The Annals of Probability
1990, Vol. 18, No. 3, 1269–1283

THE TIGHT CONSTANT IN THE DVORETZKY–KIEFER–WOLFOWITZ INEQUALITY

BY P. MASSART

Université de Paris-Sud

Let $\hat{F}_n$ denote the empirical distribution function for a sample of n i.i.d. random variables with distribution function F. In 1956 Dvoretzky, Kiefer and Wolfowitz proved that

$$P\Big(\sqrt{n}\sup_x\big(\hat{F}_n(x) - F(x)\big) > \lambda\Big) \le C\exp(-2\lambda^2),$$

where C is some unspecified constant. We show that C can be taken as 1 (as conjectured by Birnbaum and McCarty in 1958), provided that $\exp(-2\lambda^2) \le \frac{1}{2}$. In particular, the two-sided inequality

$$P\Big(\sqrt{n}\sup_x|\hat{F}_n(x) - F(x)| > \lambda\Big) \le 2\exp(-2\lambda^2)$$

holds without any restriction on λ. In the one-sided as well as in the two-sided case, the constants cannot be further improved.

1. Introduction. Let $x_1, \dots, x_n$ be independent, identically distributed real valued random variables with *continuous* distribution function F. Let $\hat{F}_n$ be the empirical distribution function which is defined by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(x_i \le x)}.$$

Next, we denote by Z_n the centered and normalized empirical process $\sqrt{n}(\hat{F}_n - F)$. To test goodness of fit, the following statistics are commonly used:

$$D_n^+ = \sup_{x\in\mathbb{R}} Z_n(x), \qquad D_n^- = \sup_{x\in\mathbb{R}} -Z_n(x) \quad \text{and} \quad D_n = \sup_{x\in\mathbb{R}}|Z_n(x)|.$$

Smirnov (1944) and Kolmogorov (1933) introduced and studied these statistics for the first time and showed that their distributions do not depend on F. Moreover the one-sided statistics D_n^+ and D_n^- have the same law and the following asymptotic results hold:

$$\lim_{n\to\infty} P(D_n^- > \lambda) = \exp(-2\lambda^2), \tag{1.1}$$

$$\lim_{n\to\infty} P(D_n > \lambda) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2\lambda^2). \tag{1.2}$$

These results were elucidated later on by Donsker's functional central limit

Received July 1988; revised July 1989.
AMS 1980 *subject classifications*. Primary 62E15; secondary 62G15.
Key words and phrases. Brownian bridge, empirical process, Kolmogorov–Smirnov statistics.

theorem, which implies the weak convergence of D_n^- and D_n to $\sup_{x \in [0,1]} Z(x)$ and $\sup_{x \in [0,1]} |Z(x)|$, respectively, where Z is a Brownian bridge.

In connection with (1.1), Dvoretzky, Kiefer and Wolfowitz (1956) proved a bound of the form

$$P(D_n^- > \lambda) \le C \exp(-2\lambda^2),$$

where C is some unspecified constant.

Birnbaum and McCarty (1958) conjectured that C can be taken as 1. The conjecture is substantiated on the one hand by the asymptotic expansion [which is due to Smirnov (1944)],

$$(1.3) \quad P(D_n^- > \lambda) = \exp(-2\lambda^2)\left(1 - \frac{2\lambda}{3\sqrt{n}} + \mathbb{O}(n^{-1})\right), \quad \text{with } \lambda = \mathbb{O}(n^{1/6}),$$

and by numerical computations on the other hand [see Birnbaum and McCarty (1958)]. Several attempts were then made in order to calculate the best constant C. Devroye and Wise (1979) showed that $C \le 306$. Shorack and Wellner (1986) gave $C \le 29$. Finally, the best result known to the author is that of Hu (1985), who proved that $C \le 2\sqrt{2}$.

What we show below is that the conjecture of Birnbaum and McCarty is true when subject to a mild constraint on λ. (This condition is needed to make our proof work; however, it does not seem to be too restrictive for statistical applications.)

THEOREM 1. *For any integer n and any* λ *not less than* $\sqrt{[\log(2)]/2} \wedge \gamma n^{-1/6}$, *where* $\gamma = 1.0841$, *we have*

$$(1.4) \quad P(D_n^- > \lambda) \le \exp(-2\lambda^2).$$

COMMENT 1. In particular, theorem 1 implies that inequality (1.4) holds whenever $\exp(-2\lambda^2) \le \frac{1}{2}$.

Since

$$P(D_n > \lambda) \le 2P(D_n^- > \lambda),$$

a bound for the two-sided Kolmogorov–Smirnov statistic follows easily from Theorem 1 and Comment 1.

COROLLARY 1. *For all integer n and any positive* λ, *we have*

$$(1.5) \quad P(D_n > \lambda) \le 2\exp(-2\lambda^2).$$

COMMENT 2. (i) It follows from Comment 1 and from (1.1) that the true level of significance of the one-sided Kolmogorov–Smirnov test for goodness of fit is not greater than the level of the asymptotic test (at least if this level does not exceed 50%, which is of course the case in practice!), whatever the sample size can be.

(ii) The numerical constants 1 and 2 appearing, respectively, in bounds (1.4) and (1.5) cannot be further improved because of the asymptotic formulae (1.1) and (1.2).

