The Annals of Statistics
1985, Vol. 13, No. 2, 821–826

A UNIFORM BOUND FOR THE TAIL PROBABILITY OF KOLMOGOROV-SMIRNOV STATISTICS[1]

BY INCHI HU

Stanford University

Using an argument developed in Siegmund (1982), we give a bound for the tail probability of Kolmogorov-Smirnov statistics in the following form

$$P(\inf_x(F_n(x) - F(x)) > \zeta) \le 2\sqrt{2}e^{-2n\zeta^2}.$$

1. Introduction. Let $X_1, X_2, \cdots$ be independent, identically distributed random variables with a continuous but unknown distribution function F. Denote the empirical distribution function for sample $X_1, X_2, \cdots, X_n$ by $\hat{F}(x) = (1/n)\{\#\text{ of } X_i \le x, i = 1, \cdots, n\}$. In testing goodness of fit, that is, to test $F = F_0$ for some specific choice of F_0, the commonly used test statistics are

$$D_n^+ = \sqrt{n}\ \sup_x(\hat{F}_n(x) - F(x)), \quad D_n^- = \sqrt{n}\ \inf_x(\hat{F}(x) - F(x))$$
$$D_n = \sqrt{n}\ \sup_x|\hat{F}(x) - F(x)|.$$

The purpose of this paper is to give a bound for the tail probability of D_n^- in the following form.

THEOREM 1. $p\{D_n^- > \sqrt{n}\zeta\} \le 2\sqrt{2}e^{-2n\zeta^2}$.

A bound of the form $p\{D_n^- > \sqrt{n}\zeta\} \le Ce^{-2n\zeta^2}$, where C is some unspecified constant, has been proven by Dvoretzky, Kiefer, and Wolfowitz (1956). There are several papers conjecturing that C can be taken as 1, cf. Birnbaum and McCarty (1958) and Csörgő and Horváth (1981). Each of them is substantiated by considerable numerical computation, although no proof is available. Devroye and Wise (1979) proved $C \le \{2 + 32/(6\pi)^{1/2} + 8/3^{1/2} + 2^{1/2}4\exp(17/18)\} \le 306$, but this bound is too large to be useful in any application. The best result known to the author (before this paper was written) is $c \le 29$, due to G. Shorack (private communication), so the result of this paper is a substantial improvement of all the results known so far and partial support of the conjecture.

2. Proof of the main result. First we introduce some notation and basic facts about exponential families. Assume the distribution function F of X_1 can be imbedded in an exponential family, i.e. for all θ in some neighborhood of 0 $\exp[\psi(\theta)] = \int \exp(\theta x)\,F(dx)$ is finite, so $\exp[\theta x - \psi(\theta)]\,F(dx)$ defines a family of probability distributions indexed by θ. It is easy to show that the mean and variance of these distributions are given by $\psi'(\theta)$ and $\psi''(\theta)$ respectively. Hence

Received July 1984; revised October 1984.
[1] Research supported by the Office of Naval Research Contract N00014-77-C-0306 (NR-042-373).
AMS 1980 *subject classifications.* Primary 62E15, secondary 62G15.
Key words and phrases. Kolmogorov-Smirnov statistics, exponential family, random walk.

$\mu = \psi'(\theta)$ is a one to one function of θ. It will be convenient to regard this family of distributions as indexed by μ and write $F_\mu(dx) = \exp[\theta x - \psi(\theta)] F(dx)$. Let P_μ denote the probability according to which $X_1, X_2 \cdots$ are independent with $P_\mu(X_i \in dx) = F_\mu(dx)$ $(i = 1, 2, \cdots)$. The density of $S_n = X_1 + \cdots + X_n$ under P_μ will be denoted by $f_{\mu,n}$. If A is an event belonging to the σ-field generated by $X_1, \cdots X_m$, the following notation will be used: $P_\zeta^{(m)}(A) = P_\mu(A \mid S_m = m\zeta)$. In this paper we consider only events of the form $A = \{\tau < k\}$, $(k = 1, 2, \cdots m)$, where τ is a stopping time.

