BOUNDARY CROSSING PROBABILITIES FOR
THE KOLMOGOROV-SMIRNOV STATISTICS

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 $O(n^{-1} \log \log n)^{1/2}$. 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 monotoncity 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 $X_1, X_2, \ldots, \text{ and } Y_1, Y_2, \ldots$ 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

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

and define $F_Y^*$ similarly. We will be concerned with the processes

$$U_n = n \sup_{t} (F_X^*(t) - F_Y^*(t)), \quad \text{and}$$

$$V_n = n \sup_{t} (F_X^*(t) - F(t)).$$

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

Received March 9, 1971; revised September 28, 1971.

1 This paper is a revised portion of the author’s doctoral dissertation which was written at Columbia University in 1970 under the direction of David O. Siegmund.
THEOREM 1. Suppose \( \{\phi_m, \phi_{m+1}, \ldots\} \in \Phi_1 \). If \( m = n_1 < n_2 < \ldots \) is any integer sequence then

\[
p(m, \phi) = P[U_n \geq n^4 \phi_n \; \text{for some} \; n \geq m] \leq 2 \sum_{i=1}^{\infty} \exp \left( -n_{i+1}^2 \phi_{n_i}^2 / n_i \right).
\]

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

COROLLARY. Suppose \( \{\phi_m, \phi_{m+1}, \ldots\} \in \Phi_2 \). Then

\[
p(m, \phi) \leq 4 \exp \left( 1 + 2 \phi_m^2 / m \right) \sum_{n=m}^{\infty} n^{-1} \phi_n^3 \exp \left( -\phi_n^2 \right).
\]

Considering the result of Chung [1], it is not unexpected that convergence of (1) is a criterion for upper class sequences in \( \Phi_1 \).

THEOREM 2. Suppose \( \{\phi_m, \phi_{m+1}, \ldots\} \in \Phi_2 \) for some \( m \) and \( \sum_{n=m}^{\infty} n^{-1} \phi_n^3 \times \exp \left( -\phi_n^2 \right) = \infty \). Suppose also that \( F \) is continuous. Then \( P[U_n > n^4 \phi_n \; \text{for infinitely many} \; n] = 1 \).

My best result for the one-sample statistic is:

THEOREM 3. Suppose \( \{\phi_m, \phi_{m+1}, \ldots\} \in \Phi_2 \) with \( \sup_{n \geq m} n^{-1} \phi_n^3 \leq b < \infty \). Then for any integer sequence \( m = n_1 < n_2 < \ldots \)

\[
q(m, \phi) = P[V_n \geq n^4 \phi_n \; \text{for some} \; n \geq m] \leq \sum_{i=1}^{\infty} \left[ 1 + \frac{4 \phi_n^2 n_i^4}{n_{i+1}^4 - 1} (1 + (2\pi n_{i+1})^4) \right] \exp \left[ -\frac{4 n_i^4 \phi_n^2}{n_{i+1}^4 - 1} \right]
\]

\[
+ \frac{16 n_i^2 \phi_n^2}{(n_{i+1}^4 - 1)} \exp \left[ -\frac{2 n_i^2 \phi_n^2}{n_{i+1}^4 - 1} + \frac{16 b}{3} \exp \left( \frac{8 n_i^4 \phi_n^2}{n_{i+1}^4 - 1} \right) \right].
\]

The assumption \( b < \infty \) is not a restrictive one; when \( b = \infty \) this result may be applied to \( \phi_n \wedge an \) in place of \( \phi_n \). However, smaller bounds may be available from [3] in such cases.

Finally, choosing \( n_i = m \) and \( n_1 = \min \{n: \sum_{j=n_1-1}^{n-1} f^{-1} \phi_j^2 > \frac{1}{2}\} \), we can eliminate the subsequence.

