THE DVORETZKY–KIEFER–WOLFOWITZ INEQUALITY WITH SHARP CONSTANT: MASSART'S 1990 PROOF

SEMINAR, SEPT. 28, 2011

R. M. Dudley

A. Dvoretzky, J. Kiefer, and J. Wolfowitz (1956) proved the "Dvoretzky–Kiefer–Wolfowitz" (DKW) inequality, namely that there is a constant $D < +\infty$ such that for any distribution function F on $\mathbb{R}$ and its empirical distribution functions F_n, we have for every $u > 0$,

$$\Pr(\sqrt{n}\sup_x |(F_n - F)(x)| > u) \leq D\exp(-2u^2). \tag{1}$$

Massart (1990, Ann. Prob.) proved the following:

Theorem 1 (Massart)**.** *The inequality (1) holds with the constant $D = 2$.*

Remark. The constant $D = 2$ is best possible because, for a Brownian bridge y and F continuous, we have as shown by Kolmogorov (1933) and as also follows from the Komlós–Major–Tusnády approximation,

$$\lim_{n \to\infty} \Pr(\sqrt{n}\sup_x |(F_n - F)(x)| > u) = \Pr(\sup_t |y_t| > u)$$

$$= 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2u^2).$$

As u becomes large, the latter sum is asymptotic to $2\exp(-2u^2)$.

The distribution of $\sup_x |(F_n - F)(x)|$ is the same for all continuous F. It will suffice to prove Theorem 1 for the $U[0,1]$ distribution U. For a discontinuous F, $\sup_x |(F_n - F)(x)|$ is stochastically smaller than for F continuous.

Set $\alpha_n(t) := \sqrt{n}(U_n - U)(t)$ for $0 \leq t \leq 1$, let $D_n^+ := \sup_t \alpha_n(t)$, $D_n^- := \sup_t(-\alpha_n(t))$, and

$$D_n := \sup_t |\alpha_n(t)| = \max(D_n^+, D_n^-).$$

We have the following symmetry:

Proposition 2. *For any $n = 1, 2, ...,$ D_n^+ and D_n^- are equal in distribution.*

Proof. (brief) Let $X_1, ..., X_n$ be the i.i.d. $U[0,1]$ variables on which U_n is based. Let $Y_j := 1 - X_j$ for $j = 1, ..., n$. Then $Y_1, ..., Y_n$ are i.i.d. $U[0,1]$ and for them, D_n^- and D_n^+ are interchanged for those of the X_j. $\square$

Massart (1990, Theorem 1) gives the following fact, which is interesting in itself and implies Theorem 1 (see the Remarks after it):

Theorem 3. *For any $n = 1, 2, ...$ and any $\lambda \geq \lambda_n$ where $\lambda_n = \min(\sqrt{\log 2/2}, \zeta n^{-1/6})$ and $\zeta := 1.0841$, we have*

$$\Pr(D_n^- > \lambda) \leq \exp(-2\lambda^2). \tag{2}$$

Remarks. If $\exp(-2\lambda^2) \leq 1/2$, then $\lambda \geq \sqrt{(\log 2)/2}$, which implies the hypothesis of Theorem 3. Also, Proposition 2 implies $\Pr(D_n > \lambda) \leq 2\Pr(D_n^- > \lambda)$. Further, Theorem 1 holds trivially if $2\exp(-2\lambda^2) > 1$, so it suffices to prove Theorem 3 to prove Theorem 1 in all cases.

Proof. If for a given $\lambda > 0$,

$$D_n^- = \sqrt{n} \sup_{0 \leq t \leq 1} (t - U_n(t)) > \lambda, \tag{3}$$

then $t - U_n(t) = \lambda/\sqrt{n}$ for some t, because between its downward jumps at the observations X_j, $t - U_n(t)$ is continuously increasing. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ (almost surely) be the order statistics of $X_1, ..., X_n$. Thus

$$\sup_{0 \leq t \leq 1} t - U_n(t) = \max_{1 \leq k \leq n} X_{(k)} - (k-1)/n$$

(the supremum occurs just to the left of some $X_{(k)}$); the supremum is strictly positive with probability 1 because $X_{(1)} > 0$. So if (3) holds there is a smallest $k = 1, ..., n$ with $X_{(k)} - (k-1)/n > \lambda/\sqrt{n}$. Letting $X_{(0)} := 0$, we must have $t - U_n(t) = \lambda/\sqrt{n}$ for some t with $X_{(k-1)} < t < X_{(k)}$. Let $\tau_n := \tau_n(\lambda)$ be the least $t \in [0,1]$ such that $t - U_n(t) = \lambda/\sqrt{n}$ if one exists ($D_n^- > \lambda$), otherwise let $\tau_n = 2$. If $\tau_n < 2$, then for some $j = 0, 1, ..., n-1$, $\tau_n - j/n = \lambda/\sqrt{n}$, i.e. $\tau_n = \frac{j}{n} + \frac{\lambda}{\sqrt{n}}$, which implies that

$$j < n - \lambda\sqrt{n}. \tag{4}$$

If $\lambda \geq \sqrt{n}$ then $\Pr(D_n^- > \lambda) = 0$, implying the conclusion of the theorem, so suppose $\lambda < \sqrt{n}$. Let $J \geq 0$ be the largest integer less than $n - \lambda\sqrt{n}$.