Siegmund (1982) derived the following fundamental identity

$$P_{\mu_0}^{(m)}(\tau < k) = \exp\{-m[(\theta_2 - \theta_0)\mu_0 + \psi(\theta_0) - \psi(\theta_2)]\} \cdot \int_{\{\tau<k\}} f_{\mu_2,m-\tau}(m\mu_0 - S_\tau)\exp[-(\theta_1 - \theta_2)S_\tau]/f_{\mu_0,m}(m\mu_0)\, dP_{\mu_1}. \tag{1}$$

The notation $\mu_i = \mu(\theta_i)$, $i = 0, 1, 2$ is used above, and θ_1, θ_2 satisfy $\psi(\theta_1) = \psi(\theta_2)$.

Let us bring our attention back to D_n^-. It is well known that the distribution of D_n^- is the same for all continuous distributions, so without loss of generality we may take F to be the uniform distribution on (0, 1). The well known representation of uniform order statistics in terms of sums of independent exponential random variables shows that

$$\begin{aligned} P\{D_n^- > \sqrt{n}\zeta\} &= P\{\sup_{0<x<1}(x - \hat{F}_n(x)) > \zeta\} \\ &= P\{\max_{1\le j\le n}(W_j - j) \ge n\zeta - 1 \mid W_{n+1} - (n+1) = -1\} \\ &= P_{\mu_0}^{(m)}\{\tau < m\} \end{aligned}$$

where $W_j = Y_1 + \cdots + Y_j$ and $Y_1, Y_2, \cdots$ are independent standard exponential, $m = n + 1$, $\mu_0 = (-1/m)$, $\tau = \inf\{i: W_i - i \ge n\zeta - 1\}$.

For reasons which will be indicated later, we divide the set $\{\tau < m\}$ into two parts $\{\tau \le n/2 + 1\} \cup \{n/2 + 1 < \tau < m\}$ and apply a time reversal argument to the later part, i.e.

$$\begin{aligned} P_{\mu_0}^{(m)}(\tau < m) &= P_{\mu_0}^{(m)}(\tau \le n/2 + 1) + P_{\mu_0}^{(m)}(n/2 + 1 < \tau < m) \\ &\le P_{\mu_0}^{(m)}(\tau \le n/2 + 1) + \tilde{P}_{\nu_0}^{(m)}(T < n/2) \end{aligned}$$

where $\nu_0 = 1/m$, $T = \inf\{i: S_i \ge n\zeta\}$ and under the probability $\tilde{P}$, S_i has the same distribution as $i - W_i$ $(i = 1, \cdots, n + 1)$. By (1) we have

$$\begin{aligned} &P_{\mu_0}^{(m)}\{\tau \le n/2 + 1\} \\ &\quad = \exp\{-m[(\theta_2 - \theta_0)\mu_0 + \psi(\theta_0) - \psi(\theta_2)]\} \\ &\qquad \cdot \int_{(\tau\le n/2+1)} f_{\mu_2,m-\tau}(m\mu_0 - S_\tau)\exp[-(\theta_1 - \theta_2)S_\tau]/f_{\mu_0,m}(m\mu_0)\, dP_{\mu_1} \end{aligned} \tag{2}$$

and

$$\tilde{P}_{\nu_0}^{(m)}(T < n/2) \tag{3}$$
$$= \exp\{-m[(\lambda_2 - \lambda_0)\nu_0 + \phi(\lambda_0) - \phi(\lambda_2)]\}$$
$$\cdot \int_{(T<n/2)} g_{\nu_2, m-T}(m\nu_0 - S_T)\exp[-(\lambda_1 - \lambda_2)S_T]/g_{\nu_0,m}(m\nu_0)\, d\tilde{P}_{\nu_1},$$

where

$$\psi(\theta) = -\theta - \log(1-\theta), \quad \phi(\lambda) = \lambda - \log(1+\lambda),$$
$$\mu(\theta) = \psi'(\theta) = \theta/(1-\theta), \quad \nu(\lambda) = \psi'(\lambda) = \lambda/(1+\lambda),$$
$$f_{\mu,k}(x) = \frac{(1-\theta)^k}{(k-1)!}(x+k)^{k-1}\exp[-(x+k)(1-\theta)], \quad x \geq -k, \quad -\infty < \theta < 1,$$
$$g_{\nu,k}(y) = \frac{(1+\lambda)^k}{(k-1)!}(k-y)^{k-1}\exp[(1+\lambda)(y-k)], \quad y \leq k, \quad -1 < \lambda < \infty,$$

$\theta_2 < 0 < \theta_1 < 1$ satisfy $\psi(\theta_2) = \psi(\theta_1)$, and $-1 < \lambda_2 < 0 < \lambda_1$ satisfy $\phi(\lambda_2) = \phi(\lambda_1)$.