COROLLARY. Let \( \phi \) be as in Theorem 3. Then

\[
q(m, \phi) \leq c(m, \Phi) \sum_{n=m}^{\infty} n^{-1} \phi_n^4 \exp \left( -2 \phi_n^2 \right),
\]

where

\[
c(m, \phi) = 4e \left[ 1 + 4 \phi_n m^{-4} (1 + (2\pi (m + 1))^4) \exp \left( \frac{1}{2} - 2 \phi_n^2 \right) \right]
\]

\[
+ 16 \exp \left[ 16b \exp \left( 8m^{-1} \phi_n \right) / 3 \right].
\]

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 \( \phi_n = \frac{1}{2} \log_n n + \alpha \log_n n \frac{1}{\phi} \) 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]:

\[
P(U_n \geq k) \leq \exp \left[ -k^2/(n + 1) \right].
\]

Let \( d_n(t) = n(F_X^{-1}(t) - F_Y^{-1}(t)) \) and

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

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

\[
P(U_N \geq U_n | X_1, \ldots, X_n, Y_1, \ldots, Y_n) \geq \frac{1}{2}.
\]

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

\[
P(U_n \geq x \text{ for some } n \leq m) \leq 2P(U_m \geq x) \quad (m = 1, 2, \ldots, x > 0).
\]

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

\[
p(m, \phi) \leq \sum_{i=1}^{\infty} P(U_n \geq n_i \phi_{n_i} \text{ for some } n_i \leq n < n_{i+1})
\]

\[
\leq 2 \sum_{i=1}^{\infty} P(U_{n_{i+1}} \geq n_i \phi_{n_i}) \leq 2 \sum_{i=1}^{\infty} \exp \left\{ -n_i \phi_{n_i}^2/n_{i+1} \right\},
\]

completing the proof.

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

\[
q(m, \phi) \leq 2 \sum_{i=1}^{\infty} \sum_{n_{i+1}}^{n_{i+1}-1} 2n_i^{-1} \phi_{n_i}^2 \exp(-\phi_{n_i}^2 + \phi_{n_i}^2 - n_i \phi_{n_i}^2/n_{i+1})
\]

and

\[
\phi_{n_i}^2 - n_i \phi_{n_i}^2/n_{i+1} \leq \phi_{n_{i+1}}^2 \left( 1 - \left( \frac{n_i}{n_{i+1}} \right)^2 \right)
\]

\[
\leq 2 \phi_{n_{i+1}}^2 (n_{i+1} - n_i)/n_{i+1} \leq 2 \sum_{n_i=n_{i+1}}^{n_{i+1}-1} \phi_{n_i}^2/n \leq 1 + 2\phi_{n_i}^2/m.
\]

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 \( 0 < x \leq n - 1 \), \( P(U_n \geq x) \geq \exp \left[ -(x + 1)^2/(n - x - 1/2) \right] \).

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

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

Using \( y \leq \sinh y \leq ye^y \), valid for \( y \geq 0 \), we find

\[
- \log P(U_n \geq k) \leq k^2 \int_0^\infty \exp \left[ -(n - k + \frac{1}{2})t \right] dt = \frac{k^2}{n - k + 1/2}.
\]

The lemma now follows from \( x < [x + 1] \leq x + 1 \leq n \).
5. Proof of Theorem 3. Just as in the proof of Theorem 1, let \( d_n'(t) = n(F^{-1}_X(t) - F(t)) \) and \( \tau_n' = \inf \{ t : d_n'(t) = V_n \} \). Then \( E(V_{n+1} | X_1, \ldots, X_n) \geq E(d_n'(\tau_n') | \tau_n') = d(\tau_n') = V_n \) 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

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

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

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

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 \( k > 0 \),

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

so

\[
E(V_n) = \int_0^x P[V_n \geq y] dy \leq 1 + \int_1^x 2 \sqrt{2} \exp \left[-(y - 1)^2/(n + 1)\right] dy = 1 + (2\pi(n + 1))^{\frac{1}{2}}.
\]

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

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

so

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

Combining (6) through (9),

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

The theorem now follows with the choice

\[
t_i = 4n_t^i \phi_{n_i}/(n_{i+1} - 1).
\]

Acknowledgment. I am deeply indebted to David Siegmund for his help and guidance.
REFERENCES


