The Canadian Journal of Statistics
Vol. 46, No. 4, 2018, Pages 656–672
La revue canadienne de statistique

The empirical identity process: Asymptotics and applications

Enrico BIBBONA[1], Giovanni PISTONE[2], and Mauro GASPARINI[1]

[1]*Dipartimento di Scienze Matematiche "G.L. Lagrange", Politecnico di Torino, 10129 Turin, Italy*
[2]*de Castro Statistics, Collegio Carlo Alberto, 10122 Turin, Italy*

Key words and phrases: Strong approximations; uniform empirical process; uniform quantile process.

MSC 2010: Primary 62G30; secondary 60F17, 62G30

Abstract: When sampling independent observations drawn from the uniform distribution on the unit interval, as the sample size gets large the asymptotic behaviour of both the empirical distribution function and empirical quantile function is well known. In this article we study analogous asymptotic results for the function that is obtained by composing the empirical quantile function with the empirical distribution function. Since the former is the generalized inverse of the latter, the result will approximate the identity function. We define a scaled and centered version of this function—the *empirical identity process*—and prove it converges to a highly irregular limit process whose trajectories are not right-continuous and impossible to study using standard probability in metric spaces. However, when this process is integrated over time, and appropriately rescaled and centered, it becomes possible to define a functional limit theorem for it, which then converges to a randomly pinned Brownian motion. By applying these theoretical results, a new goodness-of-fit test is derived. We demonstrate that this test is very efficient when it is applied to data which come from a multimodal or mixture distribution, like the classic Old Faithful dataset. *The Canadian Journal of Statistics* 46: 656–672; 2018 © 2018 Statistical Society of Canada

Résumé: Le comportement de la fonction de répartition empirique et de la fonction quantile empirique est bien connu lorsque celles-ci sont calculées avec un nombre croissant d'observations indépendantes tirées uniformément de l'intervalle unité. Les auteurs étudient des résultats asymptotiques similaires pour la fonction obtenue par la composée de la fonction quantile empirique avec la fonction de répartition empirique. Puisque la première est l'inverse généralisée de la deuxième, leur composée approximera la fonction identité. Les auteurs définissent une version centrée et normalisée de cette fonction–le *processus identité empirique*–et prouvent qu'il converge vers un processus limite très irrégulier dont les trajectoires sont discontinues à droite et impossibles à étudier avec les probabilités habituelles dans un espace métrique. Toutefois, lorsque ce processus est intégré dans le temps et qu'il est centré et réduit de façon appropriée, il devient possible de définir un théorème limite fonctionnel qui converge vers un pont brownien. Les auteurs exploitent ces nouveaux résultats théoriques afin de définir un nouveau test d'adéquation. Ils démontrent que ce test est très efficace lorsqu'il est utilisé pour des données multimodales ou des mélanges, comme pour le jeu de données classique du geyser Old Faithful. *La revue canadienne de statistique* 46: 656–672; 2018 © 2018 Société statistique du Canada

1. INTRODUCTION

The asymptotic behaviour of both the uniform empirical process and the uniform quantile process is well known, for example, Csörgő & Révész (1981), Csörgő (1983), Shorack & Wellner (1986).

Additional Supporting Information may be found in the online version of this article at the publisher's website.
* *Author to whom correspondence may be addressed.*
E-mail: enrico.bibbona@polito.it

Such processes are defined as centered and scaled versions of the empirical distribution function and the empirical quantile function, respectively. To the best of our knowledge this article is the first investigation of the process that results from applying the empirical quantile function to the empirical distribution function itself. Intuitively, such a back-and-forth operation approximates the identity function. We introduce a centered and scaled version of this process which we define as the *empirical identity process* (EIP). The asymptotic properties of the EIP are somewhat unexpected; the EIP converges in distribution to a white noise process whose finite-dimensional distributions (FDD) are products of exponential or Laplace distributions; see the discussion following Theorem 1. This limiting process is very irregular and, in particular, is not right-continuous. It is therefore impossible to build a proper weak convergence theory in any metric space. In the hope of achieving regularity, we study the integral of the EIP. The resulting limit theorem is a functional version of a classical result by Moran (1947) on the asymptotic behaviour of the sum of squared spacings.

Our results provide some unexpected asymptotic properties of a process related to the uniform empirical and quantile functions. Moreover, we demonstrate that these asymptotic results have interesting applications from a statistical point of view. Based on the asymptotics of the EIP, we propose a new test statistic that can be used in goodness-of-fit problems. This new methodology is not always superior to other methods based on the empirical distribution function, but it performs better when the true distribution is multimodal or a mixture. Using simulations, we identify cases in which this new statistic outperforms existing alternatives; we also provide a relevant application to the popular Old Faithful dataset.

2. EMPIRICAL IDENTITY PROCESSES AND ASYMPTOTIC RESULTS

Let $U_1, \dots, U_n$ be independent random variables from a uniform distribution on [0,1]. Let $U_{n,1} \le \cdots \le U_{n,n}$ be their order statistics, together with $U_{n,0} = 0$ and $U_{n,n+1} = 1$. Let $\mathbb{F}_n(t) = n^{-1}\sum_{i=1}^{n}(U_i \le t)$, $0 \le t \le 1$, denote the empirical distribution function and $\mathbb{Q}_n(u) = \inf\{t \in [0,1] : \mathbb{F}_n(t) \ge u\}$ the empirical quantile function, $0 < u \le 1$; $\mathbb{Q}_n(\cdot)$ is the left-continuous generalized inverse function of $\mathbb{F}_n(\cdot)$.