(iii) It is well known that the Kolmogorov–Smirnov statistics are stochastically smaller for all laws having atoms [this is an easy consequence of Shorack and Wellner (1986), Theorem 3, page 5, for instance]. So inequalities (1.4) and (1.5) remain valid when F is not continuous.

As a by-product of the proof of Theorem 1, we get an interesting exponential bound for binomial tails which does not seem to be known, implying the classical inequalities of Hoeffding and Bernstein [these inequalities are recorded in Shorack and Wellner (1986), page 440].

THEOREM 2. *Let S be a random variable with binomial distribution $\mathscr{B}(n, p)$. Setting $q = 1 - p$, the following inequality holds for any positive ε, with $\varepsilon \le q$:*

$$P(S - np > n\varepsilon) \le \exp\left(-\frac{n\varepsilon^2}{2(p + \varepsilon/3)(q - \varepsilon/3)}\right).$$

2. Proof of the main result. Since the distribution of D_n is the same for all continuous distributions, we take F to be the uniform distribution on $[0,1]$. We also assume that $\lambda < \sqrt{n}$ [otherwise, $P(D_n^- > \lambda) = 0$ and (1.4) is immediate].

To give the idea of the proof, we need first to introduce some notation. Let Z be a Brownian bridge; define τ_n (resp., τ) to be the first time that $-Z_n$ (resp., Z) crosses the level λ.

Setting $\varepsilon = \lambda/\sqrt{n}$ and $f_\lambda(s) = dP(\tau \le s)/ds$, we shall show that the following (rather mysterious) local comparison between the laws of τ_n and τ is *nearly* true for all t of the form $\varepsilon + j/n$, where j is an integer satisfying $0 \le j < n - \lambda\sqrt{n}$:

$$P(\tau_n = t) \le 1/n f_\lambda(t - \varepsilon/3).$$

If this inequality were really available, (1.3) would easily follow from a summation extended to all j with $0 \le j < n - \lambda\sqrt{n}$. The difficulty arises from the fact that a corrective factor is needed to make such an inequality valid. This factor will cause us some trouble, especially "on the tail," that is, when t is close to ε.

Now, let us give the exact distributions of τ_n and τ. The result at (2.1) is due to Smirnov (1944), and at (2.2) to Csáki (1974) [see also Bretagnolle and Massart (1989) for a short and direct proof of the second result]:

$$\begin{aligned} P(\tau_n = \varepsilon + j/n) &= p_{\lambda,n}(j) \\ &= \lambda\sqrt{n}\,(j + \lambda\sqrt{n})^{j-1}(n - j - \lambda\sqrt{n})^{n-j} n^{-n}\binom{n}{j} \end{aligned} \tag{2.1}$$

for all integers j with $0 \leq j < \lambda\sqrt{n}$,

$$(2.2)\quad f_\lambda(s) = \frac{\lambda}{\sqrt{2\pi}} s^{-3/2}(1-s)^{-1/2}\exp\left(-\frac{\lambda^2}{2s(1-s)}\right), \qquad s \in]0,1[.$$

Of course, summing these formulae, we also have

$$(2.3)\quad P(D_n^- > \lambda) = \sum_{0 \leq j < n - \lambda\sqrt{n}} p_{\lambda,n}(j),$$

$$(2.4)\quad \exp(-2\lambda^2) = \int_0^1 f_\lambda(s)\, ds.$$

The key result is the following comparison between $p_{\lambda,n}$ and f_λ.

PROPOSITION 1. *Let j be an integer with $0 \leq j < n - \lambda\sqrt{n}$. Let $s = 2\varepsilon/3 + j/n$ and $s' = 1 - s$. If $n\varepsilon \geq 2$, then*

$$(2.5)\quad \begin{aligned} p_{\lambda,n}(j) \leq{}& \frac{1}{n}\left(1 - \frac{\varepsilon}{3s'} + \frac{\varepsilon^2}{6s'^2}\right) \\ &\times \exp\left(\frac{0.4}{ns} - \frac{\varepsilon^2}{24n}(v_n(s) + v_n(s'))\right) f_\lambda(s), \end{aligned}$$

where $v_n(s) = (s(s^2 - 1/(4n^2)))^{-1}$.

To prove Proposition 1, we need several technical lemmas. The first one is interesting in itself.

LEMMA 1 (A lower bound for the Cramér transform of the Bernoulli law). *Let $0 < \varepsilon \leq q = 1 - p < 1$. Consider the functions*

$$h(p,\varepsilon) = (p+\varepsilon)\log\left(\frac{p+\varepsilon}{p}\right) + (q-\varepsilon)\log\left(\frac{q-\varepsilon}{q}\right),$$

$$\varphi(t) = t - \frac{t^2}{2(1+2t/3)} - \log(1+t), \qquad t \geq 0.$$

Then

(i) *φ is a positive increasing convex function with $\varphi(t)/t \to \frac{1}{4}$ as t goes to infinity,*

(ii) *$h(p,\varepsilon) \geq \varepsilon^2/[2(p+\varepsilon/3)(q-\varepsilon/3)] + \varepsilon\varphi(t)/t$, with $t = \varepsilon/(q-\varepsilon)$.*

PROOF. To prove (i), just note that $\varphi(0) = 0$ and that

$$\varphi'(t) = (t^3/9)(1+2t/3)^{-2}(1+t)^{-1} > 0,$$

for all positive t.

