AN EXPOSITION OF BRETAGNOLLE AND MASSART'S PROOF OF THE KMT THEOREM FOR THE UNIFORM EMPIRICAL PROCESS

R. M. Dudley

February 23, 2005

Preface

These lecture notes, part of a course given in Aarhus, August 1999, treat the classical empirical process defined in terms of empirical distribution functions. A proof, expanding on one in a 1989 paper by Bretagnolle and Massart, is given for the Komlós-Major-Tusnády result on the speed of convergence of the empirical process to a Brownian bridge in the supremum norm.

Herein "$A := B$" means A is defined by B, whereas "$A =: B$" means B is defined by A.

Richard Dudley
Cambridge, Mass., August 24, 1999

Contents

Chapter 1

Empirical distribution functions: the KMT theorem

1.1 Introduction

Let $U[0,1]$ be the uniform distribution on $[0,1]$ and U its distribution function. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with law U. Let $F_n(t)$ be the empirical distribution function based on $X_1, X_2, \ldots, X_n$,

$$F_n(t) \ := \ \frac{1}{n} \sum_{j=1}^{n} 1_{\{X_j \le t\}},$$

and $\alpha_n(t)$ the corresponding empirical process, i.e., $\alpha_n(t) = \sqrt{n}(F_n(t) - t), \ t \in [0,1]$. Here α_n may be called the *classical* empirical process. Recall that a *Brownian bridge* is a Gaussian stochastic process $B(t), \ 0 \le t \le 1$, with $EB(t) = 0$ and $EB(t)B(u) = t(1-u)$ for $0 \le t \le u \le 1$. Donsker (1952) proved (neglecting measurability problems) that $\alpha_n(t)$ converges in law to a Brownian bridge $B(t)$ with respect to the sup norm. Komlós, Major, and Tusnády (1975) stated a sharp rate of convergence, namely that on some probability space there exist X_i i.i.d. $U[0,1]$ and Brownian bridges B_n such that

$$P\left(\sup_{0 \le t \le 1} |\sqrt{n}(\alpha_n(t) - B_n(t))| > x + c \log n \right) \ < \ K e^{-\lambda x} \tag{1.1}$$

for all n and x, where c, K, and λ are positive absolute constants. Komlós, Major and Tusnády (KMT) formulated a construction giving a joint distribution of α_n and B_n, and this construction has been accepted by later workers. But Komlós, Major and Tusnády gave hardly any proof for (1.1). Csörgő and Révész (1981) sketched a method of proof of (1.1) based on lemmas of G. Tusnády, Lemmas 1.2 and 1.4 below. The implication from Lemma 1.4 to 1.2 is not difficult, but Csörgő and Révész did not include a proof of Lemma 1.4. Bretagnolle and Massart (1989) gave a proof of the lemmas and of the inequality (1.1) with specific constants, Theorem 1.1 below. Bretagnolle and Massart's proof was rather compressed and some readers have had difficulty following it. Csörgő and Horváth (1993), pp. 116-139, expanded the proof while making it more elementary and gave a proof of Lemma 1.4 for $n \ge n_0$ where n_0 is at least 100. The purpose of the present chapter is to give a detailed and in some minor details corrected version of the original Bretagnolle and Massart proof of the lemmas for all n, overlapping in part with the Csörgő and Horváth proof, then to prove (1.1) for some constants, as given by Bretagnolle and Massart and largely following their proof.

Mason and van Zwet (1987) gave another proof of the inequality (1.1) and an extended form of it for subintervals $0 \le t \le d/n$ with $1 \le d \le n$ and $\log n$ replaced by $\log d$, without Tusnády's inequalities and without specifying the constants c, K, λ. Some parts of the proof sketched by Mason and van Zwet are given in more detail by Mason (1998).

Acknowledgments. I am very grateful to Evarist Giné, David Mason, Jon Wellner, and Uwe Einmahl for conversations and correspondence on the topic.

1.2 Statements: the theorem and Tusnády's lemmas

The main result of the present chapter is:

Theorem 1.1. (Bretagnolle and Massart) *The approximation (1.1) of the empirical process by the Brownian bridge holds with $c = 12$, $K = 2$ and $\lambda = 1/6$ for $n \ge 2$.*

The rest of this chapter will give a proof of the theorem. In a preprint, Rio (1991, Theorem 5.1) states in place of (1.1)

$$P\left(\sup_{0\le t\le 1} |\sqrt{n}(\alpha_n(t) - B_n(t))| > ax + b\log n + \gamma \log 2\right) \quad < \quad Ke^{-x} \tag{1.2}$$

for $n \ge 8$ where $a = 3.26$, $b = 4.86$, $\gamma = 2.70$, and $K = 1$. This implies that for $n \ge 8$, (1.1) holds with $c = 5.76$, $K = 1$, and $\lambda = 1/3.26$, where all three constants are better than in Theorem 1.1.

Tusnády's lemmas are concerned with approximating symmetric binomial distributions by normal distributions. Let $\mathcal{B}(n, 1/2)$ denote the symmetric binomial distribution for n trials. Thus if B_n has this distribution, B_n is the number of successes in n independent trials with probability 1/2 of success on each trial. For any distribution function F and $0 < t < 1$ let $F^{-1}(t) := \inf\{x : F(x) \ge t\}$. Here is one of Tusnády's lemmas (Lemma 4 of Bretagnolle and Massart (1989)).

Lemma 1.2. *Let Φ be the standard normal distribution function and Y a standard normal random variable. Let Φ_n be the distribution function of $\mathcal{B}(n, 1/2)$ and set $C_n := \Phi_n^{-1}(\Phi(Y)) - n/2$. Then*

$$|C_n| \;\le\; 1 + (\sqrt{n}/2)|Y|, \tag{1.3}$$

$$|C_n - (\sqrt{n}/2)Y| \;\le\; 1 + Y^2/8. \tag{1.4}$$

Recall the following well known and easily checked facts:

Theorem 1.3. *Let X be a real random variable with distribution function F.*
(a) If F is continuous then $F(X)$ has a $U[0, 1]$ distribution.
(b) For any F, if V has a $U[0, 1]$ distribution then $F^{-1}(V)$ has distribution function F.

Thus $\Phi(Y)$ has a $U[0, 1]$ distribution and $\Phi_n^{-1}(\Phi(Y))$ has distribution $\mathcal{B}(n, 1/2)$. Lemma 1.2 will be shown (by a relatively short proof) to follow from:

Lemma 1.4. *Let Y be a standard normal variable and let β_n be a binomial random variable with distribution $\mathcal{B}(n, 1/2)$. Then for any integer j such that $0 \le j \le n$ and $n+j$ is even, we have*

$$P(\beta_n \ge (n+j)/2) \;\ge\; P(\sqrt{n}Y/2 \ge n(1-\sqrt{1-j/n})), \tag{1.5}$$

$$P(\beta_n \ge (n+j)/2) \;\le\; P(\sqrt{n}Y/2 \ge (j-2)/2). \tag{1.6}$$

Remarks. The restriction that $n+j$ be even is not stated in the formulation of the lemma by Bretagnolle and Massart (1989), but $n+j$ is always even in their proof. If (1.6) holds for $n+j$ even it also holds directly for $n+j$ odd, but the same is not clear for (1.5). It turns out that only the case $n+j$ even is needed in the proof of Lemma 1.2, so I chose to restrict the statement to that case.

The following form of Stirling's formula with remainder is used in the proof of Lemma 1.4.

Lemma 1.5. *Let $n! = (n/e)^n\sqrt{2\pi n}A_n$ where $A_n = 1 + \beta_n/(12n)$, which defines A_n and β_n for $n = 1, 2, \cdots$. Then $\beta_n \downarrow 1$ as $n \to \infty$.*

1.3 Stirling's formula: Proof of Lemma 1.5

It can be checked directly that $\beta_1 > \beta_2 > \cdots > \beta_8 > 1$. So it suffices to prove the lemma for $n \ge 8$. We have $A_n = \exp((12n)^{-1} - \theta_n/(360n^3))$ where $0 < \theta_n < 1$, see Whittaker and Watson (1927), p. 252 or Nanjundiah (1959). Then by Taylor's theorem with remainder,

$$A_n \;=\; \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \frac{1}{6(12n)^3}\phi_n e^{1/12n}\right)\exp(-\theta_n/(360n^3))$$

where $0 < \phi_n < 1$. Next,

$$\beta_{n+1} \;\le\; 12(n+1)\left[\exp\left(\frac{1}{12(n+1)}\right) - 1\right]$$

$$\le\; 1 + \frac{1}{24(n+1)} + \frac{1}{6(12(n+1))^2}e^{1/(12(n+1))},$$

from which $\limsup_{n\to\infty}\beta_n \le 1$, and

$$\beta_n \;=\; 12n[A_n - 1] \;\ge\; 12n\left[\left(1 + \frac{1}{12n} + \frac{1}{288n^2}\right)\exp(-1/(360n^3)) - 1\right].$$

Using $e^{-x} \ge 1 - x$ gives

$$\beta_n \;\ge\; 12n\left[\frac{1}{12n} + \frac{1}{288n^2} - \frac{1}{360n^3}\left(1 + \frac{1}{12n} + \frac{1}{288n^2}\right)\right]$$

$$=\; 1 + \frac{1}{24n} - \frac{1}{30n^2}\left(1 + \frac{1}{12n} + \frac{1}{288n^2}\right).$$

Thus $\liminf_{n\to\infty}\beta_n \ge 1$ and $\beta_n \to 1$ as $n \to \infty$. To prove $\beta_n \ge \beta_{n+1}$ for $n \ge 8$ it will suffice to show that

$$1 + \frac{1}{24(n+1)} + \frac{e^{1/108}}{6\cdot 144n^2} \;\le\; 1 + \frac{1}{24n} - \frac{1}{30n^2}\left[1 + \frac{1}{96} + \frac{1}{288\cdot 8^2}\right]$$

or

$$\frac{e^{1/108}}{6\cdot 144n^2}+\frac{1}{30n^2}\left[1+\frac{1}{96}+\frac{1}{288\cdot 64}\right] \;\le\; \frac{1}{24n(n+1)}$$

or that $0.035/n^2 \le 1/[24n(n+1)]$ or $0.84 \le 1-1/(n+1)$, which holds for $n \ge 8$, proving that β_n decreases with n. Since its limit is 1, Lemma 1.5 is proved. □

1.4 Proof of Lemma 1.4

First, (1.5) will be proved. For any $i = 0, 1, \cdots, n$ such that $n+i$ is even, let $k \;:=\; (n+i)/2$ so that k is an integer, $n/2 \le k \le n$, and $i = 2k-n$. Let $p_{ni} \;:=\; P(\beta_n = (n+i)/2) = P(\beta_n = k) = \binom{n}{k}/2^n$ and $x_i \;:=\; i/n$. Define $p_{ni} \;:=\; 0$ for $n+i$ odd. The factorials in $\binom{n}{k}$ will be approximated via Stirling's formula with correction terms as in Lemma 1.5. To that end, let

$$CS(u,v,w,x,n) \;:=\; \frac{1+u/(12n)}{(1+v/[6n(1-x)])(1+w/[6n(1+x)])}.$$

By Lemma 1.5, we can write for $0 \le i < n$ and $n+i$ even

$$p_{ni} \;=\; CS(x_i,n)\sqrt{2/\pi n}\,\exp(-ng(x_i)/2 \;-\; (1/2)\log(1-x_i^2)) \tag{1.7}$$

where $g(x) \;:=\; (1+x)\log(1+x)+(1-x)\log(1-x)$ and $CS(x_i,n) \;:=\; CS(\beta_n,\beta_{n-k},\beta_k,x_i,n)$. By Lemma 1.5 and since $k \ge n/2$,

$$1^+ \;:=\; 1.013251 \ge 12(e(2\pi)^{-1/2}-1) = \beta_1 \ge \beta_{n-k} \ge \beta_k \ge \beta_n > 1.$$