We next define the *lower empirical identity function* as

$$R_n^L(t) = U_{n,n\mathbb{F}_n(t)} = \begin{cases} 0 & \text{if } 0 \le t < U_{n,1} \\ \mathbb{Q}_n\left(\mathbb{F}_n(t)\right) & \text{if } U_{n,1} \le t \le 1, \end{cases}$$

the *upper empirical identity function* as

$$R_n^U(t) = U_{n,n\mathbb{F}_n(t)+1} = \begin{cases} \mathbb{Q}_n\left(\mathbb{F}_n(t) + \frac{1}{n}\right) & \text{if } 0 \le t < U_{n,n} \\ 1 & \text{if } U_{n,n} \le t \le 1, \end{cases}$$

and the *empirical identity function* as their average

$$R_n(t) = \frac{R_n^L(t) + R_n^U(t)}{2}.$$

The trajectories of $R_n(t)$, $R_n^L(t)$, and $R_n^U(t)$ for a specific sample of size $n = 2$ are shown in Figure 1. By the Glivenko-Cantelli theorem, as $n \to \infty$, the three random sequences, $R_n(t)$, $R_n^L(t)$, and $R_n^U(t)$ converge almost surely in the uniform norm to the identity function. It is therefore interesting to study their second-order asymptotics by defining the *lower* and *upper empirical identity process*

$$Y_n^L(t) = (n+1)\{R_n^L(t) - t\} \quad \text{and} \quad Y_n^U(t) = (n+1)\{R_n^U(t) - t\},$$

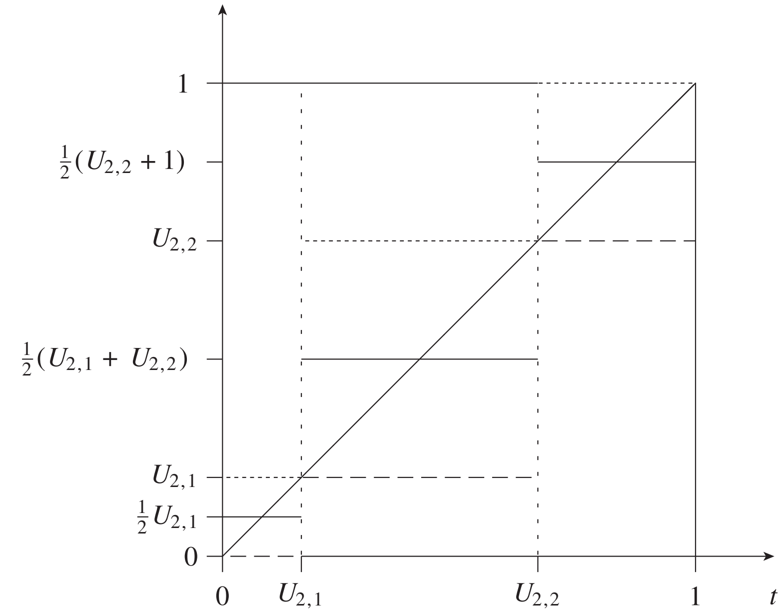

FIGURE 1: Trajectories of the EIP (solid), of the lower EIP (dashed) and of the upper EIP (dotted) with $n = 2$, $U_{2,1} = 1/6$ and $U_{2,2} = 4/6$.

and the *empirical identity process*

$$Y_n(t) = (n+1)\{R_n(t) - t\} = \{Y_n^L(t) + Y_n^U(t)\}/2,$$

for $0 \leq t \leq 1$. We use the scaling factor $n+1$ instead of n to make notation in the following sections simpler.

Theorem 1. *For any positive integer k and points $0 < u_1 < \cdots < u_k < 1$, the random vector $(Y_n^U(u_1), \ldots, Y_n^U(u_k), -Y_n^L(u_1), \ldots, -Y_n^L(u_k))$ converges in distribution to a vector of 2k independent exponential random variables, as $n \to \infty$.*

In other words, the joint FDD of the bivariate process $(Y_n^U(\cdot), -Y_n^L(\cdot))$ converge to those of two independent exponential variates. As a consequence, for the EIP itself we can conclude that

Corollary 1. *For any positive integer k and points $0 < u_1 < \cdots < u_k < 1$, as $n \to \infty$ the random vector $(Y_n(u_1), \ldots, Y_n^U(u_k))$ converges in distribution to a vector of k independent random variables with the Laplace density $f(z) = \exp(-2|z|)$.*

The proof of Theorem 1 is long and requires two preliminary lemmas; it is sketched in the appendix. A process with independent FDD cannot be right-continuous; see the criterion in Theorem 13.6 of Billingsley (1999). Therefore, the theory of weak convergence in $D(0, 1)$, the usual space of cadlag paths, does not apply. The resulting conclusion is that such limit processes are fairly intractable objects and for any statistical application we need to regularize them.

3. THE INTEGRATED EMPIRICAL IDENTITY PROCESS

The anti-derivative of a function is always more regular than the function itself. Therefore, in the hope of obtaining a more regular limit process, it is natural to look at the asymptotic behaviour of the integrals of the processes defined in the previous section.

Since it turns out that the asymptotic behaviours of the lower and upper EIPs are equivalent—see Section 3.2 for further details—it is simpler and notationally convenient to study only the integrated lower EIP,

$$I_n(t) = -\int_0^t Y_n^L(u)du = (n+1)\int_0^t \{u - R_n^L(u)\}du, \quad t \in [0,1], \tag{1}$$

which hereafter we simply call the *integrated process*; the minus sign makes it non-negative.