The following fact, according to Massart (1990), is due to Smirnov (1944).

Proposition 4 (Smirnov)**.** *For each λ with $0 < \lambda < \sqrt{n}$ and $\varepsilon := \lambda/\sqrt{n}$, and each $j = 1, ..., J$,* $\Pr(\tau_n = \varepsilon + j/n) = p_{\lambda,n}(j)$ *where*

$$p_{\lambda,n}(j) = \lambda\sqrt{n}(j + \lambda\sqrt{n})^{j-1}(n - j - \lambda\sqrt{n})^{n-j} n^{-n} \binom{n}{j}. \tag{5}$$

For $j = 0$, $\Pr(\tau_n = \varepsilon) = (1-\varepsilon)^n$.

Proof. For each n, $\varepsilon = \lambda/\sqrt{n}$ and $i = 1, ..., J$ let $A_i := \left\{X_{(i)} \leq \varepsilon + \frac{i-1}{n}\right\}$. Here is a

Claim. We have $\{\tau_n = \varepsilon\} = A_1^c$ and for $j = 1, ..., J$, $\{\tau_n = \varepsilon + \frac{j}{n}\} = \left(\bigcap_{1 \leq i \leq j} A_i\right) \cap A_{j+1}^c$.

Proof of Claim. This is straightforward and omitted.

Now continuing the proof of Proposition 4, $X_{(1)}$ has distribution function $1 - (1-x_1)^n$ and so density $n(1-x_1)^{n-1}$ for $0 \leq x_1 \leq 1$. For $1 \leq i < n$, conditional on $X_{(1)}, ..., X_{(i)}$, $X_{(i+1)}$ is the least of $n - i$ variables i.i.d. $U[X_{(i)}, 1]$ (this conditional distribution only depends on $X_{(i)}$). Thus $\Pr(X_{(i+1)} \geq t | X_{(i)}) = ((1-t)/(1-X_{(i)}))^{n-i}$ and the conditional density of $X_{(i+1)}$ given $X_{(i)} = x_i$ is $(n-i)(1-x_{i+1})^{n-i-1}/(1-x_i)^{n-i}$. Iterating, the joint density of $X_{(1)}, ..., X_{(j+1)}$ is $n!(1-x_{j+1})^{n-j-1}/(n-j-1)!$ for $0 \leq x_1 \leq x_2 \leq \cdots \leq x_j \leq x_{j+1} \leq 1$ and 0 elsewhere. Thus

$$\Pr\left(\tau_n = \varepsilon + \frac{j}{n}\right) = \frac{n!}{(n-j-1)!} I_j J_j$$

where

$$I_j := \int_0^{\varepsilon} dx_1 \int_{x_1}^{\varepsilon+1/n} dx_2 \cdots \int_{x_{j-1}}^{\varepsilon+(j-1)/n} dx_j, \tag{6}$$

$$J_j := \int_{\varepsilon+(j/n)}^{1} (1-x_{j+1})^{n-j+1} dx_{j+1} = \left(1-\varepsilon-\frac{j}{n}\right)^{n-j} / (n-j), \tag{7}$$

and a $(j+1)$-fold integral equals the given product because $x_j \leq \varepsilon + (j-1)/n$ and $x_{j+1} \geq \varepsilon + j/n$ imply $x_j \leq x_{j+1}$. Also, $(j/n) + \varepsilon < 1$ follows from $j \leq J < n - \lambda\sqrt{n}$. So, to prove Proposition 4 it remains to show that

$$I_j = \varepsilon \left(\frac{j}{n} + \varepsilon\right)^{j-1} / j!. \tag{8}$$

This will be proved for each ζ with $0 < \zeta \leq 1 - \frac{j-1}{n}$ in place of ε and by induction on j. Equation (8) holds for $j = 1$. Assume it holds for $j-1$ for some j with $2 \leq j \leq J$, and for each ζ with $0 \leq \zeta \leq 1 - (j-2)/n$ in place of ε. In the integral (6) make the changes of variables $\xi_i = x_i - x_1$ for $i = 2, ..., j$ and let $\delta := \varepsilon - x_1 + \frac{1}{n}$. Then

$$I_j = \int_0^{\varepsilon} dx_1 \int_0^{\delta} d\xi_2 \int_{\xi_2}^{\delta+1/n} d\xi_3 \cdots \int_{\xi_{j-1}}^{\delta+(j-2)/n} d\xi_j. \tag{9}$$