To prove (ii), we set $x = p + \varepsilon$, then $0 < \varepsilon \leq x$ and

$$h(p,\varepsilon) - \frac{\varepsilon^2}{2(p+\varepsilon/3)(q-\varepsilon/3)} - \frac{\varepsilon\varphi(t)}{t}$$
$$= x\log(x) - x\log(x-\varepsilon) - \frac{\varepsilon^2}{2(x-2\varepsilon/3)} - \varepsilon,$$

the derivative of the right-hand side of this identity w.r.t. ε is equal to

$$(\varepsilon^3/9)(x-\varepsilon)^{-1}(x-2\varepsilon/3)^{-2} > 0.$$

Thus, being increasing w.r.t. ε and equal to zero when $\varepsilon = 0$, the right-hand side of the above identity is a nonnegative function of ε for any fixed x. Hence (ii) holds. □

COMMENT 3. Using the usual Cramér–Chernoff computation, if S has the binomial distribution $\mathscr{B}(n,p)$, we have [see, for instance, Shorack and Wellner (1986), page 440]

$$P(S - np > n\varepsilon) \leq \exp(-nh(p,\varepsilon)),$$

where $h(\cdot,\cdot)$ is defined as in Lemma 1. So Lemma 1 implies immediately Theorem 2.

PROOF OF PROPOSITION 1 (where $j \geq 1$). Using Stirling's formula with upper and lower bounds [as given in Feller (1968), page 54], we get

$$\binom{n}{j} \leq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{j(n-j)}}\, n^n j^{-j}(n-j)^{-(n-j)}C_j,$$

where $C_j = \exp(-1/(12j+1))$; then

$$p_{\lambda,n}(j) \leq \frac{\lambda}{\sqrt{2\pi}}\frac{n}{\sqrt{j(n-j)}\,(j+\lambda\sqrt{n})}\left(\frac{j+\lambda\sqrt{n}}{j}\right)^j\left(\frac{n-j-\lambda\sqrt{n}}{n-j}\right)^{n-j}C_j.$$

Let $h(\cdot,\cdot)$ be defined as in Lemma 1, recalling that $s = 2\varepsilon/3 + j/n$ and $s' = 1 - s$, the above inequality becomes

$$p_{\lambda,n}(j) \leq \frac{\lambda}{n\sqrt{2\pi}}\left(s - \frac{2\varepsilon}{3}\right)^{-1/2}\left(s + \frac{\varepsilon}{3}\right)^{-1}\left(s' + \frac{2\varepsilon}{3}\right)^{-1/2}$$
$$\times\exp\left(-nh\left(s' - \frac{\varepsilon}{3},\varepsilon\right)\right)C_j.$$

Let ψ be the function defined on $\mathbb{R}^+$ by

$$\psi(t) = -\log(1+t) + \frac{3}{2}\log\left(1 + \frac{2t}{3}\right). \tag{2.6}$$

Setting $t = \varepsilon/(s - 2\varepsilon/3)$, Lemma 1 implies

$$p_{\lambda,n}(j) \le \frac{1}{n}\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} \frac{s^{3/2}}{\sqrt{s - 2\varepsilon/3}\,(s + \varepsilon/3)} C_j \exp\left(-\frac{n\varepsilon\varphi(t)}{t}\right) f_\lambda(s),$$

which can also be written

$$(2.7)\qquad p_{\lambda,n}(j) \le \frac{1}{n}\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} C_j \exp\left(-\frac{n\varepsilon\varphi(t)}{t} + \psi(t)\right) f_\lambda(s).$$

To control the error term $-n\varepsilon\varphi(t)/t + \psi(t)$, we shall use the following lemma.

LEMMA 2. *Let $\theta = 0.4833$. Let φ and ψ be the functions defined, respectively, in Lemma* 1 *and by* (2.6). *For positive t and ν, let*

$$T(\nu, t) = \nu^2\varphi(t) - \nu t\psi(t) + \frac{\theta t^2}{1 + 2t/3}.$$

Then the function T is positive on the set $\{(\nu, t)/0 < t \le \nu\}$.

PROOF. (a) *In the boundary case where $\nu = t$:* An elementary computation gives

$$\frac{d}{dt}\left(\frac{T(t,t)}{t^2}\right) = \frac{-2\theta/3 - t/3 + t^2/9}{(1 + 2t/3)^2}$$

so $T(t,t)/t^2$ is minimum at the point $t_0 = (3 + \sqrt{9 + 24\theta})/2$. Recalling that $\theta = 0.4833$, one may easily verify that $T(t_0, t_0)/t_0^2 \ge 5 \times 10^{-6}$; thus, T is positive on the diagonal.

(b) *In the general case:* From (a), we know that $T(t,t) > 0$. But, as a function of ν, T is polynomial with degree 2. Hence, $T(\nu, t)$ is positive for all $\nu \ge t$ if and only if one of the two following conditions is satisfied:

$$\Delta(t) = (1 + 2t/3)\psi^2(t) - 4\theta\varphi(t) < 0;$$

$$2\varphi(t) - \psi(t) > 0.$$

So, to get Lemma 2 it is enough to show that

(i) $\Delta(t) < 0$ for all $0 < t \le 3.37$,
(ii) $2\varphi(t) - \psi(t) > 0$ for all $t \ge 3.37$.