Thus, for $x \;:=\; x_i$, by clear or easily checked monotonicity properties,

$$CS(x,n) \;\le\; CS(\beta_n,\beta_k,\beta_k,x,n) \;=$$

$$\left(1+\frac{\beta_n}{12n}\right)\left[1+\frac{\beta_k}{3n(1-x^2)}+\frac{\beta_k^2}{36n^2(1-x^2)}\right]^{-1}$$

$$\le\; CS(\beta_n,\beta_k,\beta_k,0,n) \;\le\; CS(\beta_n,\beta_n,\beta_n,0,n)$$

$$\le\; CS(1,1,1,0,n) \;=\; \left(1+\frac{1}{12n}\right)\left[1+\frac{1}{3n}+\frac{1}{36n^2}\right]^{-1}.$$

It will be shown next that $\log(1+y) - 2\log(1+2y) \le -3y + 7y^2/2$ for $y \ge 0$. Both sides vanish for $y = 0$. Differentiating and clearing fractions, we get a clearly true inequality. Setting $y \;:=\; 1/(12n)$ then gives

$$\log CS(x_i,n) \;\le\; -1/(4n) \;+\; 7/(288n^2). \tag{1.8}$$

To get a lower bound for $CS(x,n)$ we have by an analogous string of inequalities

$$CS(x,n) \;\ge\; \left(1+\frac{1}{12n}\right)\left\{1+\frac{1^+}{3n(1-x^2)}+\frac{(1^+)^2}{36n^2(1-x^2)}\right\}^{-1}. \tag{1.9}$$

The inequality (1.5) to be proved can be written as

$$\sum_{i=j}^{n} p_{ni} \ \geq\ 1 - \Phi(2\sqrt{n}(1 - \sqrt{1 - j/n})). \tag{1.10}$$

When $j = 0$ the result is clear. When $n \leq 4$ and $j = n$ or $n - 2$ the result can be checked from tables of the normal distribution. Thus we can assume from here on

$$n \geq 5. \tag{1.11}$$

CASE I. Let $j^2 \geq 2n$, in other words $x_j \geq \sqrt{2/n}$. Recall that for $t > 0$ we have $P(Y > t) \leq (t\sqrt{2\pi})^{-1} \exp(-t^2/2)$, e.g. Dudley (1993), Lemma 12.1.6(a). Then (1.10) follows easily when $j = n$ and $n \geq 5$. To prove it for $j = n - 2$ it is enough to show

$$n(2 - \log 2) - 4\sqrt{2n} + \log(n + 1) + 4 + \log[2\sqrt{2\pi}(\sqrt{n} - \sqrt{2})] \ \geq\ 0, \quad n \geq 5.$$

The left side is increasing in n for $n \geq 5$ and is ≥ 0 at $n = 5$.

For $5 \leq n \leq 7$ we have $(n - 4)^2 < 2n$, so we can assume in the present case that $2n \leq j^2 \leq (n - 4)^2$ and $n \geq 8$. Let $y_i \ :=\ 2\sqrt{n}(1 - \sqrt{1 - i/n})$. Then it will suffice to show

$$p_{ni} \ \geq\ \int_{y_i}^{y_{i+2}} \phi(u)du, \quad i = j, j + 2, \cdots, n - 4, \tag{1.12}$$

where ϕ is the standard normal density function. Let

$$f_n(x) \ :=\ \sqrt{n/2\pi(1 - x)} \exp(-2n(1 - \sqrt{1 - x})^2). \tag{1.13}$$

By the change of variables $u = 2\sqrt{n}(1 - \sqrt{1 - x})$, (1.12) becomes

$$p_{ni} \ \geq\ \int_{x_i}^{x_{i+2}} f_n(x)dx. \tag{1.14}$$

Clearly $f_n > 0$. To see that $f_n(x)$ is decreasing in x for $\sqrt{2/n} \leq x \leq 1 - 4/n$, note that

$$2(1 - x)f_n'/f_n \ =\ 1 - 4n[\sqrt{1 - x} - 1 + x],$$

so f_n is decreasing where $\sqrt{1 - x} - (1 - x) > 1/(4n)$. We have $\sqrt{y} - y \geq y$ for $y \leq 1/4$, so $\sqrt{y} - y > 1/(4n)$ for $1/(4n) < y \leq 1/4$. Let $y \ :=\ 1 - x$. Also $\sqrt{1 - x} - (1 - x) > x/4$ for $x < 8/9$, so $\sqrt{1 - x} - (1 - x) > 1/(4n)$ for $1/n < x < 8/9$. Thus $\sqrt{1 - x} - (1 - x) > 1/(4n)$ for $1/n < x < 1 - 1/(4n)$, which includes the desired range. Thus to prove (1.14) it will be enough to show that

$$p_{ni} \ \geq\ (2/n)f_n(x_i), \quad i = j, j + 2, \cdots, n - 4. \tag{1.15}$$

So by (1.7) it will be enough to show that for $\sqrt{2/n} \leq x \leq 1 \ - \ 4/n$ and $n \geq 8$,

$$CS(x, n)(1 + x)^{-1/2} \exp[n\{4(1 - \sqrt{1 - x})^2 - g(x)\}/2] \ \geq\ 1. \tag{1.16}$$

Let

$$J(x) \ :=\ 4(1 - \sqrt{1 - x})^2 - g(x). \tag{1.17}$$

Then J is increasing for $0 < x < 1$, since its first and second derivatives are both 0 at 0, while its third derivative is easily checked to be positive on $(0, 1)$. In light of (1.9), to prove (1.16) it suffices to show that

$$\left(1 + \frac{1}{12n}\right) e^{nJ(x)/2} \;\geq\; \sqrt{1+x}\left(1 + \frac{1^+}{3n(1-x^2)} + \frac{(1^+)^2}{36n^2(1-x^2)}\right). \tag{1.18}$$

When $x \leq 1 - 4/n$ and $n \geq 8$ the right side is less than 1.5, using first $\sqrt{1+x} \leq \sqrt{2}$, next $x \leq 1 - 4/n$, and lastly $n \geq 8$. For $x \geq 0.55$ and $n \geq 8$ the left side is larger than 1.57, so (1.18) is proved for $x \geq 0.55$. We will next need the inequality

$$J(x) \;\geq\; x^3/2 + 7x^4/48, \quad 0 \leq x \leq 0.55. \tag{1.19}$$

To check this one can calculate $J(0) = J'(0) = J''(0) = 0$, $J^{(3)}(0) = 3$, $J^{(4)}(0) = 7/2$, so that the right side of (1.19) is the Taylor series of J around 0 through fourth order. One then shows straightforwardly that $J^{(5)}(x) > 0$ for $0 \leq x < 1$.

It follows since $nx^2 \geq 2$ and $n \geq 8$ that $nJ(x)/2 \geq x/2 + 7/24n$. Let $K(x) := \exp(x/2)/\sqrt{1+x}$ and $\kappa(x) := (K(x) - 1)/x^2$. We will next see that $\kappa(\cdot)$ is decreasing on $[0, 1]$. To show $\kappa' \leq 0$ is equivalent to $e^{x/2}[4 + 4x - x^2] \geq 4(1 + x)^{3/2}$, which is true at $x = 0$. Differentiating, we would like to show $e^{x/2}[6 - x^2/2] \geq 6\sqrt{1+x}$, or squaring that and multiplying by 4, $e^x(144 - 24x^2 + x^4) \geq 144(1 + x)$. This is true at $x = 0$. Differentiating, we would like to prove $e^x(144 - 48x - 24x^2 + 4x^3 + x^4) \geq 144$. Using $e^x \geq 1 + x$ and algebra gives this result for $0 \leq x \leq 1$.

It follows that $K(x) \geq 1 + 0.3799/n$ when $\sqrt{2/n} \leq x \leq 0.55$. It remains to show that for $x \leq 0.55$,

$$\left(1 + \frac{1}{12n}\right)\left(1 + \frac{0.3799}{n}\right) e^{7/(24n)} \;\geq\; 1 + \frac{1^+}{3n(1-x^2)} + \frac{(1^+)^2}{36n^2(1-x^2)}.$$

At $x = 0.55$ the right side is less than $1 + 0.543/n$, so Case I is completed since $0.543 \leq 1/12 + 0.3799 + 7/24$.

CASE II. The remaining case is $j < \sqrt{2n}$. For any integer k, $P(\beta_n \geq k) = 1 - P(\beta_n \leq k-1)$. For $k = (n+j)/2$ we have $k - 1 = (n + j - 2)/2$. If n is odd, then $P(\beta_n \geq n/2) = 1/2 = P(Y \geq 0)$. If n is even, then $P(\beta_n \geq n/2) - p_{n0}/2 = 1/2 = P(Y \geq 0)$. So, since $p_{n0} = 0$ for n odd, (1.5) is equivalent to

$$\frac{1}{2}p_{n0} + \sum_{0<i\leq j-2} p_{ni} \;\leq\; P(0 \leq Y \leq 2\sqrt{n}(1 - \sqrt{1 - j/n})). \tag{1.20}$$

Given $j < \sqrt{2n}$, a family $I_0, I_1, \cdots, I_K$ of adjacent intervals will be defined such that for n odd,

$$p_{ni} \;\leq\; P(\sqrt{n}Y/2 \in I_k) \;\text{ with }\; i = 2k + 1,\; 0 \leq k \leq K \;:=\; (j - 3)/2, \tag{1.21}$$

while for n even,

$$p_{ni} \;\leq\; P(\sqrt{n}Y/2 \in I_k) \;\text{ with }\; i = 2k,\; 1 \leq k \leq K \;:=\; (j - 2)/2, \tag{1.22}$$

and

$$p_{n0}/2 \;\leq\; P(\sqrt{n}Y/2 \in I_0). \tag{1.23}$$

In either case,

$$I_0 \cup I_1 \cup \cdots \cup I_K \subset [0, n(1-\sqrt{1-j/n})]. \tag{1.24}$$

The intervals will be defined by

$$\delta_{k+1} \;:=\; (k+1)/n + k(k+1/2)(k+1)/n^{3/2}, \quad k \geq 0, \tag{1.25}$$

$$\Delta_{k+1} \;:=\; \delta_{k+1} + k + 1/2 = \delta_{k+1} + (i+1)/2, \quad i = 2k, \quad n \text{ even}, \tag{1.26}$$

$$\Delta_{k+1} \;:=\; \delta_{k+1} + k + 1 = \delta_{k+1} + (i+1)/2, \quad i = 2k+1, \quad n \text{ odd}, \tag{1.27}$$

$$I_k \;:=\; [\Delta_k, \Delta_{k+1}] \text{ with } \Delta_0 = 0. \tag{1.28}$$

It will be shown that $I_0, I_1, \cdots, I_K$ defined by (1.25) through (1.28) satisfy (1.21) through (1.24). Recall that $n \geq 5$ (1.11) and $x_i \;:=\; i/n$.