Applying the induction hypothesis to the inner $(j-1)$-fold integral in (9) we get

$$\begin{aligned} I_j &= \int_0^\varepsilon dx_1 \delta \left(\frac{j-1}{n} + \delta\right)^{j-2} /(j-1)! \\ &= \int_0^\varepsilon dx_1 \left(\varepsilon - x_1 + \frac{1}{n}\right)\left(\frac{j}{n} + \varepsilon - x_1\right)^{j-2} /(j-1)!, \end{aligned}$$

so setting $y := \varepsilon - x_1$ gives

$$(j-1)! I_j = \int_0^\varepsilon \left(y + \frac{1}{n}\right)\left(y + \frac{j}{n}\right)^{j-2} dy.$$

An integration by parts and calculation then give that $(j-1)! I_j$ equals $\varepsilon\left(\varepsilon + \frac{j}{n}\right)^{j-1}/j$, which proves (8) and thus Proposition 4. $\square$

Next, for a Brownian bridge $y = \{y_t\}_{0 \le t \le 1}$, and $\lambda > 0$, let $\tau_\lambda := \inf\{s > 0 : \ y_s \ge \lambda\}$ if this is less than 1, otherwise let $\tau_\lambda = 2$. The following fact is known, e.g. it follows from part of Lemma 1.3 of Bretagnolle and Massart, Ann. Prob. 1989: for $0 < s < 1$,

$$\begin{aligned} \Pr(\tau_\lambda \le s) &= 1 - \Phi\left(\frac{\lambda}{\sqrt{s(1-s)}}\right) \\ &\quad + \exp(-2\lambda^2)\left(1 - \Phi\left(\frac{(1-2s)\lambda}{\sqrt{s(1-s)}}\right)\right), \end{aligned} \tag{10}$$

where Φ is the standard normal distribution function. Let $f_\lambda(s) := d\Pr(\tau_\lambda \le s)/ds$. From (10) and a calculation one gets

$$f_\lambda(s) = \frac{\lambda}{\sqrt{2\pi}} \frac{1}{s^{3/2}\sqrt{1-s}} \exp\left(-\frac{\lambda^2}{2s(1-s)}\right). \tag{11}$$

From the definitions we have for each $\lambda > 0$

$$\Pr(D_n^- > \lambda) = \sum_{0 \le j < n - \lambda\sqrt{n}} p_{\lambda,n}(j). \tag{12}$$

A well-known, simple reflection proof gives

$$\exp(-2\lambda^2) = \Pr(\tau_\lambda \le 1) = \int_0^1 f_\lambda(s)ds. \tag{13}$$

The next fact is one of the main steps in the proof of Theorem 3.

Proposition 5. *Let j be a nonnegative integer with $j < n - \lambda\sqrt{n}$. Let $s = (2\varepsilon/3) + j/n$, $s' = 1 - s$, and $v_n(s) = 1/\left(s(s^2 - 1/(4n^2)\right))$. If $n\varepsilon \ge 2$, then*

$$p_{\lambda,n}(j) \le \frac{1}{n}\left(1 - \frac{\varepsilon}{3s'} + \frac{\varepsilon^2}{6s'^2}\right) E_{n,\lambda,s} \tag{14}$$

where

$$E_{n,\lambda,s} = \exp\left(\frac{0.4}{ns} - \frac{\varepsilon^2}{24n}(v_n(s) + v_n(s'))\right) f_\lambda(s). \tag{15}$$

Some lemmas and other facts will be used to prove Proposition 5. The first one has implications for the binomial distribution.

Lemma 6. *Let $0 < \varepsilon < q = 1 - p < 1$. Let*

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

For $t \ge 0$ let

$$g(t) = t - \frac{t^2}{2(1 + 2t/3)} - \log(1 + t).$$

Then
(i) g is a strictly increasing convex function with $g(t)/t \to 1/4$ as $t \to +\infty$,

(ii) For $t := \varepsilon/(q-\varepsilon)$,

$$h(p,\varepsilon) \geq \frac{\varepsilon^2}{2(p+\varepsilon/3)(q-\varepsilon/3)} + \varepsilon g(t)/t.$$

Proof. For (i), we have $g(0) = 0$, and for all $t > 0$

$$g'(t) = (t^3/9)(1+2t/3)^{-2}(1+t)^{-1} > 0.$$

To see that g' is increasing, note that $(3+2t)^2(1+t)/t^3$ is decreasing. So g is convex. As $t \to +\infty$, $g(t)/t \to 1 - 1/(4/3) = 1/4$ as stated.