PROOF OF (i). Consider the functions defined on $\mathbb{R}^+$ by

$$R_0(t) = \frac{\psi^2(t)}{\varphi(t)},$$

$$R_1(t) = 6\varphi(t) - t^2\left(1 + \frac{2t}{3}\right)^{-1}\psi(t)$$

and

$$R_2(t) = -6\psi(t) + \frac{t^2}{1+t}.$$

Note that the following relations between these functions are available:

$$\frac{d}{dt}R_0(t) = \frac{t}{9}\left(1 + \frac{2t}{3}\right)^{-1}(1+t)^{-1}R_1(t),$$

$$\frac{d}{dt}R_1(t) = \frac{t}{3}\left(1 + \frac{2t}{3}\right)^{-2}R_2(t)$$

and

$$\frac{d}{dt}R_2(t) = \frac{t^2}{3}(1+2t)(1+t)^{-2}\left(1 + \frac{2t}{3}\right)^{-1}.$$

Since $R_0(0) = R_1(0) = R_2(0) = 0$, it comes from the above relations that R_2 is a positive and increasing function; thus, R_1 has the same property, which in turn implies that the same is true for R_0. Hence, setting $R(t) = (1 + 2t/3)R_0(t)$, R is a fortiori increasing. So, for any $t \in [0, 3.37]$, we have

$$R(t) \le R(3.37) \le 0.4723 < 0.4833;$$

thus (i) is satisfied. □

PROOF OF (ii). Note that

$$2\varphi'(t) - \psi'(t) = \frac{t}{9}(2t^2 - 2t - 3)(1+t)^{-1}\left(1 + \frac{2t}{3}\right)^{-2};$$

hence, $2\varphi - \psi$ is an increasing function on the interval $[(1+\sqrt{7})/2, +\infty[$. Then, since $3.37 > (1+\sqrt{7})/2$, for any $t \ge 3.37$, we have

$$2\varphi(t) - \psi(t) \ge 2\varphi(3.37) - \psi(3.37) \ge 7 \times 10^{-4} > 0,$$

completing the proof of (ii) and, therefore, that of Lemma 2. □

In order to deduce Proposition 1 from inequality (2.7), we need some more technical results.

CLAIM 1. Let $\beta = 0.826$, then

$$(1 + 2x)^{-1/2} \le \left(1 - x + \frac{3x^2}{2}\right)\exp(-\beta x^3), \quad \text{for any } x \in [0,1].$$

PROOF OF CLAIM 1. Let $\alpha = 135/32$. Note that $\rho: y \to 1 + \alpha y - \exp(2\beta y)$ is a concave function with $\rho(0) = 0$ and $\rho(1) \ge 1.3 \times 10^{-3} > 0$. So ρ is nonnegative on the interval [0, 1]. Hence,

$$\exp(2\beta x^3) \le 1 + \alpha x^3 \le 1 + \alpha x^3 + \frac{x^3(5 - 12x)^2}{32} = \left(1 - x + \frac{3x^2}{2}\right)^2 (1 + 2x).$$

Elevating both sides of this inequality to the power $\frac{1}{2}$, we easily get Claim 1. □

CLAIM 2. Let v_n be defined as in the statement of Proposition 1. Then, for any positive ε and s' such that $n\varepsilon \ge 2$ and $ns' \ge 1$, we have

$$\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} \le \left(1 - \frac{\varepsilon^2}{3s'} + \frac{\varepsilon}{6s'^2}\right)\exp\left(-\frac{\varepsilon^2 v_n(s')}{24n}\right). \tag{2.8}$$

PROOF OF CLAIM 2. Applying Claim 1 with $x = \varepsilon/(3s')$, (2.8) reduces to

$$\frac{\beta\varepsilon^3}{27s'^3} \ge \frac{\varepsilon^2 v_n(s')}{24n},$$

which is equivalent to

$$\frac{8\beta}{9} \ge \left(n\varepsilon\left(1 - (2ns')^{-2}\right)\right)^{-1}.$$

Now, since $n\varepsilon \ge 2$ and $ns' \ge 1$, we have

$$\left(n\varepsilon\left(1 - (2ns')^{-2}\right)\right)^{-1} \le \frac{2}{3} < \frac{8\beta}{9};$$

hence, Claim 2 is verified. □

CLAIM 3. Let v_n be defined as in the statement of Proposition 1. Then, for any positive ε and s with $n^{-1} \le \varepsilon \le 3s/2$, we have

$$\left(1 + 12n\left(s - \frac{2\varepsilon}{3}\right)\right)^{-1} \ge (12ns)^{-1} + \frac{\varepsilon^2 v_n(s)}{24n}. \tag{2.9}$$

PROOF OF CLAIM 3. Since

$$\left(1 + 12n\left(s - \frac{2\varepsilon}{3}\right)\right)^{-1} - (12ns)^{-1}$$
$$= (8n\varepsilon - 1)(12ns)^{-1}\left(1 + 12n\left(s - \frac{2\varepsilon}{3}\right)\right)^{-1},$$

(2.9) reduces to

$$\frac{8n\varepsilon - 1}{1 + 12n(s - 2\varepsilon/3)} \ge \frac{\varepsilon^2}{2}\Big/\left(s^2 - \frac{1}{4n^2}\right),$$

which is equivalent to

$$2(8n\varepsilon - 1)s^2 - 12n\varepsilon^2 s + (8n\varepsilon - 1)\left(\varepsilon^2 - \frac{1}{2n^2}\right) \geq 0,$$

as the left-hand side of this inequality is increasing w.r.t. s, it is enough to consider the case where $s = 2\varepsilon/3$. So, we have to verify that $8n\varepsilon - 1 \geq (\frac{8}{9} - (2n\varepsilon)^{-2})^{-1}$.