Proof of (1.24). It needs to be shown that $\Delta_{K+1} \leq n(1-\sqrt{1-x_j})$. Since $j < \sqrt{2n}$, we have $K \leq j/2 - 1 < \sqrt{n/2} - 1$ and

$$\delta_{K+1} \;\leq\; (K+1)/n \;+\; K(K+1/2)/(n\sqrt{2}) \;\leq\; x_j/2 + nx_j^2/(4\sqrt{2}).$$

We have $\Delta_{K+1} = nx_j/2 - 1/2 + \delta_{K+1}$. It will be shown next that

$$1 - \sqrt{1-x} \;\geq\; x/2 + x^2/8, \quad 0 \leq x \leq 1. \tag{1.29}$$

The functions and their first derivatives agree at 0 while the second derivative of the left side is clearly larger.

It then remains to prove that

$$1/2 + nx_j^2(1/8 \;-\; 1/4\sqrt{2}) - x_j/2 \;\geq\; 0.$$

This is true since $nx_j^2 \leq 2$ and $x_j \leq (2/8)^{1/2} = 1/2$, so (1.24) is proved.

Proof of (1.21)-(1.23). First it will be proved that

$$p_{ni} \;\leq\; \frac{\sqrt{2}}{\sqrt{\pi n}} \exp\left[-\frac{1}{4n} + \frac{7}{288n^2} - \frac{(n-1)i^2}{2n^2} + \frac{(i/n)^{2n}}{2n(1-i^2/n^2)}\right]. \tag{1.30}$$

In light of (1.7) and (1.8), it is enough to prove, for $x \;:=\; i/n$, that

$$-[ng(x) + \log(1-x^2) - (n-1)x^2]/2 \;\leq\; x^{2n}/2n(1-x^2). \tag{1.31}$$

It is easy to verify that for $0 \leq t < 1$,

$$g(t) \;=\; (1+t)\log(1+t) + (1-t)\log(1-t) \;=\; \sum_{r=1}^{\infty} t^{2r}/r(2r-1).$$

Thus the left side of (1.31) can be expanded as $\sum_{r\geq 2} x^{2r}(1-n/(2r-1))/2r = A + B$ where $A = \sum_{r=2}^{n-1}$ and $B = \sum_{r\geq n}$. We have

$$d^2A/dx^2 = \sum_{2\leq r\leq (n+1)/2} (2r-n-1)(x^{2r-2} - x^{2n-2r})$$

which is ≤ 0 for $0 \leq x \leq 1$. Since $A = dA/dx = 0$ for $x = 0$ we have $A \leq 0$ for $0 \leq x \leq 1$. Then, $2nB \leq x^{2n}/(1-x^2)$, so (1.30) is proved.

We have for $n \geq 5$ and $x \leq (\sqrt{2n}-2)/n$ that $x^{2n}/(1-x^2) < 10^{-3}$, since $n \mapsto (\sqrt{2n}-2)/n$ is decreasing in n for $n \geq 8$ and the statement can be checked for $n = 5, 6, 7, 8$. So (1.30) yields

$$p_{ni} \ \leq \ \sqrt{2/\pi n} \exp[-0.249/n + 7/288n^2 - (n-1)i^2/2n^2]. \tag{1.32}$$

Next we will need:

Lemma 1.6. *For any $0 \leq a < b$ and a standard normal variable Y,*

$$P(Y \in [a,b]) \ \geq \ \sqrt{1/2\pi}(b-a)\exp[-a^2/4 - b^2/4]\phi(a,b) \tag{1.33}$$

where $\phi(a,b) \ := \ [4/(b^2-a^2)]\sinh[(b^2-a^2)/4] \ \geq \ 1.$

Proof. Since the Taylor series of sinh around 0 has all coefficients positive, and $(\sinh u)/u$ is an even function, clearly $\sinh u/u \geq 1$ for any real u. The conclusion of the lemma is equivalent to

$$\frac{a+b}{2}\int_a^b \exp(-u^2/2)du \ \geq \ \exp(-a^2/2) - \exp(-b^2/2). \tag{1.34}$$

Letting $x \ := \ b-a$ and $v \ := \ u-a$ we need to prove

$$\left(a + \frac{x}{2}\right)\int_0^x \exp(-av - v^2/2)dv \ \geq \ 1 - \exp(-ax - x^2/2).$$

This holds for $x = 0$. Taking derivatives of both sides and simplifying, we would like to show

$$\int_0^x \exp(-av - v^2/2)dv \ \geq \ x\exp(-ax - x^2/2).$$

This also holds for $x = 0$, and differentiating both sides leads to a clearly true inequality, so Lemma 1.6 is proved. □

For the intervals I_k, Lemma 1.6 yields

$$P(\sqrt{n}Y/2 \in I_k) \ \geq \ \sqrt{2/\pi n}\phi_k \exp[-(\Delta_{k+1}^2 + \Delta_k^2)/n + \log(\Delta_{k+1} - \Delta_k)] \tag{1.35}$$

where $\phi_k \ := \ \phi(2\Delta_k/\sqrt{n}, 2\Delta_{k+1}/\sqrt{n})$. The aim is to show that the ratio of the bounds (1.35) over (1.32) is at least 1.

First consider the case $k = 0$. If n is even, this means we want to prove (1.23). Using (1.32) and (1.35) and $\phi_0 \geq 1$, it suffices to show that

$$0.249/n - 7/288n^2 - 1/4n - 1/n^2 - 1/n^3 + \log(1 + 2/n) \ \geq \ 0.$$

Since $\log(1+u) \geq u - u^2/2$ for $u \geq 0$ by taking a derivative, it will be enough to show that

$$(E)_n \ := \ 1.999/n - 3/n^2 - 7/288n^2 - 1/n^3 \ \geq \ 0,$$

and it is easily checked that $n(E)_n > 0$ since $n \geq 5$.

If n is odd, then (1.32) applies for $i = 2k+1 = 1$ and we have $\Delta_0 = 0$, $\Delta_1 = \delta_1+1 = 1 + 1/n$ so (1.35) yields

$$P(\sqrt{n}Y/2 \in I_0) \;\geq\; \sqrt{2/\pi n}\exp[-(1+1/n)^2/n + \log(1+1/n)].$$

Using $\log(1+u) \geq u - u^2/2$ again, the desired inequality can be checked since $n \geq 5$. This completes the case $k = 0$.

Now suppose $k \geq 1$. In this case, $i < \sqrt{2n} - 2$ implies $n \geq 10$ for n even and $n \geq 13$ for n odd. Let $s_k := \delta_k + \delta_{k+1}$ and $d_k := \delta_{k+1} - \delta_k$. Then for i as in the definition of Δ_{k+1},

$$\Delta_{k+1} + \Delta_k \;=\; i + s_k, \tag{1.36}$$

$$\Delta_{k+1} - \Delta_k \;=\; 1 + d_k, \tag{1.37}$$

$$s_k \;=\; \frac{2k+1}{n} + \frac{2k^3+k}{n^{3/2}}, \tag{1.38}$$

and

$$d_k \;=\; \frac{1}{n} + \frac{3k^2}{n^{3/2}}. \tag{1.39}$$

From the Taylor series of sinh around 0 one easily sees that $(\sinh u)/u \geq 1 + u^2/6$ for all u. Letting $u := (\Delta_{k+1}^2 - \Delta_k^2)/n \geq i/n$ gives

$$\log\phi_k \;\geq\; \log(1 + i^2/6n^2). \tag{1.40}$$

We have

$$d_k \leq 3/(2\sqrt{n}) \tag{1.41}$$

since $2k \leq \sqrt{2n} - 2$ and $n \geq 10$. Next we have another lemma:

Lemma 1.7. *$\log(1+x) \geq \lambda x$ for $0 \leq x \leq \alpha$ for each of the pairs $(\alpha,\lambda) = (0.207, 0.9)$, $(0.195, 0.913)$, $(0.14, 0.93)$, $(0.04, 0.98)$.*

Proof. Since $x \mapsto \log(1+x)$ is concave, or equivalently we are proving $1 + x \geq e^{\lambda x}$ where the latter function is convex, it suffices to check the inequalities at the endpoints, where they hold. □

Lemma 1.7 and (1.40) then give

$$\log\phi_k \;\geq\; 0.98 i^2/6n^2 \tag{1.42}$$

since $i^2/(6n^2) \leq 1/3n \leq 0.04$, $n \geq 10$. Next,

Lemma 1.8. *We have $\log(\Delta_{k+1} - \Delta_k) \;\geq\; \lambda d_k$ where $\lambda = 0.9$ when n is even and $n \geq 20$, $\lambda = 0.93$ when n is odd and $n \geq 25$, and $\lambda = 0.913$ when $k = 1$ and $n \geq 10$. Only these cases are possible (for $k \geq 1$).*

Proof. If n is even and $k \geq 2$, then $4 \leq i = 2k < \sqrt{2n} - 2$ implies $n \geq 20$. If n is odd and $k \geq 2$, then $5 \leq i = 2k+1 < \sqrt{2n} - 2$ implies $n \geq 25$. So only the given cases are possible.

We have $k \leq k_n := \sqrt{n/2} - 1$ for n even or $k_n := \sqrt{n/2} - 3/2$ for n odd. Let $d(n) := 1/n + 3k_n^2/n^{3/2}$ and $t := 1/\sqrt{n}$. It will be shown that $d(n)$ is decreasing in n, separately for n even and odd. For n even we would like to show that $3t/2+(1-3\sqrt{2})t^2+3t^3$ is increasing for $0 \le t \le 1/\sqrt{20}$ and in fact its derivative is > 0.04. For n odd we would like to show that $3t/2+(1-9/\sqrt{2})t^2+27t^3/4$ is increasing. We find that its derivative has no real roots and so is always positive as desired.

Since $d(\cdot)$ is decreasing for $n \ge 20$, its maximum for n even, $n \ge 20$ is at $n = 20$ and we find it is less than 0.207 so Lemma 1.7 applies to give $\lambda = 0.9$. Similarly for n odd and $n \ge 25$ we have the maximum $d(25) < 0.14$ and Lemma 1.7 applies to give $\lambda = 0.93$.

If $k = 1$ then $n \mapsto n^{-1} + 3/n^{3/2}$ is clearly decreasing. Its value at $n = 10$ is less than 0.195 and Lemma 1.7 applies with $\lambda = 0.913$. So Lemma 1.8 is proved. □

It will next be shown that for $n \ge 10$

$$s_k \le n^{-1} + k/\sqrt{n}. \tag{1.43}$$

By (1.38) this is equivalent to $2/\sqrt{n} + (2k^2+1)/n \le 1$. Since $k \le \sqrt{n/2} - 1$ one can check that (1.43) holds for $n \ge 14$. For $n = 10, 11, 12, 13$ note that k is an integer, in fact $k \le 1$, and (1.43) holds.

After some calculations, letting $s := s_k$ and $d := d_k$ and noting that

$$\Delta_k^2 + \Delta_{k+1}^2 = \frac{1}{2}[(\Delta_{k+1} - \Delta_k)^2 + (\Delta_k + \Delta_{k+1})^2],$$

to show that the ratio of (1.35) to (1.32) is at least 1 is equivalent to showing that

$$-\frac{is}{n} - \frac{d}{n} - \frac{s^2}{2n} - \frac{d^2}{2n} - \frac{1}{2n} - \frac{7}{288n^2} - \frac{i^2}{2n^2} + \frac{0.249}{n} + \log(1+d) + \log\phi_k \ge 0. \tag{1.44}$$

Proof of (1.44). First suppose that n is even and $n \ge 20$ or n is odd and $n \ge 25$. Apply the bound (1.41) for $d^2/2n$, (1.42) for $\log\phi_k$, (1.43) for s and Lemma 1.8 for $\log(1+d)$. Apply the exact value (1.39) of d in the d/n and λd terms. We assemble together terms with factors k^2, k and no factor of k, getting a lower bound A for (1.44) of the form

$$A := \alpha[k^2/n^{3/2}] - 2\beta[k/n^{5/4}] + \gamma[1/n] \tag{1.45}$$

where, if n is even, so $i = 2k$ and $\lambda = 0.9$, we get

$$\alpha = 0.7 - [2.5 - 2(0.98)/3]/\sqrt{n} - 3/n,$$

$$\beta = n^{-3/4} + n^{-5/4}/2,$$

$$\gamma = 0.649 - [17/8 + 7/288]/n - 1/2n^2.$$

Note that for each fixed n, A is $1/n$ times a quadratic in $k/n^{1/4}$. Also, α and γ are increasing in n while β is decreasing. Thus for $n \ge 20$ the supremum of $\beta^2 - \alpha\gamma$ is attained at $n = 20$ where it is < -0.06. So the quadratic has no real roots and since $\alpha > 0$ it is always positive, thus (1.44) holds.