We work with (2) first. Under P_{μ_1} the increment of the random walk S_i has an exponential right tail. The following Lemma is a direct consequence. The proof is omitted.

LEMMA 1. *Under P_{μ_1}, $R_m = S_\tau - (n\zeta - 1)$ is independent of τ and has an exponential distribution with parameter $(1 - \theta_1)$.*

By Lemma 1

$$\int_{(\tau \leq n/2+1)} f_{\mu_2, m-\tau}(m\mu_0 - S_\tau)\exp[-(\theta_1 - \theta_2)S_\tau]\, dP_{\mu_1}/f_{\mu_0,m}(m\mu_0)$$
$$= \sum_{k=1}^{[n/2+1]} \int_{(\tau=k)} f_{\mu_2,m-k}(-n\zeta - R_m)\exp[-(\theta_1 - \theta_2)R_m]\, dP_{\mu_1}$$
$$\cdot \exp[-(\theta_1 - \theta_2)(n\zeta - 1)]/f_{\mu_0,m}(m\mu_0)$$
$$= (1 - \theta_1)\exp[-(\theta_1 - \theta_2)(n\zeta - 1)] \sum_{k=1}^{[n/2+1]} P_{\mu_1}(\tau = k)$$
$$\cdot \int_0^{m-k-n\zeta} f_{\mu_2,m-k}(-n\zeta - x)\exp[-x(1-\theta_2)]\, dx/f_{\mu_0,m}(m\mu_0)$$
$$= \exp[-(\theta_1 - \theta_2)n\zeta] \sum_{k=1}^{[n/2+1]} P_{\mu_1}(\tau = k) f_{\mu_2,m-k+1}(-n\zeta - 1)/f_{\mu_0,m}(\mu_0 m).$$

Observe that $f_{\mu_2,m-k+1}(x)$ is maximized at $x = ((m-k+1)\theta_2 - 1)/(1-\theta_2)$ and the maximized value is

$$\frac{(1-\theta_2)(m-k)^{m-k}e^{-(m-k)}}{(m-k)!},$$

and

$$f_{\mu_0,m}(m\mu_0) = \frac{m^m e^{-m}}{(m-1)[(m-1)!]}.$$

Substituting these results into the expression above, we have an upper bound of the form

$$(1-\theta_2)\exp[-(\theta_1-\theta_2)n\zeta] \sum_{k=1}^{[n/2+1]} P_{\mu_1}(\tau = k)\,\frac{(m-k)^{m-k}e^{-(m-k)}(m-1)[(m-1)!]}{(m-k)!\,m^m e^{-m}}.$$

Using Stirling's formula with upper and lower bound (see e.g. Feller, Vol. I, page 54), we find the expression above is bounded by

$$(1-\theta_2)\exp[-(\theta_1-\theta_2)n\zeta]e\left(\frac{m-1}{m}\right)^m \sum_{k=1}^{[n/2+1]} P_{\mu_1}(\tau = k)\left(\frac{m-1}{m-k}\right)^{1/2}$$

$$\le (1-\theta_2)\exp[-(\theta_1-\theta_2)n\zeta]e\left(\frac{m-1}{m}\right)^m \sqrt{2}.$$

So $P_{\mu_0}^{(m)}\{\tau \le n/2+1\} \le \sqrt{2}\exp\{-n[(\theta_1-\theta_2)\zeta - \psi(\theta_2)]\}$.

The process for bounding (3) is more or less the same, although we lose the independence of $S_T - n\zeta$ and T.

$$\int_{(T<n/2)} g_{\nu_2,m-T}(m\nu_0 - S_T)\exp[-(\lambda_1-\lambda_2)S_T]\,d\tilde{P}_{\nu_1}/g_{\nu_0,m}(\nu_0 m)$$

$$\le \sum_{k=1}^{[n/2]} \int_{n\zeta}^{n\zeta+1} g_{\nu_2,m-T}(1-y)\exp[-(\lambda_1-\lambda_2)y] \cdot \tilde{P}_{\nu_1}(T=k,\, S_T \in dy)/g_{\nu_0,m}(\nu_0 m)$$

$$\le \exp[-(\lambda_1-\lambda_2)n\zeta] \sum_{k=1}^{[n/2]} \int_{n\zeta}^{n\zeta+1} g_{\nu_2,m-T}(1-y) \cdot \tilde{P}_{\nu_1}(T=k,\, S_T \in dy)/g_{\nu_0,m}(\nu_0 m).$$