Since the left-hand side (resp. right-hand side) of the latter inequality is increasing (resp. decreasing) w.r.t. $n\varepsilon$, we may only check the case where $n\varepsilon = 1$. The verification is easy. Thus (2.9) holds. □

We may now finish the proof of Proposition 1 in the case where $j \geq 1$. Starting from (2.7) and noting that $t = n\varepsilon/j \leq n\varepsilon$, we may apply Lemma 2 with $\nu = n\varepsilon$ to get

$$p_{\lambda,n}(j) \leq \frac{1}{n}\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} C_j \exp\left(\frac{\theta}{ns}\right) f_\lambda(s).$$

The end of the proof is then straightforward, using Claims 2 and 3 to bound, respectively, $(1 + 2\varepsilon/(3s'))^{-1/2}$ and C_j, and noting finally that $\theta - \frac{1}{12} < 0.4$. □

Proof of Proposition 1 (where $j = 0$). Here $s = 2\varepsilon/3$, and Lemma 1 gives

$$\begin{aligned} p_{\lambda,n}(0) &= (1-\varepsilon)^n \leq \exp\left(-\frac{n\varepsilon}{4} - \frac{\lambda^2}{2ss'}\right) \\ &\leq \frac{\lambda}{n\sqrt{2\pi}} s^{-3/2} \exp\left(\frac{0.4}{n\varepsilon}\right) \exp\left(-\frac{\lambda^2}{2ss'}\right) \exp(-H(n\varepsilon)), \end{aligned}$$

where the function H is defined, for any positive ν, by

$$H(\nu) = \frac{3\log(3/2)}{2} - \frac{\log(2\pi)}{2} + \frac{\nu}{4} + \frac{0.4}{\nu} - \frac{\log(\nu)}{2},$$

it is elementary to see that H is minimum at the point $\nu_0 = 1 + \sqrt{2.6}$, with $H(\nu_0) \geq 1.5 \times 10^{-2} > 0$. Therefore, H is a positive function and so

$$\begin{aligned} p_{\lambda,n}(0) &\leq \frac{\lambda}{n\sqrt{2\pi}} s^{-3/2} \exp\left(\frac{0.4}{n\varepsilon}\right) \exp\left(-\frac{\lambda^2}{2ss'}\right) \\ &\leq \frac{1}{n}\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} \exp\left(\frac{0.4}{n\varepsilon}\right) f_\lambda(s). \end{aligned}$$

Now, since $n\varepsilon \geq 2$, we have

$$\frac{\varepsilon^2 v_n(s)}{24n} \leq (16n\varepsilon(4/9 - 1/16))^{-1} \leq 9/(55n\varepsilon)$$

hence

$$\frac{0.4}{n\varepsilon} + \frac{\varepsilon^2 v_n(s)}{24n} \le \frac{0.4 + 9/55}{n\varepsilon} \le \frac{0.4}{ns};$$

thus,

$$p_{\lambda,n}(0) \le \frac{1}{n}\left(1 + \frac{2\varepsilon}{s'}\right)^{-1/2} \exp\left(\frac{0.4}{ns} - \frac{\varepsilon^2 v_n(s)}{24n}\right) f_\lambda(s).$$

So, using Claim 2, we easily get (2.5) and the proof of Proposition 1 is complete. □

Let us bring our attention to the proof of Theorem 1. To derive Theorem 1 from Proposition 1, we first have to bound a Riemann sum by the corresponding integral with an explicit corrective factor. To do this we shall interpolate the mean value of a certain function on a small interval by its value at the midpoint. The following lemma will be useful.

LEMMA 3. *Let $0 < \delta \le s \le 1 - \delta$ and $s' = 1 - s$. If g is a positive function, defined on the interval $[s - \delta, s + \delta]$, such that $\log(g)$ is convex, then, for any positive λ, a lower bound for*

$$\frac{1}{2\delta}\int_{s-\delta}^{s+\delta} g(u)\exp\left(-\frac{\lambda^2}{2u(1-u)}\right) du$$

is given by

$$g(s)\exp\left(-\frac{\lambda^2}{2ss'}\right)\exp\left(-\frac{\lambda^2\delta^2}{6}\left(\left(s(s^2-\delta^2)\right)^{-1} + \left(s'(s'^2-\delta^2)\right)^{-1}\right)\right).$$

PROOF. Applying Jensen's inequality twice, we get

$$\begin{aligned}
&\frac{1}{2\delta}\int_{s-\delta}^{s+\delta} g(u)\exp\left(-\frac{\lambda^2}{u(1-u)}\right) du \\
&\quad\ge \exp\left(\frac{1}{2\delta}\int_{s-\delta}^{s+\delta}\left(\log(g(u)) - \frac{\lambda^2}{2u(1-u)}\right) du\right) \\
&\quad\ge \exp\left(\log(g(s)) - \frac{1}{2\delta}\int_{s-\delta}^{s+\delta}\frac{\lambda^2}{2}\left(u^{-1} + (1-u)^{-1}\right) du\right).
\end{aligned}$$