For (ii), the proof by straightforward calculation is omitted. □

A consequence for the binomial distribution is:

Theorem 7. *Let* S_n *be a binomial* (n,p) *random variable and suppose that* $0 < \varepsilon < q = 1-p$. *Then*

$$\Pr(S_n - np \geq n\varepsilon) \leq \exp\left(-\frac{n\varepsilon^2}{2(p+\varepsilon/3)(q-\varepsilon/3)}\right).$$

Proof. The probability is less or equal to

$$\left\{\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon}\left(\frac{q}{q-\varepsilon}\right)^{q-\varepsilon}\right\}^n = \exp(-nh(p,\varepsilon))$$

by Chernoff's inequality (1952, Ann. Math. Statist.; Massart also mentions Cramér's name here). Then we can apply Lemma 6(ii). □

Now, let's begin the proof of Proposition 5 for $j \geq 1$. By Stirling's formula with error bounds (Feller, vol. I), for $1 \leq j < n$,

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

where $C_j := \exp(-1/(12j+1))$. One plugs that bound into (5). Recalling $h(\cdot,\cdot)$ as in Lemma 6 and that $s = 2\varepsilon/3 + j/n$ and

$s' = 1 - s$, it follows that

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

Define $\psi(t)$ for $0 \leq t < \infty$ by

$$\psi(t) = -\log(1+t) + \frac{3}{2}\log\left(1 + \frac{2t}{3}\right). \tag{16}$$

Setting $t = \varepsilon/(s - 2\varepsilon/3)$, which agrees with the definition of t in Lemma 6(ii)), that Lemma gives
(17)

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

To bound the exponentiated term, we have the following.

Lemma 8. *Let $\theta := 0.4833$. Let g and ψ be the functions defined in Lemma 6 and (16) respectively. For $t > 0$ and $\nu > 0$ let*

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

Then $T(\nu, t) > 0$ for $0 < t \leq \nu$.

Proof. (a) First suppose $\nu = t > 0$. Then, straightforwardly we have

$$\frac{d}{dt}\left(\frac{T(t,t)}{t^2}\right) = \frac{-2\theta/3 - t/3 + t^2/9}{(1 + 2t/3)^2}. \tag{18}$$

The quadratic in the numerator has two roots, one being negative, and a positive root $t_0 = (3 + \sqrt{9 + 24\theta})/2$. The derivative in (18) equals $-2\theta/3 < 0$ when $t = 0$, so t_0 is a relative minimum of $T(t,t)/t^2$ and is the absolute minimum for $t > 0$. We find $T(t_0,t_0)/t_0^2 \doteq 5.05 \cdot 10^{-6} > 5 \cdot 10^{-6} > 0$, so $T(t,t) > 0$ for all $t > 0$.

(b) Now for general $0 < t \le \nu$, T is a quadratic polynomial in ν for fixed t. Its derivative with respect to ν is $2\nu g(t) - t\psi(t)$. If

$$2g(t) - \psi(t) > 0 \tag{19}$$

then $T(\nu,t)$ is increasing in ν for all $\nu > t$ and so $T(\nu,t) > 0$. Or, if the discriminant $D(t)$ of the quadratic polynomial satisfies

$$D(t) = t^2\psi^2(t) - 4g(t)\theta t^2/(1 + 2t/3) < 0$$

or equivalently

$$\Delta(t) := (1 + 2t/3)\psi^2(t) - 4\theta g(t) < 0 \tag{20}$$

then $T(t,\nu)$ remains positive for all $\nu \ge t$ as it is for $\nu = t$ and has no roots. So to prove Lemma 8 it will suffice to show that

(i) $2g(t) - \psi(t) > 0$ for all $t \ge 3.37$.

(ii) $\Delta(t) < 0$ for $0 < t \le 3.37$.

Proof of (i). We have

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

We have $2t^2 - 2t - 3 > 0$ for $t > (1+\sqrt{7})/2$, thus $2g - \psi$ is increasing for such t. Since $(1+\sqrt{7})/2 \doteq 1.823 < 3.37$ we have that for $t \ge 3.37$,

$$2g(t) - \psi(t) \ge 2g(3.37) - \psi(3.37) \doteq .000775 > 7 \cdot 10^{-4} > 0,$$

proving (i).

For (ii), Massart states that $R(t) = \left(1 + \frac{2t}{3}\right)\psi^2(t)/g(t)$ is increasing for $t > 0$ (which is only needed for $t \le 3.37$). The proof given in the first half of p. 1275 of his paper is not correct, as the function $R_0 = \psi^2/g$ is not increasing. Nevertheless it appears true that $R(t)$ is increasing by examining computed values of it on a grid, $0, 0.01, 0.02, ..., 3.37$. (In a recent email, Massart said he had independently verified the increasing property with a Matlab plot, but as of today, neither of us seems to have a rigorous proof.) It follows that (ii) holds.

To continue the proof of Proposition 5, three Claims will be used.

Claim 1. Let $\beta := 0.826$. Then for any $x \in [0, 1]$,

$$(1+2x)^{-1/2} \leq \left(1 - x + \frac{3x^2}{2}\right) \exp(-\beta x^3).$$

This is proved by a straightforward calculation, omitted.