3.1. The Integrated Process and Its Relation to Spacings

A simple geometric inspection of Figure 1 shows that the integrated process is strictly related to the uniform *spacings*, which are defined as

$$D_{n,i} = U_{n,i} - U_{n,i-1}, \quad i = 1, \ldots, n+1. \tag{2}$$

We obtain

$$\begin{aligned} I_n(t) &= (n+1)\int_0^t \{u - R_n^L(u)\} \ du \\ &= (n+1)\sum_{i=1}^{nF_n(t)} \int_{U_{n,i-1}}^{U_{n,i}} (u - U_{n,i-1})du + (n+1)\int_{U_{n,nF_n(t)}}^t \{u - U_{n,nF_n(t)}\}du \\ &= \frac{n+1}{2}\sum_{i=1}^{nF_n(t)} D_{n,i}^2 + \frac{n+1}{2}\{t - R_n^L(t)\}^2. \end{aligned} \tag{3}$$

In particular, at $t = 1$ the integrated process equals

$$I_n(1) = (n+1)\int_0^1 \{u - R_n^L(u)\} \ du = \frac{n+1}{2}\sum_{i=1}^{n+1} D_{n,i}^2.$$

This is the well-known Greenwood statistic, for which a classical theorem due to Moran (1947) established convergence to normality in the following way:

Theorem 2. *The following convergence in law holds: $n \to \infty$,*

$$M_n = \sqrt{n+1}\{I_n(1) - 1\} = \sqrt{n+1}\left(\frac{n+1}{2}\sum_{i=1}^{n+1} D_{n,i}^2 - 1\right) \xrightarrow{L} N(0,1). \tag{4}$$

Theorem 2 refers to the convergence of $I_n(1)$, the value of the integrated process when $t = 1$. More can be said about the convergence of $I_n(\cdot)$ as a process in the functional space $D(0,1)$. In other words, we are in a position to extend Moran's theorem into a *functional version*. To do so, we need some technical steps which involve strong approximation theorems. The rest of this section can be skipped by those not interested in the particular probabilistic details; the important result is Theorem 4, which is found in the next subsection.

Recall first the following strong approximation results of Aly (1983, 1988):

Theorem 3. *There exists a probability space on which both*

- *a two-dimensional Wiener process* $(W_1(\cdot), W_2(\cdot))$ *with zero mean and autocovariance matrix*

$$\mathbb{E}\left[\begin{pmatrix} W_1(s) \\ W_2(s)) \end{pmatrix} \; (W_1(t)W_2(t))\right] = \min\{s,t\}\begin{pmatrix} 1 & 4 \\ 4 & 20 \end{pmatrix}, \quad t,s>0,$$

- *and a vector* $\{D_{n,i}\}_{i=1,\ldots,n+1}$ *of random variables of arbitrary size n with the same law as the uniform spacings defined in (2)*

are defined, such that the two processes

$$E_n(t) = \begin{cases} 0 & \text{if } 0 \le t < \frac{2}{n+1} \\ \sqrt{n+1}\left\{(n+1)\sum_{i=1}^{\lfloor (n+1)t \rfloor} D_{n,i}^2 - 2t\right\} & \text{if } \frac{2}{n+1} \le t \le 1, \end{cases}$$

$$V_n(t) = \frac{1}{\sqrt{n+1}}\{W_2((n+1)t) - 4tW_1(n+1)\}\ , \quad 0 \le t \le 1, \tag{5}$$

are so close to each other that the following condition holds: for every $\varepsilon > 0$ *there are constants A, B such that*

$$P\left\{\sup_{0\le t\le 1} |E_n(t) - V_n(t)| > A\frac{\log(n+1)}{\sqrt{n+1}}\right\} \le B(n+1)^{-\varepsilon}. \tag{6}$$

Theorem 3 was first stated in Aly (1983), but a minor step of the proof was not fully justified. This led the author to write Aly (1988), which provided a rigorous proof of Theorem 3 based on a multivariate Hungarian construction due to Einmahl (1989). However, the rate of convergence in Einmahl (1989) is suboptimal and not fast enough to allow for the rate $\log(n+1)/\sqrt{n+1}$ found in (6). Theorem 3 is then provided in Aly (1988) with a rate of convergence that is slowed to $\{\log(n+1)\}^2/\sqrt{n+1}$. We here conclude that the statement in Aly (1983) is actually correct. The reason is that a stronger multivariate Hungarian construction is now available that allows us to prove Theorem 3 using the same proof that appears in Aly (1988), provided Zaitsev (1998) is cited in place of Einmahl (1989). The process $V_n(\cdot)$ defined in (5) is constructed by applying such a bivariate Hungarian construction. Different processes arise for different choices of n, but their law is actually the same irrespective of n, since they are all centered Gaussian processes with the covariance function

$$\mathbb{E}\{V_n(t)V_n(s)\} = 20\min(t,s) - 16st.$$

Moreover, each process V_n has the same law as the process

$$V(t) = 2\sqrt{5}W(t) - 2(\sqrt{5}-1)tW(1), \quad 0 \le t \le 1, \tag{7}$$

where $W(\cdot)$ is a standard one-dimensional Wiener process. A consequence of Theorem 3 and of the previous remarks is the following corollary:

Corollary 2. *The process* $E_n(\cdot)$ *converges weakly in* $D(0,1)$ *to the process* $V(\cdot)$ *defined in* (7).

3.2. A New Functional Version of Moran's Theorem

Since the statistic M_n introduced in (4) equals $E_n(1)/2$, the previous corollary may already be considered a functional version of Moran's theorem. However, it involves the process $E_n(\cdot)$ and not the process $I_n(\cdot)$ which is of interest here. We therefore need another functional version of Moran's theorem which provides the asymptotics for $I_n(\cdot)$ directly.

Theorem 4. *The process* $2\sqrt{n+1}\{I_n(\cdot) - \mathbb{F}_n(\cdot)\}$ *converges weakly in D(0, 1) to the Gaussian process* $V(\cdot)$ *defined in (7).*

A full proof of the theorem is given in the appendix. Due to the continuous mapping theorem and to the continuity of the supremum operator and of the absolute value, we obtain the following corollary, which we will use below to construct a new goodness-of-fit test.

Corollary 3. *The random variable* $2\sup_{0\le t\le 1}\sqrt{n+1}\ |\ I_n^L(t) - \mathbb{F}_n(t)\ |$ *converges weakly to* $\sup_{0\le t\le 1} |\ V(t)|$ *as* $n \to \infty$.