When n is odd, $i = 2k+1$, $\lambda = 0.93$ and $n \ge 25$. We get a lower bound A for (1.44) of the same form (1.45) where now

$$\alpha = 0.79 - [2.5 - 2(0.98)/3]/\sqrt{n} - 3/n,$$

$$\beta = 1/2n^{1/4} + 2(1 - 0.98/6)/n^{3/4} + 1/2n^{5/4},$$
$$\gamma = 0.679 - (3.625 + 7/288 - 0.98/6)/n - 1/2n^2.$$

For the same reasons, the supremum of $\beta^2 - \alpha\gamma$ for $n \geq 25$ is now attained at $n = 25$ and is negative (less than -0.015), so the conclusion (1.44) again holds.

It remains to consider the case $k = 1$ where n is even and $n \geq 10$ or n is odd and $n \geq 13$. Here instead of bounds for s_k and d_k we use the exact values (1.38) and (1.39) for $k = 1$. We still use the bounds (1.42) for $\log \phi_k$ and Lemma 1.8 for $\log(1+d_k)$. When n is even, $i = 2k = 2$, and we obtain a lower bound A' for (1.44) of the form $a_1/n + a_2/n^{3/2} + \cdots$. All terms n^{-2} and beyond have negative coefficients. Applying the inequality $-n^{-(3/2)-\alpha} \geq -n^{-3/2} \cdot 10^{-\alpha}$ for $n \geq 10$ and $\alpha = 1/2, 1, \cdots$, I found a lower bound $A' \geq 0.662/n - 1.115/n^{3/2} > 0$ for $n \geq 10$. The same method for n odd gave $A' \geq 0.662/n - 1.998/n^{3/2} > 0$ for $n \geq 13$. The proof of (1.5) is complete.

Proof of (1.6). For n odd, (1.6) is clear when $j = 1$, so we can assume $j \geq 3$. For n even, (1.6) is clear when $j = 2$. We next consider the case $j = 0$. By symmetry we need to prove that $p_{n0} \leq P(\sqrt{n}|Y|/2 \leq 1)$. This can be checked from a normal table for $n = 2$. For $n \geq 4$ we have $p_{n0} \leq \sqrt{2/\pi n}$ by (1.32). The integral of the standard normal density from $-2/\sqrt{n}$ to $2/\sqrt{n}$ is clearly larger than the length of the interval times the density at the endpoints, namely $2\sqrt{2/\pi n}\exp(-2/n)$. Since $\exp(-2/n) \geq 1/2$ for $n \geq 4$ the proof for n even and $j = 0$ is done.

We are left with the cases $j \geq 3$. For $j = n$, we have $p_{nn} = 2^{-n}$ and can check the conclusion for $n = 3, 4$ from a normal table. Let ϕ be the standard normal density. We have the inequality, for $t > 0$,

$$P(Y \geq t) \;\geq\; \psi(t) \;:=\; \phi(t)[t^{-1} - t^{-3}], \tag{1.46}$$

Feller (1968), p. 175. Feller does not give a proof. For completeness, here is one:

$$\psi(t) \;=\; -\int_t^\infty \psi'(x)dx \;=\; \int_t^\infty \phi(x)(1 - 3x^{-4})dx \;\leq\; P(Y \geq t).$$

To prove (1.6) via (1.46) for $j = n \geq 5$ we need to prove

$$1/2^n \;\leq\; \phi(t_n)t_n^{-1}(1 - t_n^{-2})$$

where $t_n \;:=\; (n-2)/\sqrt{n}$. Clearly $n \mapsto t_n$ is increasing. For $n \geq 5$ we have $1 - t_n^{-2} \geq 4/9$ and $(2\pi)^{-1/2}e^{2-2/n} \cdot 4/9 \geq 0.878$. Thus it suffices to prove

$$n(\log 2 - 0.5) + 0.5\log n - \log(n-2) + \log(0.878) \;\geq\; 0, \quad n \geq 5.$$

This can be checked for $n = 5, 6$ and the left side is increasing in n for $n \geq 6$, so (1.6) for $j = n \geq 5$ follows.

So it will suffice to prove $p_{ni} \leq P(\sqrt{n}Y/2 \in [(i-2)/2, i/2])$ for $j \leq i < n$. From (1.30) and Lemma 1.6, and the bound $\phi_k \geq 1$, it will suffice to prove, for $x \;:=\; i/n$,

$$-\frac{1}{4n} + \frac{7}{288n^2} - \frac{(n-1)x^2}{2} + \frac{x^{2n}}{2n(1-x^2)} \;\leq\; -\frac{n[(x-2/n)^2 + x^2]}{4}$$

where $3/n \leq x \leq 1 - 2/n$. Note that $2n(1-x^2) \geq 4$. Thus it is enough to prove that

$$x - x^2/2 - x^{2n}/4 \;\geq\; 3/4n \;+\; 7/288n^2$$

for $3/n \le x \le 1$ and $n \ge 5$, which holds since the function on the left is concave, and the inequality holds at the endpoints. Thus (1.6) and Lemma 1.4 are proved. □

1.5 Proof of Lemma 1.2

Let $G(x)$ be the distribution function of a normal random variable Z with mean $n/2$ and variance $n/4$ (the same mean and variance as for $\mathcal{B}(n,1/2)$). Let $B(k,n,1/2) := \sum_{0\le i\le k}\binom{n}{i}2^{-n}$. Lemma 1.4 directly implies

$$G(\sqrt{2kn}-n/2) \;\le\; B(k,n,1/2) \;\le\; G(k+1) \quad\text{for}\quad k \le n/2. \tag{1.47}$$

Specifically, letting $k := (n-j)/2$, (1.6) implies

$$B(k,n,1/2) \;\le\; P(Z \ge n-k-1) \;=\; P(k+1 \ge n-Z) \;=\; G(k+1)$$

since $n-Z$ has the same distribution as Z. (1.5) implies

$$B(k,n,1/2) \;\ge\; P\left(\frac{n}{2}-\frac{\sqrt{n}}{2}Y \le -\frac{n}{2}+\sqrt{2kn}\right) \;=\; G(\sqrt{2kn}-n/2).$$

Let

$$\eta \;:=\; \Phi_n^{-1}(G(Z)). \tag{1.48}$$

This definition of η from Z is called a quantile transformation. By Theorem 1.3, $G(Z)$ has a $U[0,1]$ distribution and η a $\mathcal{B}(n,1/2)$ distribution. It will be shown that

$$Z-1 \;\le\; \eta \;\le\; Z+(Z-n/2)^2/2n+1 \quad\text{if}\quad Z \le n/2, \tag{1.49}$$

and

$$Z-(Z-n/2)^2/2n \;-\; 1 \;\le\; \eta \;\le\; Z+1 \quad\text{if}\quad Z \ge n/2. \tag{1.50}$$

Define a sequence of extended real numbers $-\infty = c_{-1} < c_0 < c_1 < \cdots < c_n = +\infty$ by $G(c_k) = B(k,n,1/2)$ Then one can check that $\eta = k$ on the event $A_k := \{\omega : \; c_{k-1} < Z(\omega) \le c_k\}$. By (1.47), $G(c_k) = B(k,n,1/2) \le G(k+1)$ for $k \le n/2$. So, on the set A_k for $k \le n/2$ we have $Z-1 \le c_k - 1 \le k = \eta$. Note that for n even, $n/2 < c_{n/2}$ while for n odd, $n/2 = c_{(n-1)/2}$. So the left side of (1.49) is proved.

If Y is a standard normal random variable with distribution function Φ and density ϕ then $\Phi(x) \le \phi(x)/x$ for $x > 0$, e.g. Dudley (1993), Lemma 12.1.6(a). So we have

$$P(Z \le -n/2) \;=\; P\left(\frac{n}{2}+\frac{\sqrt{n}}{2}Y \le -\frac{n}{2}\right) \;=$$

$$P\left(\frac{\sqrt{n}}{2}Y \le -n\right) \;=\; \Phi(-2\sqrt{n}) \;\le\; \frac{e^{-2n}}{2\sqrt{2\pi n}} \;<\; \frac{1}{2^n}.$$

So $G(-n/2) < G(c_0) = 2^{-n}$ and $-n/2 < c_0$. Thus if $Z \le -n/2$ then $\eta = 0$. Next note that $Z+(Z-n/2)^2/2n = (Z+n/2)^2/2n \ge 0$ always. Thus the right side of (1.49) holds when $Z \le -n/2$ and whenever $\eta = 0$. Now assume that $Z \ge -n/2$. By (1.47), for $1 \le k \le n/2$

$$G((2(k-1)n)^{1/2}-n/2) \;\le\; B(k-1,n,1/2) \;=\; G(c_{k-1}),$$

from which it follows that $(2(k-1)n)^{1/2} - n/2 \ \le\ c_{k-1}$ and

$$k-1 \ \le\ (c_{k-1}+n/2)^2/2n. \tag{1.51}$$

The function $x \mapsto (x+n/2)^2$ is clearly increasing for $x \ge -n/2$ and thus for $x \ge c_0$. Applying (1.51) we get on the set A_k for $1 \le k \le n/2$

$$\eta \ = \ k \ \le \ (Z+n/2)^2/2n \ + \ 1 \ = \ Z + (Z-n/2)^2/2n \ + \ 1.$$

Since $P(Z \le n/2) = 1/2 \le P(\eta \le n/2)$, and η is a non-decreasing function of Z, $Z \le n/2$ implies $\eta \le n/2$. So (1.49) is proved.

It will be shown next that (η, Z) has the same joint distribution as $(n-\eta, n-Z)$. It is clear that η and $n-\eta$ have the same distribution and that Z and $n-Z$ do. We have for each $k = 0, 1, \cdots, n$, $n-\eta = k$ if and only if $\eta = n-k$ if and only if $c_{n-k-1} < Z \le c_{n-k}$. We need to show that this is equivalent to $c_{k-1} \le n-Z < c_k$, in other words $n-c_k < Z \le n-c_{k-1}$. Thus we want to show that $c_{n-k-1} = n-c_k$ for each k. It is easy to check that $G(n-c_k) = P(Z \ge c_k) = 1-G(c_k)$ while $G(c_k) = B(k, n, 1/2)$ and $G(c_{n-k-1}) = B(n-k-1, n, 1/2) = 1-B(k, n, 1/2)$. The statement about joint distributions follows. (1.49) thus implies (1.50).