From this step on the argument is the same as above. Substituting in the maximal value of $g_{\nu,m-T}$ and using Stirling's formula carefully, we arrive at

$$\tilde{P}_{\nu_0}^{(m)}(T < n/2) \le \sqrt{2}\exp\{-n[(\lambda_1-\lambda_2)\zeta - \phi(\lambda_2)]\}.$$

To complete the proof it is sufficient to show

LEMMA 2.

$$\max_{\{(\theta_2,\theta_2):\psi(\theta_1)=\psi(\theta_2)\}}[(\theta_1-\theta_2)\zeta - \psi(\theta_2)] \ge 2\zeta^2$$

or equivalently

$$\max_{\{(\lambda_1,\lambda_2):\phi(\lambda_1)=\phi(\lambda_2)\}}[(\lambda_1-\lambda_2)\zeta - \phi(\lambda_2)] \ge 2\zeta^2.$$

PROOF OF LEMMA 2. Using the method of Lagrange's multiplier, it is easy to show that $(\theta_1 - \theta_2)\zeta - \psi(\theta_2)$ is maximized at θ_1 and θ_2 satisfying

$$\begin{cases} 1/\theta_1 + 1/|\theta_2| = 1/\zeta \\ \psi(\theta_1) = \psi(\theta_2). \end{cases} \tag{4}$$

Equation (4) involves a transcendental equation which is difficult to solve explicitly, but here is an easy way out. Dvoretzky, Kiefer, and Wolfowitz (1956) proved

$$P\{D_n^- > \sqrt{n}\zeta\} \le C_1 e^{-2n\zeta^2}.$$

Siegmund (1982) showed

$$P\{D_n^- > \sqrt{n}\zeta\} \sim C_2(\zeta)\exp(-n[(\theta_1 - \theta_2) - \psi(\theta_2)])$$

where θ_1 and θ_2 satisfy (4) and $C_2(\zeta)$ is a constant depending only on ζ. These two results imply

$$\lim_{n\to\infty}\sup C_1\exp(-2n\zeta^2)/C_2(\zeta)\exp(-n[(\theta_1 - \theta_2)\zeta - \psi(\theta_2)]) \ge 1.$$

Suppose $(\theta_2 - \theta_1)\zeta - \psi(\theta_2) < 2\zeta^2$ for some ζ, then

$$\lim_{n\to\infty} C_1\exp(-2n\zeta^2)/C_2(\zeta)\exp(-n[(\theta_1 - \theta_2)\zeta - \psi(\theta_2)] = 0.$$

This is a contradiction. Consequently Lemma 2 is true, and the proof of Theorem 1 is completed.

3. Concluding remarks.

(i) Birnbaum and Tingey (1951) gave the exact distribution of D_n^-, but their formula is inconvenient for numerical calculation.

(ii) At first sight, the conjecture mentioned in Section 1 seems unlikely to be true, when compared with the asymptotic result $\lim_{n\to\infty}P(D_n^- > \zeta) = e^{-2\zeta^2}$, but Smirnov's (1944) result $P\{D_n^- > \zeta\} = \exp[-2\zeta(\zeta + (3n^{1/2})^{-1})] + o(n^{-1/2})$, which suggests that D_n^- approaches the asymptotic distribution from below, served as analytical support of the conjecture.

(iii) The usefulness of a bound of the form $p\{D_n^- > \sqrt{n}\zeta\} \le ce^{-2n\zeta^2}(*)$ can be argued as follows: On the one hand we have an asymptotic result but without accurate estimation of the error term; on the other hand we have an exact formula but even for a moderately large sample size it is not easy to do the numerical computation. A bound of the form (*) with a reasonable constant c can serve as an easily calculated and conservative confidence bound. The result of Theorem 1 represents substantial progress in this direction. Also, in some cases the constant c appears as a component of a more complicated procedure for determining confidence bounds. (See e.g. Burke et al., 1981; and Csörgő, Horváth, 1981, Lemma 2.1), so it is helpful to know the value of c even approximately.