Claim 2. For $\varepsilon = \lambda/\sqrt{n}$ as usual and j, s, s', and $v_n(\cdot)$ as defined in Proposition 5, if $n\varepsilon \geq 2$ and $ns' \geq 1$, we have

$$\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2} \leq \left(1 - \frac{\varepsilon}{3s'} + \frac{\varepsilon^2}{6s'^2}\right) \exp\left(-\frac{\varepsilon^2 v_n(s')}{24n}\right). \tag{21}$$

Proof of Claim 2. One can first check easily that $\varepsilon \leq 3s'$. Then we can apply Claim 1 with $x = \varepsilon/(3s')$. The rest of the proof is a straightforward calculation, using the hypotheses. □

Claim 3. For v_n as defined in Proposition 5, and any $\varepsilon > 0$ and $s > 0$ satisfying $1/n \leq \varepsilon \leq 3s/2$, we have

$$\left(1 + 12n\left(s - \frac{2\varepsilon}{3}\right)\right)^{-1} \geq \frac{1}{12ns} + \frac{\varepsilon^2 v_n(s)}{24n}. \tag{22}$$

Proof of Claim 3. This is another routine calculation. □

Now to finish the proof of Proposition 5 for $j \geq 1$, recall (17). Note that $t = n\varepsilon/j \leq n\varepsilon$. Lemma 8 with $\nu = n\varepsilon$ gives

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

Claim 2 gives an upper bound for $\left(1 + \frac{2\varepsilon}{3s'}\right)^{-1/2}$ and Claim 3 gives one for C_j. Noting that $\theta - 1/12 < 0.4$ finishes the proof for $j \geq 1$.

Proof of Proposition 5 for $j = 0$. In this case $s = 2\varepsilon/3$. We have $p_{\lambda,n}(0) = (1-\varepsilon)^n$ by Proposition 4. It's natural to define $h(p,\delta)$ when $\delta = q$ as $-\log(p)$ since for fixed q with $0 < q < 1$, $(q-\delta)\log((q-\delta)/q) \to 0$ as $\delta \uparrow q$, so that will be done. Then $(1-\varepsilon)^n = \exp(-nh(1-\varepsilon,\varepsilon))$. To apply Lemma 6, we will have $t = \varepsilon/(q-\varepsilon)$ which it's natural to define as $+\infty$ in this case with $q = \varepsilon$; for $t \to +\infty$ the limit of $g(t)/t$ is $1/4$. Then Lemma 1 gives

$$p_{\lambda,n}(0) = (1-\varepsilon)^n = \exp(-nh(1-\varepsilon,\varepsilon) \leq \exp\left(-\frac{\lambda^2}{2ss'} - \frac{n\varepsilon}{4}\right).$$

Define $H(\nu)$ for $\nu > 0$ by

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

Then it's straightforward to check that

$$(1-\varepsilon)^n \leq \frac{\lambda}{\sqrt{2\pi n}} s^{-3/2} \exp\left(\frac{0.4}{n\varepsilon}\right) \exp\left(-\frac{\lambda^2}{2ss'}\right) \exp(-H(n\varepsilon).$$

We have $H'(\nu) = \frac{1}{4} - \frac{0.4}{\nu^2} - \frac{1}{2\nu} = 0$ if and only if $\nu^2 - 2\nu - 1.6 = 0$. The only positive root of this is at $\nu_0 = 1 + \sqrt{2.6}$. This is the minimum of H for $\nu > 0$ because $H(\nu) \to +\infty$ as $\nu \to +\infty$.

Thus $H(\nu) \geq H(\nu_0) \geq 0.01534 > 0$ for all $\nu > 0$. It follows that

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

where the second equation follows on expanding $f_\lambda(s)$ by (11).

Since $n\varepsilon \geq 2$, it follows that

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

from which it follows that

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

Combining gives

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

Via Claim 2, (14) follows and Proposition 5 is proved. $\square$

In the proof of Theorem 3, the integral in (13) will be compared to Riemann sums and thereby to the sums in (12). The comparison will use the next lemma.

Lemma 9. *Let $0 < \delta \leq s \leq 1-\delta$ and $s' = 1-s$. If G is a continuous function with $G(x) > 0$ for $s-\delta \leq x \leq s+\delta$ and $\log(G)$ is convex, then for any $\lambda > 0$,*

$$\frac{1}{2\delta}\int_{s-\delta}^{s+\delta} G(u)\exp\left(-\frac{\lambda^2}{2u(1-u)}\right) du \geq G(s)\cdot\exp\left(-\frac{\lambda^2}{2ss'}\right)\cdot$$
$$\cdot\exp\left(-\frac{\lambda^2\delta^2}{6}\left(\left(s(s^2-\delta^2)\right)^{-1}\left(s'(s'^2-\delta^2)\right)^{-1}\right)\right).$$

Proof. Jensen's inequality will be applied twice. Both times the probability measure is the uniform distribution $U[s-\delta, s+\delta]$. First the convex function is exp, then second it is $\log(G)$. We get