At the beginning of Section 3 we introduced the integrated process by focusing on the integral of the lower EIP. However it is legitimate to consider what would happen if we instead used the integral of the upper EIP, that is,

$$I_n^U(t) = \int_0^t Y_n^U(u)du = (n+1)\int_0^t \{R_n^U(u) - u\}du, \qquad t \in [0, 1],$$

or if we considered the joint distribution of these two processes. To provide a satisfactory answer a more formal proof would be required, but it can be seen from Figure 1 that the difference between $I_n(t)$ and $I_n^U(t)$ is uniformly bounded by $(n+1)/2$ times the squared maximal spacing. Therefore we can apply the classical results in Slud (1978) concerning the almost sure rate of convergence to zero of the maximal spacing to show that the difference between $\sqrt{n+1}\{I_n(t) - \mathbb{F}_n(t)\}$ and $\sqrt{n+1}\{I_n^U(t) - \mathbb{F}_n(t)\}$ vanishes almost surely as the sample size tends to infinity. Consequently, the pair $(\sqrt{n+1}\{I_n(t) - \mathbb{F}_n(t)\}, \sqrt{n+1}\{I_n^U(t) - \mathbb{F}_n(t)\})$ converges weakly to two identical copies of the process $V(\cdot)$ identified in (7).

3.3. Characterization of the Limit Process and the Asymptotic Distribution of the Maximum

To use Corollary 3 we need to compute the exact distribution of the supremum of the limit process. That required result is identified below in Theorem 5.

Let $W(\cdot)$ be standard Brownian motion and $B(t) = W(t) - tW(1)$, $t \ge 0$ be a Brownian bridge. For every t, $B(t)$ is independent of $W(1)$ and it has the distribution of Brownian motion which is constrained to visit zero (or ''pinned to 0'') at time 1. A process $B(t) + ty$, $t \ge 0$ also represents Brownian motion pinned to y when $t = 1$, cf. Revuz & Yor (1999). Therefore, the limit process $V(\cdot)$ that we identified in (7) admits the equivalent representation $V(t) = 2\sqrt{5}\left\{B(t) + t\frac{W(1)}{\sqrt{5}}\right\}$, $t \ge 0$, which constitutes Brownian motion, scaled by a factor $2\sqrt{5}$, and which at $t = 1$ is pinned to a random position $W(1)/\sqrt{5}$. Moreover, we can derive the two-sided maximal probability distribution for $V(t)$ from that of the pinned Brownian motion; see either Beghin & Orsingher (1999), Equation 4.12 or Borodin & Salminen (2002), Part II, Chapter 1, Equation 1.15.8(1).

Theorem 5. *The distribution function of the maximum of the absolute value of the stochastic process $V(\cdot)$ is given by*

$$\mathbb{P}\{\sup_{0\le t\le 1} |V(t)| < b\} = \sum_{h=-\infty}^{\infty} (-1)^h e^{-\frac{4}{50}h^2 b^2}. \tag{8}$$

A detailed proof of this result is given in the appendix.

4. A NEW GOODNESS-OF-FIT TEST

Let $X_1, \dots, X_n$ be a sample of independent continuous random variables, with possibly different distributions. Under the null hypothesis H_0 that the distribution functions of the X_i are some given $F_i(x)$, $i = 1, \dots, n$, the transformed sample $\{\widehat{U}_i = F_i(X_i)\}$, $i = 1, \dots, n$ is composed of independent uniform random variables. The integrated process of the transformed sample can be used to construct a new goodness-of-fit test of H_0 in the same spirit as the Kolmogorov–Smirnov goodness-of-fit test.

4.1. A Statistic Derived From the Supremum of the Integrated EIP

Let $\widehat{\mathbb{F}}_n(t)$ be the empirical distribution function of the transformed sample and let $\widehat{I}_n(t)$ be the related integrated process that we first identified in (1). We now define the test statistic

$$d_n = 2 \sup_{0\le t\le 1} \sqrt{n+1} \mid \widehat{I}_n(t) - \widehat{\mathbb{F}}_n(t) \mid . \tag{9}$$

Under H_0, the sequence d_n converges weakly to $\sup_{0\le t\le 1} \mid V(t)|$, whose exact distribution is identified in (8). Notice that the distribution of d_n is the same, irrespective of the distributions $F_i(x)$ of the individual sample observations.

Now define a new goodness-of-fit test which rejects H_0 if the value of d_n exceeds a critical value. An asymptotic critical value can be derived by numerically inverting (8), for example the 95th percentile of $\sup_{0\le t\le 1} \mid V(t) \mid$ equals $b = 6.790494$. Numerical simulations show that the convergence is slow. For intermediate values of n, say 100, the 0.95 quantiles of d_n are not very well approximated by the asymptotic values, but suitable values can easily be derived using Monte Carlo methods. The results of such computations are summarized in Table 1.

The critical values displayed in Table 1 were obtained by simulating from the uniform distribution. There might be some loss in accuracy when the sample data are generated from other distributions and transformed through the cumulative distribution function. To check for such a possibility we performed the following consistency check: we tested the goodness-of-fit with respect to the true distribution of the data, looking for the rate of occurrence of type I errors; see the source code provided in the Supplementary material. We do not report the observed values here, but did not detect any relevant discrepancies with respect to the nominal significance level of 5%.

4.2. Numerical Simulations

We carried out various numerical experiments to benchmark the performance of these new tests relative to other well-known goodness-of-fit tests such as that based on the original statistic M_n of Moran (1947), that is, (4). Other possible competitors are the classical Kolmogorov–Smirnov test, which is based on the statistic

$$D_n = \sup_t \sqrt{n} \mid \mathbb{F}_n(t) - F(t) \mid = \sup_t \sqrt{n} \mid \widehat{\mathbb{F}}_n(t) - t \mid,$$

TABLE 1: Approximate values of the 0.95 quantiles of the distributions of d_n, M_n, A_n and D_n for different sample sizes n, computed by Monte Carlo simulations. The values in the final column are the asymptotic ones.