Some elementary algebra, (1.49) and (1.50) imply

$$|\eta - Z| \ \le \ 1 + (Z-n/2)^2/2n \tag{1.52}$$

and since $Z < n/2$ implies $\eta \le n/2$ and $Z > n/2$ implies $\eta \ge n/2$,

$$|\eta - n/2| \ \le \ 1 + |Z-n/2|. \tag{1.53}$$

Letting $Z = (n+\sqrt{n}Y)/2$ and noting that then $G(Z) \ \equiv \ \Phi(Y)$, (1.48), (1.52), and (1.53) imply Lemma 1.2 with $C_n = \eta - n/2$. □

1.6 Inequalities for the separate processes

We will need facts providing a modulus of continuity for the Brownian bridge and something similar for the empirical process (although it is discontinuous). Let $h(t) \ := \ +\infty$ if $t \le -1$ and

$$h(t) \ := \ (1+t)\log(1+t) - t, \quad t > -1. \tag{1.54}$$

Lemma 1.9. *Let ξ be a binomial random variable with parameters n and p. Then for any $x \ge 0$ and $m \ := \ np$ we have*

$$P(\xi - m \ge x) \ \le \ \inf_{s>0} e^{-sx} E e^{s(\xi-m)} \ = \ \left(\frac{m}{m+x}\right)^{m+x} \left(\frac{n-m}{n-m-x}\right)^{n-m-x}. \tag{1.55}$$

If $p \le 1/2$ then bounds for the right side of (1.55) give

$$P(\xi \ge m+x) \ \le \ \exp\left(-\frac{m}{1-p} h\left(\frac{x}{m}\right)\right) \tag{1.56}$$

and

$$P(\xi \le m-x) \ \le \ \exp(-x^2/[2p(1-p)]). \tag{1.57}$$

Proof. The first inequality in (1.55) is clear. Let $E(k, n, p)$ denote the probability of at least k successes in n independent trials with probability p of success on each trial, and $B(k, n, p)$ the probability of at most k successes. According to Chernoff's inequalities (Chernoff, 1954), we have with $q \ := \ 1 - p$

$$E(k, n, p) \ \leq \ (np/k)^k (nq/(n-k))^{n-k} \ \text{ if } \ k \geq np,$$

and symmetrically

$$B(k, n, p) \ \leq \ (np/k)^k (nq/(n-k))^{n-k} \ \text{ if } \ k \leq np.$$

These inequalities hold for k not necessarily an integer; for this and the equality in (1.55) see also Hoeffding (1963). Then for $p \leq 1/2$, (1.56) is a consequence proved by Bennett (1962), see also Shorack and Wellner (1986, p. 440, (3)), and (1.57) is a consequence proved by Okamoto (1958) and extended by Hoeffding (1963). □

Let F_n be an empirical distribution function for the uniform distribution on $[0, 1]$ and $\alpha_n(t) \ := \ \sqrt{n}(F_n(t) - t)$, $0 \leq t \leq 1$, the corresponding empirical process. The previous lemma extends via martingales to a bound for the empirical process on intervals.

Lemma 1.10. *For any b with $0 < b \leq 1/2$ and $x > 0$,*

$$\begin{aligned} P(\sup_{0 \leq t \leq b} |\alpha_n(t)| > x/\sqrt{n}) \ &\leq \ 2 \exp\left(-\frac{nb}{1-b} h\left(\frac{x(1-b)}{nb} \right) \right) \\ &\leq \ 2 \exp(-nb(1-b)h(x/(nb))). \end{aligned} \tag{1.58}$$

Remark. The bound given by (1.58) is Lemma 2 of Bretagnolle and Massart (1989). Lemma 1.2 of Csörgő and Horváth (1993), p. 116, has instead the bound $2\exp(-nbh(x/(nb)))$. This does not follow from Lemma 1.10, while the converse implication holds by (1.83) below, but I could not follow Csörgő and Horváth's proof of their form.

Proof. From the binomial conditional distributions of multinomial variables we have for $0 \leq s \leq t < 1$

$$\begin{aligned} E(F_n(t)|F_n(u), \ u \leq s) \ &= \ E(F_n(t)|F_n(s)) \\ = \ F_n(s) + \frac{t-s}{1-s}(1 - F_n(s)) \ &= \ \frac{t-s}{1-s} + \frac{1-t}{1-s} F_n(s), \end{aligned}$$

from which it follows directly that

$$E\left(\frac{F_n(t) - t}{1-t} \Big| F_n(u), \ u \leq s \right) \ = \ \frac{F_n(s) - s}{1-s},$$

in other words, the process $(F_n(t) - t)/(1-t)$, $0 \leq t < 1$ is a martingale in t (here n is fixed). Thus, $\alpha_n(t)/(1-t)$, $0 \leq t < 1$, is also a martingale, and for any real s the process $\exp(s\alpha_n(t)/(1-t))$ is a submartingale, e.g. Dudley (1993), 10.3.3(b). Then

$$P(\sup_{0 \leq t \leq b} \alpha_n(t) > x/\sqrt{n}) \ \leq \ P(\sup_{0 \leq t \leq b} \alpha_n(t)/(1-t) > x/\sqrt{n})$$

which for any $s > 0$ equals

$$P\left(\sup_{0\le t\le b} \exp(s\alpha_n(t)/(1-t)) > \exp(sx/\sqrt{n})\right).$$

By Doob's inequality (e.g. Dudley (1993), 10.4.2, for a finite sequence increasing up to a dense set) the latter probability is

$$\le \inf_{s>0} \exp(-sx/\sqrt{n}) E\exp(s\alpha_n(b)/(1-b)) \ \le\ \exp\left(-\frac{nb}{1-b} h\left(\frac{x(1-b)}{nb}\right)\right)$$

by Lemma 1.9, (1.56). In the same way, by (1.57) we get

$$P(\sup_{0\le t\le b}(-\alpha_n(t)) > x/\sqrt{n}) \ \le\ \exp(-x^2(1-b)/(2nb))). \tag{1.59}$$

It is easy to check that $h(u) \le u^2/2$ for $u \ge 0$, so the first inequality in Lemma 1.10 follows. It is easily shown by derivatives that $h(qy) \ge q^2 h(y)$ for $y \ge 0$ and $0 \le q \le 1$. For $q = 1-b$, the bound in (1.58) then follows. □

We next have a corresponding inequality for the Brownian bridge.

Lemma 1.11. *Let $B(t)$, $0 \le t \le 1$, be a Brownian bridge, $0 < b < 1$ and $x > 0$. Let Φ be the standard normal distribution function. Then*

$$P(\sup_{0\le t\le b} B(t) > x) \ =\ 1 - \Phi(x/\sqrt{b(1-b)}) + \exp(-2x^2)\left(1 - \Phi\left(\frac{(1-2b)x}{\sqrt{b(1-b)}}\right)\right). \tag{1.60}$$

If $0 < b \le 1/2$, then for all $x > 0$,

$$P(\sup_{0\le t\le b} B(t) > x) \ \le\ \exp(-x^2/(2b(1-b))). \tag{1.61}$$

Proof. Let $X(t)$, $0 \le t < \infty$ be a Wiener process. For some real α and value of $X(1)$ let $\beta := X(1) - \alpha$. It will be shown that for any real α and y

$$P\{\sup_{0\le t\le 1} X(t) - \alpha t > y | X(1)\} \ =\ 1_{\{\beta>y\}} + \exp(-2y(y-\beta))1_{\{\beta\le y\}}. \tag{1.62}$$

Clearly, if $\beta > y$ then $\sup_{0\le t\le 1} X(t) - \alpha t > y$ (let $t = 1$). Suppose $\beta \le y$. One can apply a reflection argument as in the proof of Dudley (1993), Proposition 12.3.3, where details are given on making such an argument rigorous. Let $X(t) = B(t) + tX(1)$ for $0 \le t \le 1$, where $B(\cdot)$ is a Brownian bridge. We want to find $P(\sup_{0\le t\le 1} B(t) + \beta t > y)$. But this is the same as $P(\sup_{0\le t\le 1} Y(t) > y | Y(1) = \beta)$ for a Wiener process Y. For $\beta \le y$, the probability that $\sup_{0\le t\le 1} Y(t) > y$ and $\beta \le Y(1) \le \beta + dy$ is the same by reflection as $P(2y - \beta \le Y(1) \le 2y - \beta + dy)$. Thus the desired conditional probability, for the standard normal density ϕ, is $\phi(2y-\beta)/\phi(\beta) = \exp(-2y(y-\beta))$ as stated. So (1.62) is proved.

We can write the Brownian bridge B as $W(t) - tW(1)$, $0 \le t \le 1$, for a Wiener process W. Let $W_1(t) := b^{-1/2}W(bt)$, $0 \le t < \infty$. Then W_1 is a Wiener process. Let $\eta := W(1) - W(b)$. Then η has a normal $N(0, 1-b)$ distribution and is independent of $W_1(t)$, $0 \le t \le 1$. Let $\gamma := ((1-b)W_1(1) - \sqrt{b}\eta)\sqrt{b}/x$. We have

$$P(\sup_{0\le t\le b} B(t) > x|\eta, W_1(1)) = P\left(\sup_{0\le t\le 1}(W_1(t) - (bW_1(1) + \sqrt{b}\eta)t) > x/\sqrt{b}|\eta, W_1(1)\right).$$

Now the process $W_1(t) - (bW_1(t) + \sqrt{b}\eta)t$, $0 \le t \le 1$, has the same distribution as a Wiener process $Y(t)$, $0 \le t \le 1$, given that $Y(1) = (1-b)W_1(1) - \sqrt{b}\eta$. Thus by (1.62) with $\alpha = 0$,

$$P(\sup_{0\le t\le b} B(t) > x|\eta, W_1(1)) = 1_{\{\gamma>1\}} + 1_{\{\gamma\le 1\}}\exp(-2x^2(1-\gamma)/b). \tag{1.63}$$

Thus, integrating gives

$$P(\sup_{0\le t\le b} B(t) > x) = P(\gamma > 1) + \exp(-2x^2/b)E\left(\exp(2x^2\gamma/b)1_{\{\gamma\le 1\}}\right).$$

From the definition of γ it has a $N(0, b(1-b)/x^2)$ distribution. Since x is constant, the latter integral with respect to γ can be evaluated by completing the square in the exponent and yields (1.60).

We next need the inequality, for $x \ge 0$,

$$1 - \Phi(x) \le \frac{1}{2}\exp(-x^2/2). \tag{1.64}$$

This is easy to check via the first derivative for $0 \le x \le \sqrt{2/\pi}$. On the other hand we have the inequality $1 - \Phi(x) \le \phi(x)/x$, $x > 0$, e.g. Dudley (1993), 12.1.6(a), which gives the conclusion for $x \ge \sqrt{2/\pi}$.

Applying (1.64) to both terms of (1.60) gives (1.61), so the Lemma is proved. □

1.7 Proof of Theorem 1.1

For the Brownian bridge $B(t)$, $0 \le t \le 1$, it is well known that for any $x > 0$

$$P(\sup_{0\le t\le 1}|B(t)| \ge x) \le 2\exp(-2x^2),$$

e.g. Dudley (1993), Proposition 12.3.3. It follows that

$$P(\sqrt{n}\sup_{0\le t\le 1}|B(t)| \ge u) \le 2\exp(-u/3)$$

for $u \ge n/6$. We also have $|\alpha_1(t)| \le 1$ for all t and

$$P(\sup_{0\le t\le 1}|\alpha_n(t)| \ge x) \le D\exp(-2x^2), \tag{1.65}$$

which is the Dvoretzky-Kiefer-Wolfowitz inequality with a constant D. Massart (1990) proved (1.65) with the sharp constant $D = 2$. Earlier Hu (1985) proved it with $D = 4\sqrt{2}$. $D = 6$ suffices for present purposes. Given D, it follows that for $u \geq n/6$,

$$P(\sqrt{n} \sup_{0\leq t\leq 1} |\alpha_n(t)| \geq u) \ \leq \ D\exp(-u/3).$$

For $x < 6\log 2$, we have $2e^{-x/6} > 1$ so the conclusion of Theorem 1.1 holds. holds. For $x > n/3 - 12\log n$, $u \ := \ (x + 12\log n)/2 > n/6$ so the left side of (1.1) is bounded above by $(2 + D)n^{-2}e^{-x/6}$. We have $(2 + D)n^{-2} \leq 2$ for $n \geq 2$ and $D \leq 6$.