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

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

The function $u \mapsto 1/u$ has a positive fourth derivative. We can apply Simpson's rule with remainder as given by Davis and Polonsky (1974, 25.4.5 p. 886, in Abramowitz and Stegun, *Handbook of Mathematical Functions*): if f has a continuous fourth derivative $f^{(4)}$, $h > 0$, $x_i = x_0 + ih$ for $i = 0, 1, 2$, and $f_i = f(x_i)$, then

$$\int_{x_0}^{x_2} f(x)dx = \frac{h}{3}[f_0 + 4f_1 + f_2] - \frac{h^5}{90}f^{(4)}(\xi)$$

for some $\xi \in [x_0, x_2]$. Thus if $f^{(4)} \geq 0$,

$$\int_{x_0}^{x_2} f(x)dx \leq \frac{h}{3}[f_0 + 4f_1 + f_2]. \tag{23}$$

Thus

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

Next,

$$\frac{1}{2\delta}\int_{s-\delta}^{s+\delta}(1-u)^{-1}du = \frac{1}{2\delta}\int_{s'-\delta}^{s'+\delta} v^{-1}dv,$$

and the Lemma follows. $\square$

Next are some identities for special integrals.

Lemma 10. *For any $a \geq 0$, $b \geq 0$, and $\lambda > 0$, let $I_{a,b}(\lambda) =$*

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

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

Proof. Clearly $I_{a,b} \equiv I_{b,a}$. For any u with $0 < u < 1$,

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

which implies for any $a \geq 1$ and $b \geq 1$ that

$$I_{a,b} = I_{a-1,b} + I_{a,b-1}. \tag{24}$$

For $a = b = 1$, using $\exp(-2\lambda^2) = \int_0^1 f_\lambda(s)ds$ (13) where

$$f_\lambda(s) = \frac{\lambda}{\sqrt{2\pi}} \frac{1}{s^{3/2}\sqrt{1-s}} \exp\left(-\frac{\lambda^2}{2s(1-s)}\right)$$

(11), we get (i). Next, differentiating with respect to λ gives (ii). Then, applying (24) with $a = 1$ and $b = 2$, (iii) follows from (i) and (ii). So Lemma 10 is proved. □

Now we can prove Theorem 3 under some conditions.

Proof of Theorem 3 for $n \geq 39$ and $\lambda \leq \sqrt{n}/2$. Since $n \geq 39$, the hypothesis on λ in Theorem 3 becomes $\lambda \geq \zeta n^{-1/6}$. Thus $n\varepsilon = \lambda\sqrt{n} \geq 3.6764$. Define the function $y(\cdot)$ by $y(x) := (e^x - 1)/x$ for $x > 0$. It's easily seen (e.g. from the Taylor series) that this function is increasing. Recalling that $s \geq 2\varepsilon/3$ we have

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

Let $\mu := 0.4345$. Then $\exp\left(\frac{0.4}{ns}\right) \le 1 + \frac{\mu}{ns}$. Applying Proposition 5, we get another upper bound for $p_{\lambda,n}(j)$:

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

Preparing to apply Lemma 9, note that $z(\cdot)$ defined by $z(x) = \log(6x^2 - 2x + 1) - (5/2)\log(x)$ is convex for $x > 0$: calculation gives $z''(x)x^2(6x^2 - 2x + 1)^2 = h(x)$ where h is a positive quartic polynomial. Let $G(u) :=$

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

Then for $c := \log(\lambda/(6\sqrt{2\pi\varepsilon})$, a constant with respect to u, we have

$$\log(G(u)) = c + \log\left(1 + \frac{\mu}{nu}\right) - \frac{3}{2}\log(u) + z\left(\frac{1-u}{\varepsilon}\right),$$

in which each term is convex, so $\log G(\cdot)$ is convex. So Lemma 9 with $\delta = 1/(2n)$ gives $p_{\lambda,n}(j) \le$

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

Summing over j in (12) and using also (13) we get, in the notation of Lemma 10,

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

By Lemma 4 and simple calculations we then get

$$\frac{3\sqrt{n}}{2\lambda}(\exp(2\lambda^2)\Pr(D_n^- > \lambda) - 1)$$

$$\leq \eta_n(\lambda) := -1 + \left(\lambda + \frac{1}{4\lambda} + \frac{3\mu}{\lambda} + \frac{3\mu}{2\lambda^3}\right) n^{-1/2}$$

$$-\frac{\mu}{2}\left(4 + \frac{1}{\lambda^2}\right) n^{-1} + \frac{\mu}{2}\left(4\lambda + \frac{1}{\lambda}\right) n^{-3/2}. \tag{25}$$

Remark. Smirnov (1944), as quoted by Massart, had given the asymptotic expansion

$$\Pr(D_n^- > \lambda) \sim \exp(-2\lambda^2)\left(1 - \frac{2\lambda}{3\sqrt{n}} + O(1/n)\right) \tag{26}$$

if $\lambda = O(n^{1/6})$. By contrast, Massart's inequality (25) is one-sided, but the first term -1 on the right confirms that the term $-2\lambda/(3\sqrt{n})$ in Smirnov's expansion is valid non-asymptotically in a one-sided sense, which is what one wants.