Now the function $u \to 1/u$ has a positive fourth derivative. Simpson's interpolation method at the second order (parabolic interpolation) then gives

$$\frac{1}{2\delta}\int_{s-\delta}^{s+\delta}\frac{1}{u}\, du \le \frac{1}{6}\left(\frac{1}{s+\delta} + \frac{1}{s-\delta} + \frac{4}{s}\right) = \frac{1}{s} + \frac{\delta^2}{3s(s^2-\delta^2)}.$$

Substituting s' for s gives a bound for the mean value of $1/(1-u)$, finishing the proof of Lemma 3. □

To integrate the right-hand side of inequality (2.7), the following exact formulae will be of some help.

LEMMA 4. *For any nonnegative a, b and positive λ, let*

$$I_{a,b}(\lambda) = \frac{\lambda \exp(2\lambda^2)}{\sqrt{2\pi}} \int_0^1 u^{-1/2-a}(1-u)^{-1/2-b} \exp\left(-\frac{\lambda^2}{2u(1-u)}\right) du.$$

Then, the following relations hold:

(i) $I_{1,1}(\lambda)/2 = I_{1,0}(\lambda) = 1$;
(ii) $I_{2,2}(\lambda)/2 = I_{2,1}(\lambda) = 4 + \lambda^{-2}$;
(iii) $I_{2,0}(\lambda) = 2 + \lambda^{-2}$.

PROOF. Clearly $I_{a,b} = I_{b,a}$. Then note that, for any $u \in [0,1]$,

$$u^{-1/2-a}(1-u)^{-1/2-b} - (1-u)^{-1/2-b}u^{-1/2-a+1}$$
$$= u^{-1/2-a}(1-u)^{-1/2-b+1},$$

which implies immediately

$$I_{a,b} = I_{a-1,b} + I_{a,b-1}. \tag{2.10}$$

From (2.10) (with $a = b = 1$) and (2.4), we deduce (i). Next, deriving (i) w.r.t. λ gives (ii). We finally apply (2.10) with $a = 1$ and $b = 2$ to get (iii) from (i) and (ii). □

We are now in position to prove Theorem 1.

PROOF OF THEOREM 1 (where $n \geq 39$ and $\lambda \leq \sqrt{n}/2$). Since $n \geq 39$, the condition on λ which is given in the statement of Theorem 1 reduces to $\lambda \geq \gamma n^{-1/6}$. In particular, we also have that $n\varepsilon = \lambda\sqrt{n} \geq 3.6764$. Hence, taking into account the fact that the function $\mathscr{y}(x) = (\exp(x) - 1)/x$ is increasing for positive x and recalling that $s \geq 2\varepsilon/3$, we have

$$\exp\left(\frac{0.4}{ns}\right) \leq 1 + \mathscr{y}\left(\frac{0.6}{3.6764}\right)\frac{0.4}{ns}.$$

So, setting $\mu = 0.4345$,

$$\exp\left(\frac{0.4}{ns}\right) \leq 1 + \frac{\mu}{ns}.$$

This, combined with Proposition 1, provides the following upper bound for $p_{\lambda,n}(j)$:

$$\frac{1}{n}\left(1-\frac{\varepsilon}{3s'}+\frac{\varepsilon^2}{6s'^2}\right)\left(1+\frac{\mu}{ns}\right)\exp\left(-\frac{\varepsilon^2(v_n(s)+v_n(s'))}{24n}\right)f_\lambda(s).$$

Next, to apply Lemma 3, just note that $z\colon x \to \log(6x^2-2x+1)-5/2\log(x)$ is convex [in fact, $2z''(x)x^2(6x^2-2x+1)^2 = 36(x^2-1)^2+5(2x-1)^2+40x^2>0$]. So we may use Lemma 3 with $\delta = 1/(2n)$ and

$$g(u)=\frac{\lambda}{\sqrt{2\pi}}\left(1+\frac{\mu}{nu}\right)u^{-3/2}(1-u)^{-1/2}\left(1-\frac{\varepsilon}{3(1-u)}+\frac{\varepsilon^2}{6(1-u)^2}\right).$$

It is easy to verify that $\log(g)$ is convex when noticing that $\log(g(u)) = \log(\lambda/(6\sqrt{2\pi\varepsilon})) + \log(1+\mu/(nu)) - 3/2\log(u) + z((1-u)/\varepsilon))$, getting, for any integer j with $0 \le j < n - \lambda\sqrt{n}$,

$$p_{\lambda,n}(j) \le \int_{s-1/(2n)}^{s+1/(2n)}\left(1-\frac{\varepsilon}{3(1-u)}+\frac{\varepsilon^2}{6(1-u)^2}\right)\left(1+\frac{\mu}{nu}\right)f_\lambda(u)\,du.$$

Summing this inequality, we obtain, in the notation of Lemma 4,

$$\begin{aligned}\exp(2\lambda^2)P(D_n^- > \lambda) \le I_{1,0}(\lambda) &- \frac{\varepsilon}{3}I_{1,1}(\lambda)+\frac{\varepsilon^2}{6}I_{2,1}(\lambda)+\frac{\mu}{n}I_{2,0}(\lambda)\\ &-\frac{2\mu}{3n}I_{2,1}(\lambda)+\frac{\varepsilon^2\mu}{6n}I_{2,2}(\lambda).\end{aligned}$$