(iv) Carefully examining the proof of Theorem 1, it is clear that actually we have proved a better result, i.e.

$$P\{D_n^- > \sqrt{n}\zeta\}$$
$$\le \sqrt{2}(P_{\mu_1}(\tau \le n/2) + \tilde{P}_{\nu_1}(T < n/2))\exp(-n[(\theta_1 - \theta_2)\zeta - \psi(\theta_1)]).$$

It is easy to see that $\nu_1 = |\mu_2|$ and from (4) we know $|\mu_2|^{-1} + \mu_1^{-1} = \zeta^{-1}$. This implies either $|\mu_2| < 2\zeta < \mu_1$ or $\mu_1 < 2\zeta < |\mu_2|$. For fixed ζ as $n \to \infty$ by strong law of large number we have

$$\lim_{n\to\infty}(P_{\mu_1}(\tau \le n/2) + \tilde{P}_{\nu_1}(T < n/2)) = 1.$$

Since the author is unaware of any nontrivial uniform upper bound on $P_{\mu_1}(\tau \le n/2) + \tilde{P}_{\nu_1}(T < n/2)$ the trivial upper bound $P_{\mu_1}(\tau \le n/2) + \tilde{P}_{\nu_1}(T < n/2) \le 2$ is used in the proof. This is the place where we might lose a factor of ½. The other place we might lose precision is where we replace $f_{\mu_2,m-k+1}(-n\zeta - 1)$ by the maximum of $f_{\mu_2,m-k+1}(\cdot)$, and replace $(1 - \tau/m)^{-1/2}1_{(\tau\le n/2)}$ by its maximum $\sqrt{2}$. The former introduces serious inaccuracy when $-n\zeta - 1$ is away from the peak of $f_{\mu_2,m-k+1}(\cdot)$, although a simple calculation shows that when ζ is fixed and $n \to \infty$, by law of large numbers $\tau \approx n\zeta/\mu_1$ under P_{μ_1}, so $-n\zeta - 1$ is approximately at the peak. The use of the rather crude upper bound above reflects the author's unawareness of a sharper inequality.

(v) It is also possible to derive a bound of the form $P(D_n^- > \sqrt{n}\zeta) \le \zeta^{-1/2}e^{-2n\zeta^2}$ by working on (2) only. This bound is strictly better than Theorem 1 when $\zeta > 1/8$, but the result is poor when ζ is small. This is the reason why we split the set $\{\tau < m\}$ into two parts and use a different argument on each part.

(vi) One might want to use $p\{D_n^- > \sqrt{n}\zeta\} \le 2\sqrt{2}\exp(-n[(\theta_1 - \theta_2)\zeta - \psi(\theta_1)])$. This bound is slightly better than Theorem 1, but more numerical computation is required.

Acknowledgement. The author is greatly indebted to Professor David Siegmund for suggesting this problem and providing many helpful comments.

REFERENCES

BIRNBAUM, Z. W. and MCCARTY, R. C. (1958). A distribution-free upper confidence bound for $\Pr\{Y < X\}$, based on independent samples of X and Y. *Ann. Math. Statist.* **29** 558–562.

BIRNBAUM, Z. W. and TINGEY, F. H. (1951). One-sided confidence contours for probability distribution function. *Ann. Math. Statist.* **22** 592–596.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1981). A strong approximation of some biometric estimates under random censorship. *Z. Wahrsch. verw. Gebiete* **56** 87–112.

CSÖRGŐ, S. and HORVÁTH, L. (1981). On the Koziol-Green model for random censorship. *Biometrika* **68** 391–401.

DEVROYE, L. P. and WISE, G. L. (1979). On the recovery of discrete probability densities from imperfect measurements. *J. Franklin Inst.* **307** 1–20.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications I.* Wiley, New York.

SIEGMUND, D. (1982). Large deviations for boundary crossing probabilities. *Ann. Probab.* **10** 581–588.

SIEGMUND, D. (1983). Corrected diffusion approximations and their applications. Stanford University Technical Report.

SMIRNOV, N. V. (1944). An approximation to the distribution laws of random quantities determined by empirical data. *Uspehi Mat. Nauk* **10** 179–206.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305