFERENCES

[1] CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

[2] CASKI, E. (1968). An iterated logarithm law for semimartingales and its application to empirical distribution functions. *Studia Sci. Math. Hungar.* **3** 287–292.

[3] DARLING, D. A. and ROBBINS, H. E. (1968). Some nonparametric tests with power one. *Proc. Nat. Acad. Sci.* **61** 804–809.

[4] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[5] FELLER, W. (1943). The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc.* **54** 373–402.