	Sample size, n					
Statistic	30	50	100	200	272	∞
d_n	5.857	6.127	6.349	6.493	6.536	6.790
M_n	1.449	1.559	1.637	1.679	1.684	1.645
A_n	2.493	2.497	2.495	2.494	2.490	2.492
D_n	1.322	1.332	1.341	1.346	1.346	1.358

and the Anderson−Darling test, which belongs to the Cramer−von Mises family of tests and is based on the statistic

$$A_n = n \int_0^1 w(u)\{\widehat{\mathbb{F}}_n(u) - u\}^2 du,$$

which, with the weight function $w(x) = 1/\{x(1 - x)\}$, is designed to identify possible departures from H_0 in the tails.

For goodness-of-fit tests of this kind, the alternative hypothesis is completely nonparametric. For data not generated according to the null distribution, the power of the test is strongly influenced by the choice of the *alternative distribution* from which sampled observations arise. For example, if we test the normality of a sample which was generated from a Student-t distribution with the same mean, we expect a test based on A_n to have greater power than one based on D_n due to the differences in the tails between the null and alternative distributions.

Now, if the dataset is generated from a mixture of two normal distributions having the same variance σ^2 but different means μ_1 and μ_2, we expect that a test based on the empirical distribution may not easily find significant discrepancies between the data and a normal distribution with mean equal to a linear combination of μ_1 and μ_2, and variance somewhat larger then σ^2, for example, the variance of the mixture. However, a test based on the spacings, such as the one using d_n, could exhibit greater power. The intuitive justification is that the gap in the regularized data between the two models would easily give rise to some large spacings that could make d_n significantly larger than is likely to arise in the uniform case.

We therefore checked the power of the various tests via simulation in situations where the sample data were generated using mixture distributions. Our purpose was, first, to highlight situations where d_n performs better than competing goodness-of-fit tests. For completeness, we also studied a second set of examples involving less favourable conditions.

In all cases we evaluated the estimated power of the test via simulation using the ratio of the number of rejections and the number of simulated samples, since the distribution of the observed sample was always different from the distribution under the null hypothesis.

All numerical experiments were carried out using the R environment for statistical computing; see R Core Team (2017). We used built-in functions to generate samples and to compute the value of the test statistic D_n. For A_n we relied on the `goftest` package, whereas to compute M_n and d_n we wrote our own source code. All of our R scripts have been included as Supplementary material in order to ensure reproducibility and to allow interested readers to try other scenarios as well.

We carried out Monte Carlo evaluations of the 0.95 quantiles of the null distributions for the different test statistics using 10^6 uniform samples of lengths 30, 50, 100, 200 and 272, respectively. These simulated values were then used as critical values for the tests, in order to provide a fair comparison of the methodologies. The various quantiles that we obtained and subsequently used are summarized in Table 1. Note that while the quantiles of A_n and D_n are already very close to their asymptotic values when $n = 30$, the same statement is not true for d_n and M_n; their convergence with respect to sample size was much slower.

We next evaluated the power of the different tests for selected examples. In particular, we chose a first set of examples whose common feature is that the sample data originated from one of the following mixture distributions, while the null hypothesis specified that they arose from a single, common source:

- The null hypothesis specified that our sample arose from the standard normal distribution. We generated the 10,000 simulated samples, each of size 100, from a mixture of two normal distributions with equal weights, a common standard deviation 0.45 and means 0.88 and −0.88. We chose these particular values so that the mixture distribution would have an approximate mean and variance of zero and one, respectively. We call this alternative distribution a symmetric normal mixture.
- The null hypothesis specified that our sample arose from the standard normal distribution. We generated the 10,000 simulated samples, each of size 30, from a mixture of two normal distributions. The weights of the mixture were $p_1 = 1/5$ and $p_2 = 4/5$, the means were $\mu_1 = 1.68$ and $\mu_2 = 0.42$, and the standard deviations were $\sigma_1 = 0.2$ and $\sigma_2 = 0.6$, respectively. Again, we chose these particular values so that the mixture distribution would have an approximate mean and variance of zero and one, respectively. We call this alternative distribution an asymmetric normal mixture.
- The null hypothesis specified that our sample arose from the uniform distribution on (0, 1). We generated the 10,000 simulated samples, each of size 100, from an equally weighted mixture of two beta distributions with parameters (2, 8) and (8, 2), respectively.

The results of these various simulations are summarized in Table 2. Both the d_n and M_n statistics, which are based on spacings, outperformed the competing tests based on the empirical distribution. Moreover, the test based on d_n had the greatest estimated power in all three examples.

Of course, no test is uniformly superior against all types of alternatives. In more standard situations we would not expect our methodology to outperform the classical methods based on

TABLE 2: A first group of examples, involving mixtures. Estimated power of each test at a nominal significance level of 5% for the goodness-of-fit test based on $d_n = 2\sup_{0 \le t \le 1} \sqrt{n+1} \mid \widehat{I}_n(t) - \widehat{\mathbb{F}}_n(t) \mid$ compared to the Moran test based on M_n, the Kolmogorov−Smirnov test based on D_n and the Anderson−Darling test based on A_n.

		Estimated power for nominal 5% significance			
H_0	True distribution	d_n	M_n	D_n	A_n
Normal	Symmetric normal mixture	0.698	0.644	0.618	0.521
Normal	Asymmetric normal mixture	0.553	0.527	0.232	0.179
Uniform	Beta mixture	0.858	0.788	0.792	0.753

TABLE 3: A second set of examples, not involving mixtures. Estimated power of each test at a nominal significance level of 5% for the goodness-of-fit test based on $d_n = 2\sup_{0\le t\le 1}\sqrt{n+1}\ |\widehat{I}_n(t) - \widehat{\mathbb{F}}_n(t)|$ compared to the Moran test based on M_n, the Kolmogorov–Smirnov test based on D_n and the Anderson–Darling test based on A_n.