Applying Lemma 4, this becomes, via some easy calculations,

$$\begin{aligned}&\frac{3\sqrt{n}}{2\lambda}(\exp(2\lambda^2)P(D_n^- > \lambda)-1)\\ &\quad\le -1+\left(\lambda+\frac{1}{4\lambda}+\frac{3\mu}{\lambda}+\frac{3\mu}{2\lambda^3}\right)n^{-1/2}\\ &\qquad-\frac{\mu}{2}\left(4+\frac{1}{\lambda^2}\right)n^{-1}+\frac{\mu}{2}\left(4\lambda+\frac{1}{\lambda}\right)n^{-3/2}.\end{aligned}\tag{2.11}$$

As a matter of fact, inequality (2.11) is quite satisfactory since it is a *nonasymptotic* version of Smirnov's asymptotic expansion (1.3).

Let us denote by $\eta_n(\lambda)$ the right-hand side of inequality (2.11). Then η_n is convex w.r.t. λ. Thus, the negativity of η_n on the interval $[\gamma n^{-1/6}, \sqrt{n}/2]$ reduces to the negativity of $a_n = \eta_n(\gamma n^{-1/6})$ and $b_n = \eta_n(\sqrt{n}/2)$.

Now, it is easy to see that a_n and b_n are both decreasing. So we have just to verify negativity for $n = 39$, that is,

$$a_{39} \le -6\times 10^{-3} < 0,$$
$$b_{39} \le -0.4 < 0.$$

Hence, Theorem 1 is proved in the case where $n \ge 39$ and $\lambda \le \sqrt{n}/2$. □

The end of the proof of Theorem 1 is based on numerical computations on the one hand and on the following considerations about the slope of the function $C_{\lambda,n} = \exp(2\lambda^2)P(D_n^- > \lambda)$ on the other hand.

PROPOSITION 2. *Let $n \geq 2$ and λ be such that $0 < \lambda < \sqrt{n}$. Then, we have*

(i) $$\frac{d}{d\lambda}C_{\lambda,n} \geq 0, \quad \textit{whenever } \frac{\sqrt{n}}{2} \leq \lambda,$$

(ii) $$\sum_{j=1}^{n-1} j^{j-1}(n-j)^{(n-j)}n^{-n}\binom{n}{j} \leq 1,$$

(iii) $$\frac{d}{d\lambda}C_{\lambda,n} \leq 3.61, \quad \textit{whenever } \lambda \geq \frac{1}{2}.$$

PROOF. (i) Setting $L_{\lambda,n}(j) = \log(\exp(2\lambda^2)p_{\lambda,n}(j))$, it is enough to show that the derivatives w.r.t. λ of the functions $L_{\lambda,n}(j)$ are negative when $\sqrt{n}/2 \leq \lambda$, for *all* integers j with $0 \leq j < n - \lambda\sqrt{n}$. Now,

$$\frac{d}{d\lambda}L_{\lambda,n}(j) = 4\lambda + \frac{j}{\lambda(j+\lambda\sqrt{n})} - \frac{\lambda n^2}{(j+\lambda\sqrt{n})(n-j-\lambda\sqrt{n})},$$

recalling that $\varepsilon = \lambda/\sqrt{n}$ and setting $x = \varepsilon + j/n$ (note that $x \in [0,1]$), we get

$$\frac{d}{d\lambda}L_{\lambda,n}(j) = -\frac{n\varepsilon^2(1-2x)^2 - (x-\varepsilon)(1-x)}{\lambda x(1-x)}. \tag{2.12}$$

From (2.12), it is clear that $dL_{\lambda,n}(0)/d\lambda \leq 0$, so we may now assume that $x \geq \varepsilon + 1/n$. Thus, $\varepsilon \geq \frac{1}{2}$ implies

$$\begin{aligned}\frac{d}{d\lambda}L_{\lambda,n}(j) &\leq -\frac{n(1-2x)^2 - 2(2x-1)(1-x)}{4\lambda x(1-x)}\\ &= -\frac{(n+1)(x-\frac{1}{2})(x-(n+2)/(2n+2))}{\lambda x(1-x)}.\end{aligned}$$

Then, noticing that $1 \geq x \geq \varepsilon + 1/n \geq \frac{1}{2} + 1/n > (n+2)/(2n+2)$, we get $dL_{\lambda,n}(j)/d\lambda < 0$, for any j with $0 \leq j < n - \lambda\sqrt{n}$, which proves (i).

(ii) Note that

$$\left.\frac{d}{d\lambda}P(D_n^- > \lambda)\right|_{\lambda=0} = \sqrt{n}\left(\sum_{j=1}^{n-1} j^{j-1}(n-j)^{(n-j)}n^{-n}\binom{n}{j} - 1\right),$$

since $P(D_n^- > \lambda)$ is nonincreasing w.r.t. λ, (ii) follows easily.

(iii) The proof of (iii) involves some crude bounds on $C_{\lambda,n}$ [namely, (2.13) and (2.14)] that we establish first.