Thus it will be enough to prove Theorem 1.1 when

$$6\log 2 \leq x \leq n/3 \ - \ 12\log n. \tag{1.66}$$

The function $t \mapsto t/3 - 12\log t$ is decreasing for $t < 36$, increasing for $t > 36$. Thus one can check that for (1.66) to be non-vacuous is equivalent to

$$n \ \geq \ 204. \tag{1.67}$$

Let N be the largest integer such that $2^N \leq n$, so that $\nu \ := \ 2^N \leq n < 2\nu$. Let Z be a ν-dimensional normal random variable with independent components, each having mean 0 and variance $\lambda \ := \ n/\nu$. For integers $0 \leq i < m$ let $A(i,m) \ := \ \{i+1, \cdots, m\}$. For any two vectors $a \ := \ (a_1, \cdots, a_\nu)$ and $b \ := \ (b_1, \cdots, b_\nu)$ in $\mathbb{R}^\nu$, we have the usual inner product $(a, b) \ := \ \sum_{i=1}^{\nu} a_i b_i$. For any subset $D \subset A(0, \nu)$ let 1_D be its indicator function as a member of $\mathbb{R}^\nu$. For any integers $j = 0, 1, 2, \cdots$ and $k = 0, 1, \cdots$, let

$$I_{j,k} \ := \ A(2^j k, 2^j(k + 1)), \tag{1.68}$$

let $e_{j,k}$ be the indicator function of $I_{j,k}$ and for $j \geq 1$, let $e'_{j,k} \ := \ e_{j-1,2k} - e_{j,k}/2$. Then one can easily check that the family $\mathcal{E} \ := \ \{e'_{j,k} : \ 1 \leq j \leq N, \ 0 \leq k < 2^{N-j}\} \cup \{e_{N,0}\}$ is an orthogonal basis of $\mathbb{R}^\nu$ with $(e_{N,0}, e_{N,0}) = \nu$ and $(e'_{j,k}, e'_{j,k}) = 2^{j-2}$ for each of the given j, k. Let $W_{j,k} \ := \ (Z, e_{j,k})$ and $W'_{j,k} \ := \ (Z, e'_{j,k})$. Then since the elements of $\mathcal{E}$ are orthogonal it follows that the random variables $W'_{j,k}$ for $1 \leq j \leq N, \ 0 \leq k < 2^{N-j}$ and $W_{N,0}$ are independent normal with

$$EW'_{j,k} \ = \ EW_{N,0} \ = \ 0, \ \ \mathrm{Var}(W'_{j,k}) = \lambda 2^{j-2}, \ \mathrm{Var}(W_{N,0}) = \lambda\nu. \tag{1.69}$$

Recalling the notation of Lemma 1.2, let Φ_n be the distribution function of a binomial $\mathcal{B}(n, 1/2)$ random variable, with inverse Φ_n^{-1}. Now let $G_m(t) \ := \ \Phi_m^{-1}(\Phi(t))$.

We will begin defining the construction that will connect the empirical process with a Brownian bridge. Let

$$U_{N,0} \ := \ n \tag{1.70}$$

and then recursively as j decreases from $j = N$ to $j = 1$,

$$U_{j-1,2k} \ := \ G_{U_{j,k}}((2^{2-j}/\lambda)^{1/2} W'_{j,k}), \quad U_{j-1,2k+1} \ := \ U_{j,k} - U_{j-1,2k}, \tag{1.71}$$

$k = 0, 1, \cdots, 2^{N-j} - 1$. Note that by (1.69), $(2^{2-j}/\lambda)^{1/2} W'_{j,k}$ has a standard normal distribution, so Φ of it has a $U[0, 1]$ distribution. It is easy to verify successively for $j = N, N-1, \cdots, 0$ that the random vector $\{U_{j,k}, \ 0 \leq k < 2^{N-j}\}$ has a multinomial distribution with parameters

$n, 2^{j-N}, \cdots, 2^{j-N}$. Let $X := (U_{0,0}, U_{0,1}, \cdots, U_{0,\nu-1})$. Then the random vector X has a multinomial distribution with parameters $n, 1/\nu, \cdots, 1/\nu$.

The random vector X is equal in distribution to

$$\{n(F_n((k+1)/\nu) - F_n(k/\nu)),\ 0 \le k \le \nu - 1\}, \tag{1.72}$$

while for a Wiener process W, Z is equal in distribution to

$$\{\sqrt{n}(W((k+1)/\nu) - W(k/\nu)),\ 0 \le k \le \nu - 1\}. \tag{1.73}$$

Without loss of generality, we can assume that the above equalities in distribution are actual equalities for some uniform empirical distribution functions F_n and Wiener process $W = W_n$. Specifically, consider a vector of i.i.d. uniform random variables $(x_1, \cdots, x_n) \in \mathbb{R}^n$ such that

$$F_n(t) := \frac{1}{n}\sum_{j=1}^{n} 1_{\{x_j \le t\}}$$

and note that W has sample paths in $C[0,1]$. Both $\mathbb{R}^n$ and $C[0,1]$ are separable Banach spaces. Thus one can let $(x_1, \cdots, x_n)$ and W be conditionally independent given the vectors in (1.72) and (1.73) which have the joint distribution of X and Z, by the Vorob'ev-Berkes-Philipp theorem, see Berkes and Philippp (1979), Lemma A1. Then we define a Brownian bridge by $B_n(t) := W_n(t) - tW_n(1)$ and the empirical process $\alpha_n(t) := \sqrt{n}(F_n(t) - t)$, $0 \le t \le 1$. By our choices, we then have

$$\{n(F_n(j/\nu) - j/\nu)\}_{j=0}^{\nu} = \left\{\sum_{i=0}^{j-1}\left(X_i - \frac{n}{\nu}\right)\right\}_{j=0}^{\nu} \tag{1.74}$$

and

$$\{\sqrt{n}B_n(j/\nu)\}_{j=0}^{\nu} = \left\{\left(\sum_{i=0}^{j-1} Z_i\right) - \frac{j}{\nu}\sum_{r=0}^{\nu-1} Z_r\right\}_{j=0}^{\nu}. \tag{1.75}$$

Theorem 1.1 will be proved for the given B_n and α_n. Specifically, we want to prove

$$P_0 := P\left(\sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| > (x + 12\log n)/\sqrt{n}\right) \le 2\exp(-x/6). \tag{1.76}$$

It will be shown that $\alpha_n(j/\nu)$ and $B_n(j/\nu)$ are not too far apart for $j = 0, 1, \cdots, \nu$ while the increments of the processes over the intervals between the lattice points j/ν are also not too large.

Let $C := 0.29$. Let M be the least integer such that

$$C(x + 6\log n) \le \lambda 2^{M+1}. \tag{1.77}$$

Since $n \ge 204$ (1.67) and $\lambda < 2$ this implies $M \ge 2$. We have by definition of M and (1.66)

$$2^M \le \lambda 2^M \le C(x + 6\log n) \le Cn/3 < 0.1 \cdot 2^{N+1} < 2^{N-2}$$

so $M \le N - 3$.

For each $t \in [0,1]$, let $\pi_M(t)$ be the nearest point of the grid $\{i/2^{N-M},\ 0 \le i \le 2^{N-M}\}$, or if there are two nearest points, take the smaller one. Let $D := X - Z$ and $D(m) := \sum_{i=1}^m D_i$. Let $C' := 0.855$ and define

$$\Theta := \{U_{j,k} \le \lambda(1+C')2^j \ \text{ whenever } \ M+1 < j \le N,\ 0 \le k < 2^{N-j}\}$$
$$\cap \{U_{j,k} \ge \lambda(1-C')2^j \ \text{ whenever } \ M < j \le N,\ 0 \le k < 2^{N-j}\}.$$

Then

$$P_0 \le P_1 + P_2 + P_3 + P(\Theta^c)$$

where

$$P_1 := P\left(\sup_{0\le t\le 1} |\alpha_n(t) - \alpha_n(\pi_M(t))| > 0.28(x + 6\log n)/\sqrt{n}\right), \tag{1.78}$$

$$P_2 := P\left(\sup_{0\le t\le 1} |B_n(t) - B_n(\pi_M(t))| > 0.22(x + 6\log n)/\sqrt{n}\right), \tag{1.79}$$

and, recalling (1.74) and (1.75),

$$P_3 := 2^{N-M} \max_{m\in A(M)} P\left\{\left(|D(m) - \frac{m}{\nu}D(\nu)| > 0.5x + 9\log n\right) \cap \Theta\right\}, \tag{1.80}$$

where $A(M) := \{k2^M : k = 1, 2, \cdots\} \cap A(0,\nu)$.

First we bound $P(\Theta^c)$. Since by (1.71) $U_{j,k} = U_{j-1,2k} + U_{j-1,2k+1}$, we have

$$\Theta^c \subset \bigcup_{0\le k<2^{N-M-2}} \{U_{M+2,k} > (1+C')\lambda 2^{M+2}\} \cup \bigcup_{0\le k<2^{N-M-1}} \{U_{M+1,k} < (1-C')\lambda 2^{M+1}\}.$$

Since $U_{M+2,k}$ and $U_{M+1,k}$ are binomial random variables, Lemma 1.9 gives

$$P(\Theta^c) \le 2^{N-M-1}\left(\exp(-\lambda 2^{M+2}h(C')) + \exp(-\lambda 2^{M+1}h(-C'))\right).$$

Now $2h(C') \ge 0.5823 \ge h(-C') \ge 0.575$ (note that C' has been chosen to make $2h(C')$ and $h(-C')$ approximately equal). By definition of M (1.77), $\lambda 2^{M+1} \ge C(x + 6\log n)$, and $0.575C > 1/6$, so

$$P(\Theta^c) \le 2^{-M}\exp(-x/6). \tag{1.81}$$

Next, to bound P_1 *and* P_2. Let $b := 2^{M-N-1} \le 1/2$. Since $\alpha_n(t)$ has stationary increments, we can apply Lemma 1.10. Let $u := x + 6\log n$. We have by definition of M (1.77)

$$nb = n2^{M-N-1} < Cu/2. \tag{1.82}$$

By (1.66), $u < n/3$ so $b < C/6$. Recalling (1.54), note that $h'(t) \equiv \log(1+t)$. Thus h is increasing. For any given $v > 0$ it is easy to check that

$$y \mapsto yh(v/y) \ \text{ is decreasing for } \ y > 0. \tag{1.83}$$

Lemma 1.10 gives

$$P_1 \le 2^{N-M+2}\exp\left(-nb(1-b)h\left(\frac{0.28u}{nb}\right)\right)$$

$$< \ 2^{N-M+2}\exp\left(-\frac{C}{2}\left[1-\frac{C}{6}\right]uh\left(0.28\cdot\frac{2}{C}\right)\right)$$

by (1.83) and (1.82) and since $1-b > 1-C/6$, so one can calculate

$$P_1 \ \leq \ 2^{N-M+2}e^{-u/6} \ \leq \ 2^{2-M}\lambda^{-1}\exp(-x/6). \tag{1.84}$$

The Brownian bridge also has stationary increments, so Lemma 1.11, (1.61) and (1.82) give

$$P_2 \ \leq \ 2^{N-M+2}\exp(-(0.22u)^2/(2nb))$$
$$\leq \ 2^{N-M+2}\exp(-(0.22)^2u/C) \ \leq \ 2^{2-M}\lambda^{-1}e^{-x/6} \tag{1.85}$$

since $(0.22)^2/C > 1/6$.