It is easy to check that η_n is convex in λ for each n. Thus, to show that $\eta_n(\lambda) < 0$ for $\zeta n^{-1/6} \leq \lambda \leq \sqrt{n}/2$ it will suffice to show that $a_n := \eta_n(\zeta n^{-1/6}) < 0$ and $b_n := \eta_n(\sqrt{n}/2) < 0$.

It will be shown that a_n and b_n are decreasing in n for $n \geq 39$. We have

$$\begin{aligned}
a_n &= \eta_n(\zeta n^{-1/6}) = -1 + \Big(\zeta n^{-1/6} \\
&\quad + \frac{n^{1/6}}{\zeta}\left(\frac{1}{4} + 3\mu\right) + \frac{3\mu}{2}\frac{n^{1/2}}{\zeta^3}\Big) n^{-1/2} \\
&\quad -\frac{\mu}{2}\left(4 + \frac{n^{1/3}}{\zeta^2}\right) n^{-1} + \frac{\mu}{2}\left(4\zeta n^{-1/6} + \frac{n^{1/6}}{\zeta}\right) n^{-3/2} \\
&= -1 + \frac{3\mu}{2\zeta^3} + n^{-1/3}\left(\frac{1}{\zeta}\right)\left(\frac{1}{4} + 3\mu\right) + \left(\zeta - \frac{\mu}{2\zeta^2}\right) n^{-2/3} \\
&\quad -2\mu n^{-1} + \frac{\mu}{2\zeta} n^{-4/3} + 2\mu\zeta n^{-5/3}.
\end{aligned}$$

As $\zeta = 1.0841$ (Theorem 3) and $\mu = 0.4345$, $\zeta - \mu/(2\zeta)^2 > 0$. Terms with positive coefficients and negative powers of n, or with negative coefficients and positive powers of n, are decreasing. Just one term, $-2\mu n^{-1}$, is increasing. It will suffice to show that $(3/\zeta)n^{-1/3} - 2n^{-1}$ is decreasing for $n \geq 39$, or that $3x - 2.1682x^3$ is increasing for $0 < x \leq 1/39$. Indeed its derivative is positive there. A calculation shows that b_n is a linear combination of negative powers of n times positive coefficients, so it is also decreasing in n. We have $a_{39} \doteq -0.006382 < 0$ and $b_{39} \doteq -0.4238 < 0$, so both a_n and b_n are negative for all $n \geq 39$, and $\eta_n(\lambda) < 0$ for $\zeta n^{-1/6} \leq \lambda \leq \sqrt{n}/2$. So Theorem 3 is proved for $n \geq 39$ and $\lambda \leq \sqrt{n}/2$.

Let $C_{\lambda,n} = \exp(2\lambda^2)\Pr(D_n^- > \lambda)$.

Proposition 11. *Let $n \geq 2$ and let λ be such that $0 < \lambda < \sqrt{n}$. Then*
(i) For $\lambda \geq \sqrt{n}/2$, $\frac{d}{d\lambda}C_{\lambda,n} \leq 0$,
(ii) $\sum_{j=1}^{n-1} j^{j-1}(n-j)^{n-j}n^{-n}\binom{n}{j} \leq 1$,
(iii) For $\lambda \geq 1/2$, we have $\frac{d}{d\lambda}C_{\lambda,n} \leq 3.61$.

Proof. (i) Let $L_{\lambda,n}(j) = \log(\exp(2\lambda^2)p_{\lambda,n}(j)$ for $0 \leq j < n - \lambda\sqrt{n}$. Since

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

by (12), it will suffice to show that for each such j, $dL_{\lambda,n}/d\lambda < 0$ for $\lambda \geq \sqrt{n}/2$. The proof of this is by calculations, where the case $j = 0$ is relatively easy but separate, and the case $j \geq 1$ takes more but not especially long calculation. So (i) holds.

(ii) By (12) and (5) we have

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

By part (i), $\Pr(D_n^- > \lambda)$ is nonincreasing with respect to λ, so (ii) follows.

(iii) First suppose $n\varepsilon \geq 2$ and $\varepsilon \leq 1/2$. Using Proposition 5 and the bound $\exp(0.4/(ns)) \leq \exp(0.3)$, in the same way as in the proof of Theorem 3 for $n \geq 39$ and $\varepsilon \leq 1/2$, we now get

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

Using Lemma 10 and $2/\sqrt{n} \leq \lambda \leq \sqrt{n}/2$, it follows that

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

Combining with (i) gives

$$C_{\lambda,n} \leq \exp(\max((8/n), 0.3)) \tag{27}$$

for any integer $n \geq 4$ and any $\lambda > 0$.