We assume first that $n\varepsilon \geq 2$ and $\varepsilon \leq \frac{1}{2}$. Then, starting from Proposition 1 and using the bound $\exp(0.4/(ns)) \leq \exp(0.3)$, we proceed exactly as in the proof of Theorem 1 in the case where $n \geq 39$ and $\varepsilon \leq \frac{1}{2}$, to obtain this time,

$$C_{\lambda,n} \leq \exp(0.3)\left(I_{1,0}(\lambda) - \frac{\varepsilon}{3}I_{1,1}(\lambda) + \frac{\varepsilon^2}{6}I_{2,1}(\lambda)\right).$$

Using Lemma 4 and $2n^{-1/2} \leq \lambda \leq \sqrt{n}/2$, the above bound becomes

$$\begin{aligned} C_{\lambda,n} &\leq \exp(0.3) + \frac{2\lambda}{3\sqrt{n}}\exp(0.3)\left(-1 + \left(\lambda + \frac{1}{4\lambda}\right)n^{-1/2}\right) \\ &\leq \exp(0.3) + \frac{2\lambda}{3\sqrt{n}}\exp(0.3)\left(-\frac{3}{8}\right). \end{aligned}$$

Combining the latter inequality with (i), we get

$$C_{\lambda,n} \leq \exp\left(\left(\frac{8}{n}\right) \vee 0.3\right), \tag{2.13}$$

for any integer n ($n \geq 4$) and any positive λ.

With the help of (ii), we can build another bound for $U_{\lambda,n}$ which is more efficient than (2.13) for small values of n. In fact, let h be defined as in Lemma 1, then

$$p_{\lambda,n}(j) = \lambda\sqrt{n}\binom{n}{j}j^{j-1}(n-j)^{n-j}n^{-n}\left(\frac{j}{j+\lambda\sqrt{n}}\right)\exp\left(-h\left(1-\varepsilon-\frac{j}{n},\varepsilon\right)\right);$$

so, as a crude application of Lemma 1, we get

$$p_{\lambda,n}(j) \leq \lambda\sqrt{n}\,j^{j-1}(n-j)^{n-j}n^{-n}\exp(-2\lambda^2).$$

Thus, (ii) implies

$$C_{\lambda,n} \leq \lambda\sqrt{n} + p_{\lambda,n}(0)\exp(2\lambda^2). \tag{2.14}$$

We can now finish the proof of (iii). It comes from (2.12) that, on the one hand, $dL_{\lambda,n}(0)/d\lambda < 0$, and on the other hand, $dL_{\lambda,n}(j)/d\lambda \leq 1/\lambda$ for all $1 \leq j < n - \lambda\sqrt{n}$. Hence,

$$\frac{d}{d\lambda}C_{\lambda,n} \leq \frac{\left(C_{\lambda,n} - p_{\lambda,n}(0)\exp(2\lambda^2)\right)}{\lambda}.$$

We finally combine the latter inequality with either (2.13), in the case where $n \geq 14$, or (2.14), in the case where $n \leq 13$, to get, for any $\lambda \geq \frac{1}{2}$,

$$\frac{d}{d\lambda}C_{\lambda,n} \leq \left(2\exp\left(\frac{4}{7}\right)\right) \vee \sqrt{13} \leq 3.61. \qquad \square$$

PROOF OF THEOREM 1 (where $n \leq 38$ or $\lambda > \sqrt{n}/2$). Using Proposition 2(i) and the first part of the proof of Theorem 1, we may assume that $n \leq 38$. Then, the condition on λ in the statement of Theorem 1 reduces to $\lambda \geq \sqrt{\log(2)/2}$, so we may assume a fortiori that $\lambda \geq \frac{1}{2}$.

Now, setting $\eta = 10^{-2}$, let $\Lambda_{\eta,n} = \{\frac{1}{2} + k\eta / k \in \mathbb{N}\} \cap [\frac{1}{2}, \sqrt{n}[$. One can check with a computer [starting from the exact formula (2.3)] that

$$\max_{n \le 38} \sup_{\lambda \in \Lambda_{\eta,n}} C_{\lambda,n} \le 0.951. \tag{2.15}$$

Combining (2.15) with Proposition 2(iii), we finally get

$$\max_{n \le 38} \sup_{1/2 \le \lambda < \sqrt{n}} C_{\lambda,n} \le 0.951 + \eta \cdot 3.61 \le 0.9871 < 1,$$

which implies that Theorem 1 holds in the case where $n \le 38$. □

REFERENCES

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

BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256.

CSÁKI, E. (1974). Investigations concerning the empirical distribution function. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **23** 239–327. [In Hungarian; English translation in *Selected Transl. Math. Statist. Probab.* **15** (1981) 229–317.]

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. C. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **33** 642–669.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

HU, I. (1985). A uniform bound for the tail probability of Kolmogorov–Smirnov statistics. *Ann. Statist.* **13** 811–826.

KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* **4** 83–91.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SMIRNOV, N. V. (1944). Approximate laws of distribution of random variables from empirical data. *Uspekhi Mat. Nauk* **10** 179–206. (In Russian.)

UNIVERSITÉ DE PARIS–SUD
U.A. CNRS 743 "STATISTIQUE APPLIQUÉE"
MATHÉMATIQUES
BÂT. 425
91405 ORSAY CEDEX
FRANCE