It remains to bound P_3. Fix $m \in A(M)$. A bound is needed for

$$P_3(m) \ := \ P\left\{\left(|D(m)-\frac{m}{\nu}D(\nu)| > 0.5x + 9\log n\right)\cap\Theta\right\}. \tag{1.86}$$

For each $j = 1,\cdots,N$ take $k(j)$ such that $m \in I_{j,k(j)}$. By the definition (1.68) of $I_{j,k}$, $k(M) = m2^{-M}-1$ and $k(j) = [k(j-1)/2]$ for $j = 1,\cdots,N$ where $[x]$ is the largest integer $\leq x$. From here on each double subscript $j,k(j)$ will be abbreviated to the single subscript j, e.g. $e'_j \ := \ e'_{j,k(j)}$. The following orthogonal expansion holds in $\mathcal{E}$:

$$1_{A(0,m)} \ = \ \frac{m}{\nu}e_{N,0} + \sum_{M<j\leq N} c_j e'_j, \tag{1.87}$$

where $0 \leq c_j \leq 1$ for $m < j \leq N$. To see this, note that $1_{A(0,m)} \perp e'_{j,k}$ for $j \leq M$ since 2^M is a divisor of m. Also, $1_{A(0,m)} \perp e'_{j,k}$ for $k \neq k(j)$ since $1_{A(0,m)}$ has all 0's or all 1's on the set where $e'_{j,k}$ has non-zero entries, half of which are $+1/2$ and the other half $-1/2$. In an orthogonal expansion $f = \sum_j c_j f_j$ we always have $c_j = (f,f_j)/\|f_j\|^2$ where $\|v\|^2 \ := \ (v,v)$. We have $\|e'_j\| = 2^{(j-2)/2}$. Now, $(1_{A(0,m)}, e'_j)$ is as large as possible when the components of e'_j equal $= 1/2$ only for indices $\leq m$, and then the inner product equals 2^{j-2}, so $|c_j| \leq 1$ as stated. The m/ν factor is clear.

We next have

$$e_j \ = \ 2^{j-N}e_{N,0} + \sum_{i>j}(-1)^{s(i,j,m)}2^{j+1-i}e'_i \tag{1.88}$$

where $s(i,j,m) = 0$ or 1 for each i,j,m so that the corresponding factors are ± 1, the signs being immaterial in what follows. Let $\Delta_j \ := \ (D,e'_j)$. Then from (1.87),

$$\left|D(m)-\frac{m}{\nu}D(\nu)\right| \ \leq \ \sum_{M<j\leq N}|\Delta_j|. \tag{1.89}$$

Recall that $W'_j = (Z,e'_j)$ (see between (1.68) and (1.69)) and $D = X - Z$. Let $\xi_j \ := \ (2^{2-j}/\lambda)^{1/2}W'_j$ for $M < j \leq N$. Then by (1.69) and the preceding statement, $\xi_{M+1},\cdots,\xi_N$ are i.i.d. standard normal random variables. We have $U_{j,k} = (X,e_{j,k})$ for all j and k from the definitions. Then $U_j = (X,e_j)$. Let $U'_j = (X,e'_j)$. By (1.71) and Lemma 1.2, (1.4),

$$|U'_j \ - \ \sqrt{U_j}\xi_j/2| \ \leq \ 1 + \xi_j^2/8. \tag{1.90}$$

Let

$$L_j \;:=\; |W'_j - \sqrt{U_j}\xi_j/2| \;=\; |\xi_j||\sqrt{U_j} - \sqrt{\lambda 2^j}|/2$$

by definition of ξ_j. Thus

$$|\Delta_j| \;\le\; L_j + 1 + \xi_j^2/8. \tag{1.91}$$

Then we have on Θ

$$|\sqrt{U_j} - \sqrt{\lambda 2^j}| \;=\; |U_j - \lambda 2^j|/(\sqrt{\lambda 2^j} + \sqrt{U_j}) \;\le\; \frac{|U_j - \lambda 2^j|}{\sqrt{\lambda 2^j}} \cdot \frac{1}{1 + \sqrt{1 - C'}},$$

where as before $C' \;:=\; 0.855$. Then by (1.71), (1.88) and (1.3) of Lemma 1.2,

$$\begin{aligned} |U_j - \lambda 2^j| \;&\le\; 2^{j-N}|U_N - n| + 2\sum_{j<i\le N} 2^{j-i}|U'_i| \\ &\le\; 2 + (\lambda(1+C'))^{1/2}\sum_{j<i\le N} 2^{j-i/2}|\xi_i| \end{aligned}$$

on Θ, recalling that by (1.70), $U_N = U_{N,0} = n$. Let $C_2 \;:=\; 1/(1+\sqrt{1-C'})$. It follows that

$$L_j \;\le\; 2^{-j/2}C_2|\xi_j| + \frac{1}{2}C_2\sqrt{1+C'}\sum_{j<i\le N} 2^{(j-i)/2}|\xi_j||\xi_i|. \tag{1.92}$$

Applying the inequality $|\xi_i||\xi_j| \le (\xi_i^2 + \xi_j^2)/2$, we get the bound

$$\sum_{M<j\le N}\sum_{j<i\le N} 2^{(j-i)/2}|\xi_i\xi_j| \;\le\; \sum_{M<j\le N} A_j\xi_j^2 \tag{1.93}$$

where

$$A_j \;:=\; \frac{1}{2}\left(\sum_{M<r<j} 2^{(r-j)/2} + \sum_{j<i\le N} 2^{(j-i)/2}\right).$$

Then

$$\begin{aligned} A_j \;&\le\; \frac{1}{2}\left[\frac{2^{-1/2} - 2^{(M-j)/2}}{1 - 2^{-1/2}} + \frac{2^{-1/2}}{1 - 2^{-1/2}}\right] \\ &\le\; 1 + \sqrt{2} - 2^{(M-j-2)/2}/(1 - 2^{-1/2}). \end{aligned}$$

Let $C_3 \;:=\; C_2(1+\sqrt{2})\sqrt{1+C'}/2 \;\le\; 1.19067$. Then

$$\sum_{M<j\le N} L_j \;\le\; C_3\sum_{M<j\le N}\xi_j^2 + \sum_{M<j\le N} 2^{-j/2}|\xi_j|C_2\left(1 - \frac{\sqrt{1+C'}}{2}2^{(M-2)/2}|\xi_j|/(1 - 2^{-1/2})\right). \tag{1.94}$$

Let

$$C_4 \;:=\; \frac{\sqrt{1+C'}}{4(1 - 2^{-1/2})} \;=\; \frac{\sqrt{2}\sqrt{1+C'}(\sqrt{2}+1)}{4},$$

and for each M let $c_M \;:=\; 1/(4C_4 2^{M/2})$. Then for any real number x, we have $x(1 - C_4 2^{M/2}x) \le c_M$. It follows that

$$\sum_{M<j\le N} L_j \;\le\; \sum_{M<j\le N} C_3\xi_j^2 + c_M C_2 2^{-j/2}$$

$$\leq C_2 c_M 2^{-(M+1)/2}/(1-2^{-1/2}) + \sum_{M<j\leq N} C_3\xi_j^2$$

$$\leq \frac{C_2 2^{-M}}{\sqrt{2}\sqrt{1+C'}} + \sum_{M<j\leq N} C_3\xi_j^2.$$

Thus, combining (1.91) and (1.94) we get on Θ

$$\sum_{M<j\leq N} |\Delta_j| \leq N + \left(\frac{1}{8} + C_3\right) \sum_{M<j\leq N} \xi_j^2. \tag{1.95}$$

We have $E\exp(t\xi^2) = (1-2t)^{-1/2}$ for $t < 1/2$ and any standard normal variable ξ such as ξ_j for each j. Since $\xi_{M+1}, \cdots, \xi_N$ are independent we get

$$E\exp\left(\left(\frac{1}{3}\sum_{M<j\leq N} |\Delta_j|\right) 1_\Theta\right) \leq e^{N/3}\left(1 - \frac{2}{3}\left(C_3 + \frac{1}{8}\right)\right)^{(M-N)/2}$$

$$\leq e^{N/3} 2^{1.513(N-M)} \leq 2^{2N-1.5M}.$$

Markov's inequality and (1.89) then yield

$$P_3(m) \leq e^{-x/6} n^{-3} 2^{2N-1.5M}.$$

Thus

$$P_3 \leq e^{-x/6} n^{-3} 2^{3N-2.5M} \leq 2^{-2.5M} e^{-x/6}. \tag{1.96}$$

Collecting (1.81), (1.84), (1.85) and (1.96) we get that $P_0 \leq (2^{3-M}\lambda^{-1} + 2^{-M} + 2^{-2.5M})e^{-x/6}$. By (1.77) and (1.67) and since $x \geq 6\log 2$ (1.66) and $M \geq 2$, it follows that Theorem 1.1 holds. □

1.8 Another way of defining the KMT construction

Now, here is an alternate description of the KMT construction as given in the previous section. For any Hilbert space H, the *isonormal process* is a stochastic process L indexed by H such that the joint distributions of $L(f)$ for $F \in H$ are normal (Gaussian) with mean 0 and covariance given by the inner product in H, $EL(f)L(g) = (f,g)$. Since the inner product is a nonnegative definite bilinear form, such a process exists. Moreover, we have:

Lemma 1.12. *For any Hilbert space H, an isonormal process L on H is linear, that is, for any $f, g \in H$ and constant c, $L(cf+g) = cL(f) + L(g)$ almost surely.*

Proof. The variable $L(cf+g) - cL(f) - L(g)$ clearly has mean 0 and by a short calculation one can show that its variance is also 0, so it is 0 almost surely. □

The Wiener process (Brownian motion) is a Gaussian stochastic process W_t defined for $t \geq 0$ with mean 0 and covariance $EW_sW_t = \min(s,t)$. One can obtain a Wiener process easily from an isonormal process as follows. Let H be the Hilbert space $L^2([0,\infty),\lambda)$ where λ is Lebesgue measure. Let $W_t := L(1_{[0,t]})$. This process is Gaussian, has mean 0 and clearly has the correct covariance. Historically, the Wiener process was defined first, and then $L(f)$ was defined only for the particular Hilbert space $L^2([0,\infty))$ by way of a "stochastic integral" $L(f) = \int_0^\infty f(t)dW_t$, which generally doesn't exist as an ordinary integral but is defined as a limit in probability, approximating f in L^2 by step functions. Defining L first seems much easier.

The Brownian bridge process, as has been treated throughout this chapter, is a Gaussian stochastic process B_t defined for $0 \le t \le 1$ with mean 0 and covariance $EB_tB_u = t(1-u)$ for $0 \le t \le u \le 1$. Given a Wiener process W_t, it is easy to see that $B_t = W_t - tW_1$ for $0 \le t \le 1$ defines a Brownian bridge.