		Estimated power for nominal 5% significance			
H_0	True distribution	d_n	M_n	D_n	A_n
Normal	Cauchy	0.657	0.769	0.261	0.998
Normal	Student t	0.346	0.383	0.252	0.982
Normal	Shifted normal	0.554	0.207	0.732	0.824

the empirical distribution function. A second group of examples follows. The null hypothesis was always that our sample arose from the standard normal distribution, while the various alternatives that we considered were the following:

- the 10,000 simulated samples, each of size 100, were generated from a Cauchy distribution with scale parameter 0.5;
- the 10,000 simulated samples of size 100 were generated from a Student's t distribution with two degrees of freedom;
- the 10,000 simulated samples, each of size 100, were generated from a normal distribution with mean 0.3.

The results of these simulations are summarized in Table 3. Both the d_n and M_n statistics, based on spacings, were inferior to A_n, which had the greatest estimated power in all three scenarios.

4.3. An Application to the Old Faithful Dataset

Old Faithful is a geyser in Yellowstone National Park, Wyoming, USA. For centuries it has been erupting several times a day, spewing streams of hot water high into the sky. A popular dataset, consisting of 272 observations on the waiting times between eruptions and the corresponding durations of the eruptions, is distributed with R (R Core Team, 2017). We focus on the waiting times, which exhibit a bimodal distribution, as illustrated in Figure 2. The data are in minutes (integers), with a sample mean and standard deviation of 70.90 and 13.59, respectively. In order to avoid ties we jittered the raw data by adding Gaussian noise with mean zero and standard deviation 0.4. All tests reject normality of the sample if the null mean is fixed at 71 and the null standard deviation at 14. We do not report P-values, but our source code is available in the Supplementary material. If subsamples of size 50 are taken from the dataset, then rejection of the same null hypothesis is no longer assured. We selected 10,000 subsamples, each of size 50, at random and ran all four goodness-of-fit tests on each subsample. The observed results are summarized in Table 4. The test based on d_n seems to be able to reject the null hypothesis of normality with greater estimated power than each of the three alternative competitors.

5. CONCLUSIONS

We have defined the lower EIP and found that it converges to a process which has highly irregular trajectories. In the hope of achieving regularity we studied the limiting behaviour of its integral, obtaining Theorem 4 as our main result. We also computed the explicit limiting distribution of

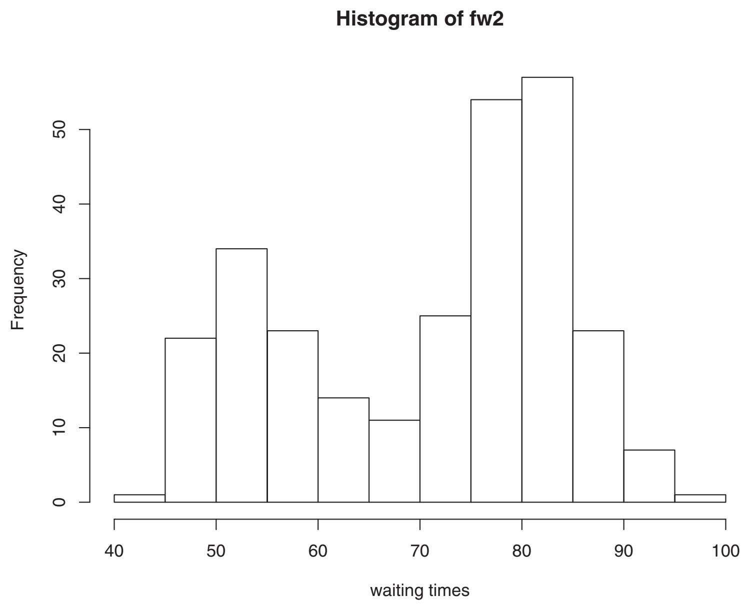

FIGURE 2: Observed histogram of the waiting times, in minutes, between successive eruptions of Old Faithful.

TABLE 4: Observed results of the analysis of subsamples of the intervals, in minutes, between successive eruptions of Old Faithful.

Total sample size	272
Number of subsamples	10,000
Size of subsamples	50
Rejections using d_n	6,692
Rejections using M_n	6,282
Rejections using D_n	4,369
Rejections using A_n	2,976

the running supremum of the integrated process, which is identified in Theorem 5. This result has an important statistical application, leading to a new goodness-of-fit test based on spacings; see Section 4. A comparative study of our newly identified goodness-of-fit test with respect to classic competitors using the Old Faithful dataset supports the conclusion that this new test is useful for multimodal data, such as those arising from mixtures of distributions.

APPENDIX

In the following we provide proofs of the various results in the article.

Proof of Theorem 1. Theorem 1 states that the FDDs of the bivariate process $(Y_n^U(\cdot), -Y_n^L(\cdot))$ converge weakly to the FDDs of two independent exponential white noise processes. Before proving the theorem, we establish two useful lemmas.