By Lemma 6 one gets an alternate bound, useful for smaller values of n,

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

Thus by (ii),

$$C_{\lambda,n} \leq \lambda\sqrt{n} + p_{\lambda,n}(0)\exp(2\lambda^2). \tag{28}$$

For $j = 0,$, $dL_{\lambda,n}(0)/d\lambda < 0$, and for $1 \le j < n - \lambda\sqrt{n}$, $dL_{\lambda,n}(j)/d\lambda < 1/\lambda$. Thus

$$\frac{d}{d\lambda} C_{\lambda,n} \le \frac{1}{\lambda}\left(C_{\lambda,n} - p_{\lambda,n}(0)\exp(2\lambda^2)\right).$$

Combining this last inequality with (27) if $n \ge 14$, or (28) for $n \le 13$, we get

$$\frac{d}{d\lambda} C_{\lambda,n} \le \max(2\exp(4/7), \sqrt{13}) \le 3.61,$$

which proves (iii) and so Proposition 11. $\square$

Proof of Theorem 3 for $n \le 38$ or $\lambda > \sqrt{n}/2$. By Proposition 11 and the first part of the proof of Theorem 3, we can assume that $n \le 38$. Then the assumption on λ in Theorem 3 reduces to $\lambda \ge \sqrt{(\log 2)/2}$ which implies $\lambda > 1/2$.

Letting $\eta := 0.01$, let $\Lambda_{\eta,m} = \{\frac{1}{2} + k\eta : \ k \in \mathbb{N}\} \cap \left[\frac{1}{2}, \sqrt{n}\right)$. A computer calculation, reported by Massart (1990), gave

$$\max_{n \le 38} \max_{\lambda \in \Lambda_{\eta,n}} C_{\lambda,n} \le 0.951. \tag{29}$$

In a confirming computation, the maximum was found to be $\doteq 0.94955$, attained at $n = 38$ and $\lambda = 1/2$. Combining (29) with Proposition 11(iii), we get

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

which finishes the proof of Theorem 3 for $n \le 38$ and so completes its proof. $\square$

Komlós, Major, and Tusnády (1975, ZW) stated a sharp rate of convergence in Donsker's theorem, namely that on some probability space there exist X_i i.i.d. $U[0,1]$ and Brownian bridges Y_n

such that

$$(30) \quad P\left(\sup_{0\le t\le 1} |(\alpha_n - Y_n)(t)| > \frac{x + c\log n}{\sqrt{n}}\right) \;<\; Ke^{-\lambda x}$$

for all $n = 1, 2, \ldots$ and $x > 0$, where c, K, and λ are positive absolute constants.

More specifically, Bretagnolle and Massart (1989, Ann. Prob.) proved the following:

Theorem 12 (Bretagnolle and Massart)**.** *The approximation (30) of empirical processes by Brownian bridges holds with $c = 12$, $K = 2$ and $\lambda = 1/6$ for $n \ge 2$.*

The Dvoretzky–Kiefer–Wolfowitz–Massart inequality, Theorem 1, gives us some crude bounds so that we can see how large n needs to be for Theorem 12 to be effective. Namely, for any empirical process α_n, any Brownian bridge Y, and any $b > 0$, we have

$$\Pr(\sup_t |(\alpha_n - Y)(t)| \ge b) \le$$

$$\Pr(\sup_t |\alpha_n(t)| \ge b/2) + \Pr(\sup_t |Y(t)| \ge b/2) \le 4\exp(-b^2/2),$$

not depending on n. If $b \ge 2.97$ then $4\exp(-b^2/2) \le 0.05$, so α_n and Y_n will very likely be within b of each other in sup norm just because both will probably be bounded in absolute value by $b/2$. If we choose $x = 3\log n$ in (30) then with $\lambda = 1/6$ the bounds for probabilities on the right in (30) will decrease just at a moderate $O(1/\sqrt{n})$ rate. Then we would like $15(\log n)/\sqrt{n} < 2.97$ to get a bound better than the crude one, in other words $(\log n)/\sqrt{n} \le 0.198$. This does not hold for $n = 1000$, or 1300, but it does hold for $n = 1320$.

Bretagnolle and Massart's theorem is proved in more detail than they gave in my notes for a summer course on empirical processes in 1999 (MaPhySto, Denmark). I plan to include the proof, as well as that of Massart's (1990) theorem, in the second edition of my book *Uniform Central Limit Theorems*.

Z. W. Birnbaum and F. H. Tingey, *Ann. Math. Statist.* **22** (1951), pp. 592-596, in Sections 2 and 3, give a brief but perhaps sufficient proof of Proposition 4. They cite two references, one of which is by Smirnov, but from 1939, not 1944. The 1944 paper seems relatively hard to access. I found the Birnbaum and Tingey reference from the natural source, namely the book by G. R. Shorack and G. Wellner, *Empirical Processes with Applications to Statistics*, originally published by Wiley in 1986, reissued in the Classics in Applied Mathematics series of the Society for Industrial and Applied Mathematics, 2009.