For $j = 0, 1, 2, ...,$ and $k = 1, ..., 2^j$ let $I_{j,k}$ be the open interval $((k-1)/2^j, k/2^j)$. Let $T_{j,k}$ be the "triangle function" defined as 0 outside $I_{j,k}$, 1 at the midpoint $(2k-1)/2^{j+1}$, and linear in between. For a function $f : [0,1] \mapsto \mathbb{R}$ and $r = 0, 1, ...,$ let $[f]_r := f$ at $k/2^r$ for $k = 0, 1, ..., 2^r$ and linear in between. Let

$$f_{j,k} \;:=\; W_{j,k}(f) \;:=\; f\left(\frac{2k-1}{2^{j+1}}\right) - \frac{1}{2}\left[f\left(\frac{k-1}{2^j}\right) + f\left(\frac{k}{2^j}\right)\right].$$

Lemma 1.13. *If f is affine, that is $f(t) \equiv a + bt$ where a and b are constants, then $f_{j,k} = 0$ for all j and k.*

Proof. One can check this easily if f is a constant or if $f(t) \equiv t$, then use linearity of the operation $W_{j,k}$ on functions for each j and k. □

Lemma 1.14. *For any $f : [0,1] \mapsto \mathbb{R}$ and $r = 0, 1, ...,$ for $0 \le t \le 1$*

$$[f]_r(t) \;=\; f(0) + t[f(1) - f(0)] + \sum_{j=0}^{r-1}\sum_{k=1}^{2^j} f_{j,k}T_{j,k}(t),$$

where the sum is defined as 0 for $r = 0$.

Proof. For $r = 0$ we have $f(0) + t[f(1) - f(0)] = f(0)$ when $t = 0$, $f(1)$ when $t = 1$, and the function is linear in between, so it equals $[f]_0$. Then by Lemma 1.13 and linearity of the operations $W_{j,k}$ we can assume in the proof for $r \ge 1$ that $f(0) = f(1) = 0$.

For $r = 1$ we have $f_{0,1}T_{0,1}(t) = 0 = f(t)$ for $t = 0$ or 1 and $f(1/2)$ for $t = 1/2$, with linearity in between, so $f_{0,1}T_{0,1} = [f]_1$, proving the case $r = 1$. Then, by induction on r, we can apply the same argument on each interval $I_{r,k}$, $k = 1, ..., 2^r$, to prove the lemma. □

The following is clear since a continuous function on $[0,1]$ is uniformly continuous:

Lemma 1.15. *If f is continuous on $[0,1]$ then $[f]_r$ converges to f uniformly as $r \to \infty$.*

It follows that for any $f \in C[0,1]$,

$$f(t) \;=\; f(0) + t[f(1) - f(0)] + \sum_{j=0}^{\infty}\sum_{k=1}^{2^j} f_{j,k}T_{j,k}(t),$$

where the sum converges uniformly on $[0,1]$. Thus, the sequence of functions

$$1, t, T_{0,1}, T_{1,1}, T_{1,2}, ..., T_{j,1}, ..., T_{j,2^j}, T_{j+1,1}, ...,$$

is known as the *Schauder basis* of $C[0,1]$. This basis fits well with a simple relation between the Brownian motion or Wiener process $W_t,\ t \geq 0$, and the Brownian bridge $B_t,\ 0 \leq t \leq 1$, given by $B_t = W_t - tW_1,\ 0 \leq t \leq 1$. Both processes are 0 at 0, and their Schauder expansions differ only in the linear "t" term where $W_.$ has the coefficient W_1 and $B_.$ has the coefficient 0, by the following fact:

Lemma 1.16. $W_{j,k}(B_.) = W_{j,k}(W_.)$ *for all* $j = 0, 1, \ldots$ *and* $k = 1, \ldots, 2^j$.

Proof. We need only note that $W_{j,k}(\cdot)$ is a linear operation on functions for each j and k and $W_{j,k}(tW_1) = 0$ by Lemma 1.13. □

Lemma 1.17. *The random variables* $W_{j,k}(B_.)$ *for* $j = 0, 1, \ldots$ *and* $k = 1, \ldots, 2^j$ *are independent with distribution* $N(0, 2^{-j-2})$.

Proof. We have by the previous lemma

$$\begin{aligned} W_{j,k}(B_.) &= W_{j,k}(W_.) = W_{(2k-1)/2^{j+1}} - \frac{1}{2}\left[W_{(k-1)/2^j} + W_{k/2^j}\right] \\ &= L(1_{[0,(2k-1)/2^{j+1}]}) - \frac{1}{2}\left[L(1_{[0,(k-1)/2^j]}) + L(1_{[0,k/2^j]})\right] \end{aligned}$$

which by linearity of the isonormal process L, Lemma 1.12, equals $L(g_{j,k})$ where

$$\begin{aligned} g_{j,k} &:= 1_{[0,(2k-1)/2^{j+1}]} - \frac{1}{2}\left[1_{[0,(k-1)/2^j]} + 1_{[0,k/2^j]}\right] \\ &= \frac{1}{2}\left[1_{((k-1)/2^j,(2k-1)/2^{j+1}]} - 1_{((2k-1)/2^{j+1},k/2^j]}\right]. \end{aligned}$$

(These functions $g_{j,k}$, multiplied by some constants, are known as Haar functions.) To finish the proof of Lemma 1.17 we will use the following:

Lemma 1.18. *The functions* $g_{j,k}$ *and* $g_{j',k'}$ *are orthogonal in* $L^2([0,1])$ *(with Lebesgue measure) unless* $(j,k) = (j',k')$.

Proof. If $j = j'$, the functions $g_{j,k}$ are orthogonal for different k since they are supported on non-overlapping intervals $I_{j,k}$. If $j \neq j'$, say $j' < j$, then $g_{j,k}$ is 0 outside of $I_{j,k}$, equal to $1/2$ on the left half of it and $-1/2$ on the right half, while $g_{j',k'}$ is constant on the interval, so the functions are orthogonal, proving Lemma 1.18. □

Returning to the proof of Lemma 1.17, we have that L of orthogonal functions are independent normal variables with mean 0, and $E(L(f)^2) = \|f\|^2$, where

$$\|g_{j,k}\|^2 = \int_0^1 g_{j,k}(t)^2 dt = 1/2^{j+2}$$

since $g_{j,k}^2$ equals $1/4$ on an interval of length $1/2^j$ and is 0 elsewhere. So Lemma 1.17 is proved. □

There are other ways of expanding functions on $[0,1]$ beside Schauder bases, for example, Fourier series. Fourier series have the advantage that the terms in the series are orthogonal functions with respect to Lebesgue measure on $[0,1]$. The Schauder basis functions are not orthogonal, for example the constant function 1 is not orthogonal to any of the other functions in the sequence, and the functions are all nonnegative, so those whose supports overlap are non-orthogonal. However, the Schauder functions are indefinite integrals of constant multiples of the orthogonal functions $g_{j,k}$ or equivalently constant multiples of Haar functions, and it turns out that the indefinite integral fits well with the processes we are considering, as in the above proof. In a sense, the Wiener process W_t is the indefinite integral of the isonormal process L via $W_t = L(1_{[0,t]})$.

Let Φ_m be the distribution function of the binomial $\text{bin}(m, 1/2)$ distribution, $\Phi_m(x) := 0$ for $x < 0$, $\Phi_m(x) := \sum_{j=0}^{k} \binom{m}{k} 2^{-m}$ for $k \leq x < k+1$, $k = 0, 1, ..., m-1$, and $\Phi_m(x) := 1$ for $x \geq m$. For a function F from $\mathbb{R}$ into itself let $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$, as in Lemma 1.3. Let $H(t|m) := \Phi_m^{\leftarrow}(t)$ for $0 < t < 1$.

Now to proceed with the KMT construction, for a given n, let $B^{(n)}$ be a Brownian bridge process. Let $V_{0,1} := n$. Let $V_{1,1} := H(\Phi(2W_{0,1}(B^{(n)}))|n)$, $V_{1,2} := V_{0,1} - V_{1,1}$. By Lemma 1.17, $2W_{0,1}(B^{(n)})$ has law $N(0,1)$, thus Φ of it has law $U[0,1]$ by Lemma 1.3(a), and $V_{1,1}$ has law $\text{bin}(n, 1/2)$ by Lemma 1.3(b). We will define empirical distribution functions U_n for the $U[0,1]$ distribution recursively over dyadic rationals, beginning with $U_n(0) = 0$, $U_n(1) = 1$, and $U_n(1/2) = V_{1,1}/n$. These values have their correct distributions so far. Now given $V_{j-1,k}$ for some $j \geq 2$ and all $k = 1, ..., 2^{j-1}$, let

$$V_{j,2k-1} := H(\Phi(2^{(j+1)/2} W_{j-1,k}(B^{(n)})|V_{j-1,k})$$

and $V_{j,2k} := V_{j-1,k} - V_{j,2k-1}$. This completes the recursive definition of the $V_{j,i}$. Then $W_{j-1,k}(B^{(n)})$ has law $N(0, 2^{-j-1})$ by Lemma 1.17, so $2^{(j+1)/2}$ times it has law $N(0,1)$, and Φ of the product has law $U[0,1]$ by Lemma 1.3(a), so $V_{j,2k-1}$ has law $\text{bin}(V_{j-1,k}, 1/2)$ by Lemma 1.3(b). Let $U_n(1/4) := V_{2,1}/n$, $U_n(3/4) := U_n(1/2) + V_{2,2}/n$, and so on. Then $U_n(k/2^j)$ for $k = 0, 1, ..., 2^j$ have their correct joint distribution and and when taken for all $j = 1, 2, ...$, they uniquely define U_n on $[0,1]$ by monotonicity and right-continuity, which has all the properties of an empirical distribution function for $U[0,1]$.

With the help of Lemma 1.2, one can show that the Schauder coefficients of the empirical process $\alpha_n := n^{1/2}(U_n - U)$, where U is the $U[0,1]$ distribution function, are close to those of $B^{(n)}$. Lemma 1.2 has to be applied not only for the given n but also for n replaced by $V_{j,k}$, and that creates some technical problems. For the present, the proof in the previous section is not rewritten here in terms of the present construction.

REFERENCES

Bennett, George W. (1962). Probability inequalities for the sum of bounded random variables. *J. Amer. Statist. Assoc.* **57**, 33–45.

Berkes, I., and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7**, 29-54.

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

Chernoff, H. (1952). A measure of efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23**, 493-507.

Csörgő, M., and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics.* Wiley, Chichester.

Csörgő, M., and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

Donsker, Monroe D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23**, 277–281.

Dudley, Richard M. (1984). *A Course on Empirical Processes.* Ecole d'été de probabilités de St.-Flour, 1982. Lecture Notes in Math. **1097**, 1-142, Springer.

Dudley, R. M. (2002). *Real Analysis and Probability.* Second ed., Cambridge University Press.

Feller, William (1968). *An Introduction to Probability Theory and Its Applications.* Vol. 1, 3d ed. Wiley, New York.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Hu, Inchi (1985). A uniform bound for the tail probability of Kolmogorov-Smirnov statistics. *Ann. Statist.* **13**, 821-826.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 111–131.

Mason, D. M. (1998). Notes on the the KMT Brownian bridge approximation to the uniform empirical process. Preprint.

Mason, D. M., and van Zwet, W. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15**, 871-884.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18**, 1269-1283.

Nanjundiah, T. S. (1959). Note on Stirling's formula. *Amer. Math. Monthly* **66**, 701-703.

Okamoto, Masashi (1958). Some Inequalities Relating to the Partial Sum of Binomial Probabilities. *Ann. Inst. Statist. Math.* **10**, 29-35.

Rio, E. (1991). Local invariance principles and its application to density estimation. Prépubl Math. Univ. Paris-Sud 91-71.

Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98**, 21-45.

Shorack, G., and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Whittaker, E. T., and Watson, G. N. (1927). *Modern Analysis*, 4th ed., Cambridge Univ. Press, Repr. 1962.