Consider k fixed distinct numbers $u_1 < u_2 < \cdots < u_k$ in the interval (0, 1) and let $u_0 = 0$ and $u_{k+1} = 1$ be the extreme points. We will be working with the $k + 1$ bins induced by these points and with the order statistics from the uniform i.i.d. process $U_1, U_2, \ldots$ that belong to the different bins. In particular, let

$$\begin{aligned} C_n &= C_n(u_1, u_2, \ldots u_k) = (C_{n,1}, C_{n,2}, \ldots, C_{n,k}, C_{n,k+1})' \\ &= n \cdot (F_n(u_1), F_n(u_2) - F_n(u_1), \ldots, F_n(u_k) - F_n(u_{k-1}), 1 - F_n(u_k))' \end{aligned}$$

be the sequence of the vectors of counts of i.i.d. uniform observations $U_1, U_2, \ldots, U_n$ that belong to the different bins, $n = 1, 2, \ldots$. In order to keep the notation simple, we consider the dependence of C_n on $u_1, u_2, \ldots, u_k$ to be understood. It is well known that the distribution of C_n is multinomial with parameters n, u_1, $u_2 - u_1, \ldots, u_k - u_{k-1}$, $1 - u_k$; that is, the probability mass function of the vector $(C_{n,1}, C_{n,2}, \ldots, C_{n,k})'$, evaluated at the vector of non-negative integers $(c_1, \ldots, c_k)$, with $\sum_{j=1}^{k} c_j \le n$, is

$$n! \prod_{j=1}^{k+1} \frac{(u_j - u_{j-1})^{c_j}}{c_j!}, \tag{A.1}$$

where $c_{k+1} = n - \sum_{j=1}^{k} c_j$. ■

Lemma 1. *Let $U_1, \ldots, U_n$ be a sequence of i.i.d. uniform random variables on* [0, 1]. *Let $U_{n,1} \le \cdots \le U_{n,n}$ be their order statistics. The constants $0 = u_0 < u_1 < u_2 < \cdots < u_k < u_{k+1} = 1$ induce a partition of the interval* (0, 1) *into bins. The order statistics $U_{n,1}, \ldots, U_{n,n}$ may be subdivided into groups that belong the different bins. The counts of the order statistics falling into each of the bins can be summarized via the multinomial vector C_n. The vectors of order statistics belonging to each bin are conditionally independent, given C_n. Each of these vectors has a conditional distribution equal to the distribution of the order statistics of $C_{n,j}$ i.i.d. uniform observations on the interval (u_{j-1}, u_j).*

Proof of Lemma 1. The density of the order statistics $U_{n,1} \le \cdots \le U_{n,n}$ evaluated at $0 < x_1 < \cdots < x_n < 1$ is $n!\ (0 < x_1 < x_2 < \cdots < x_n < 1)$. In order to write the joint density of the order statistics and the vector $C_n = (c_1, \ldots, c_k, c_{k+1})'$, a compatibility factor $\prod_{j=1}^{k} (x_{c_1+\cdots+c_j} < u_j < x_{c_1+\cdots+c_{j+1}})$ must be included. The conditional density of $U_{n,1} \le U_{n,2} \le \cdots \le U_{n,n}$ given $C_n = (c_1, \ldots, c_k, c_{k+1})'$ is

$$\begin{aligned} &\frac{n!(0 < x_1 < \cdots < x_n < 1) \prod_{j=1}^{k} (x_{c_1+\cdots+c_j} < u_j < x_{c_1+\cdots+c_{j+1}})}{n! \prod_{j=1}^{k+1} (u_j - u_{j-1})^{c_j} / c_j!} \\ &= \prod_{j=1}^{k+1} \frac{c_j!}{(u_j - u_{j-1})^{c_j}} (u_{j-1} < x_{c_0+\cdots+c_{j-1}+1} < \cdots < x_{c_0+\cdots+c_j} < u_j), \end{aligned}$$

where $c_0 = 0$, which factors into the $k + 1$ marginal densities of the vectors of adjacent order statistics belonging to the $k + 1$ bins. It should be apparent that their distributions are exactly as described in the statement of the theorem.

The following lemma is a corollary of Lemma 1.

Lemma 2. *The conditional density of the two-dimensional vector* $(R_n^U(u_{j-1}), R_n^L(u_j))'$ *given* $C_n = (c_1, \dots, c_k, c_{k+1})'$*, evaluated at (x, y), is*

$$\frac{c_j(c_j-1)(y-x)^{c_j-2}}{(u_j-u_{j-1})^{c_j}}(u_j < x < y < u_{j-1}), \tag{A.2}$$

provided $c_j > 1$*, for each* $j = 1, \dots, k+1$.

Proof of Lemma 2. Given C_n, by the previous lemma $R_n^U(u_{j-1})$ and $R_n^L(u_j)$ have the same distribution as the minimum and the maximum, respectively, of c_j uniform observations on the interval (u_{j-1}, u_j). It is easy to check that their density is given by the formula found in (A.2).

Proof of Theorem 1. Consider the moment generating function

$$\varphi_{Y_n^U(u_1),\dots,Y_n^U(u_k),-Y_n^L(u_1),\dots,-Y_n^L(u_k)}(v_1, \dots, v_k, w_1, \dots, w_k)$$

of the vector $(Y_n^U(u_1), \dots, Y_n^U(u_k), -Y_n^L(u_1), \dots, -Y_n^L(u_k))'$ evaluated at $(v_1, \dots, v_k, w_1, \dots, w_k)'$.

We will examine its limiting form as $n \to \infty$. For each n an explicit form can be found by conditioning on C_n, viz.

$$\begin{aligned}
&\varphi_{Y_n^U(u_1),\dots,Y_n^U(u_k),-Y_n^L(u_1),\dots,-Y_n^L(u_k)}(v_1, \dots, v_k, w_1, \dots, w_k)\\
&= \mathbb{E}\left(\exp\left[\sum_{j=1}^{k}\left\{v_j Y_n^U\left(u_j\right) - w_j Y_n^L(u_j)\right\}\right]\right)\\
&= \mathbb{E}\left\{\mathbb{E}\left(\exp\left[\sum_{j=1}^{k}\left\{v_j Y_n^U\left(u_j\right) - w_j Y_n^L(u_j)\right\}\right] \mid C_n\right)\right\}\\
&= \mathbb{E}\left(\prod_{j=1}^{k+1}\mathbb{E}\left[\exp\left\{v_{j-1} Y_n^U\left(u_{j-1}\right) - w_j Y_n^L(u_j)\right\} | C_n\right]\right),
\end{aligned}$$

where $v_0 = w_{k+1} = 0$ for the sake of obtaining a compact expression. To obtain the last line, we used the conditional independence property that we derived in Lemma 1 and grouped the random variables relative to the different bins.

Now we apply Lemma 2 to obtain an explicit form for the random variable $\mathbb{E}[\exp\{v_{j-1}Y_n^U(u_{j-1}) - w_j Y_n^L(u_j)\}|C_n]$.

By the strong law of large numbers, for every ε there exist a set A_ε such that $\mathbb{P}(A_\varepsilon) = 1$ and for all $\omega \in A_\varepsilon$ there exist an $N(\omega, \varepsilon)$ such that for all $n > N(\omega, \varepsilon)$

$$\min_{j=1,\dots,k+1} \frac{C_{n,j}}{n}(\omega) > u_j - u_{j-1} - \varepsilon.$$

On any such set for all $n > \max\left\{N(\omega), \frac{1}{u_j - u_{j-1} - \varepsilon}\right\}$ we have

$$\min_{j=1,\dots,k+1} C_{n,j}(\omega) > 1.$$

Without loss of generality, we can work on the set A_ε as long as only the asymptotic results are of interest. Then, on this set A_ε, for any n large enough we have

$$
\begin{aligned}
&\mathbb{E}[\exp\{v_{j-1}Y_n^U(u_{j-1}) - w_j Y_n^L(u_j)\}|C_n] \\
&= \int_{u_{j-1}}^{u_j}\int_x^{u_j} \exp\{v_{j-1}(n+1)(x-u_{j-1}) - w_j(n+1)(y-u_j)\} \\
&\quad \times \frac{C_{n,j}(C_{n,j}-1)(y-x)^{C_{n,j}-2}}{(u_j-u_{j-1})^{C_{n,j}}} dydx \\
&= \int_0^1\int_0^{1-s} \exp\{v_{j-1}(n+1)(u_j-u_{j-1})s + w_j(n+1)(u_j-u_{j-1})t\} \\
&\quad \times C_{n,j}(C_{n,j}-1)(1-s-t)^{C_{n,j}-2} dtds \\
&= 1 + I_{j1} + I_{j2} + I_{j3}
\end{aligned}
$$

after the exchange of variable $s = (x - u_{j-1})/(u_j - u_{j-1})$ and $t = (u_j - y)/(u_j - u_{j-1})$ and several integrations by parts; here

$$
\begin{aligned}
I_{j1} &= v_{j-1}(n+1)(u_j-u_{j-1})\int_0^1 \exp\{v_{j-1}(n+1)(u_j-u_{j-1})s\}(1-s)^{C_{n,j}} ds, \\
I_{j2} &= w_j(n+1)(u_j-u_{j-1})\int_0^1 \exp\{w_j(n+1)(u_j-u_{j-1})t\}(1-t)^{C_{n,j}} dt, \\
I_{j3} &= v_{j-1}w_j(n+1)^2(u_j-u_{j-1})^2 \\
&\quad \times \int_0^1\int_0^{1-t} \exp\{v_{j-1}(n+1)(u_j-u_{j-1})s + w_j(n+1)(u_j-u_{j-1})t\}(1-s-t)^{C_{n,j}} ds \;\; dt.
\end{aligned}
$$

Now, by the law of large numbers we have $C_{n,j}/(n+1) \to u_j - u_{j-1}$ almost surely (a.s.), as $n \to \infty$, for each $j = 1, \ldots, k+1$. Thus on the set A_ε introduced previously

$$
\left\{\left(1 - \frac{s}{n+1}\right)^{n+1}\right\}^{C_{n,j}/(n+1)} \le \exp(-s)^{u_j-u_{j-1}-\varepsilon}. \tag{A.3}
$$

After employing the changes of variable $s = \frac{z}{n}$ and $t = \frac{r}{n}$, we can apply the dominated convergence theorem; it follows that, as $n \to \infty$ and for $|v_j| < 1$, $j = 0, 1, \ldots, k+1$,

$$
\begin{aligned}
I_{j1} &= v_{j-1}(u_j-u_{j-1})\int_0^n \exp\{v_{j-1}(u_j-u_{j-1})z\}\left\{\left(1-\frac{z}{n+1}\right)^{n+1}\right\}^{\frac{C_{n,j}}{n+1}} dz \\
&\to v_{j-1}(u_j-u_{j-1})\int_0^\infty \exp\{(v_{j-1}-1)(u_j-u_{j-1})z\}dz \\
&= \frac{v_{j-1}}{1-v_{j-1}} \qquad \text{a.s.}, \\
I_{j2} &\to \frac{w_j}{1-w_j} \qquad \text{a.s.}, \qquad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
I_{j3} &= v_{j-1} w_j n^2 (u_j - u_{j-1})^2 \\
&\times \int_0^n \int_0^{n(1-r)} \exp\{v_{j-1}(u_j - u_{j-1})z + w_j(u_j - u_{j-1})r\} \left\{ \left(1 - \frac{z+r}{n+1}\right)^{n+1} \right\}^{\frac{C_{n,j}}{n+1}} dzdr \\
&\to \frac{v_{j-1}}{1 - v_{j-1}} \times \frac{w_j}{1 - w_j} \qquad \text{a.s.}
\end{aligned}
$$

After some additional algebra we also have the result that as $n \to \infty$

$$
\prod_{j=1}^{k+1} (1 + I_{j1} + I_{j2} + I_{j3}) \to \prod_{j=1}^{k+1} \frac{1}{1 - v_{j-1}} \times \frac{1}{1 - w_j} = \prod_{j=1}^{k} \frac{1}{1 - v_j} \times \frac{1}{1 - w_j}
$$

almost surely, since we set $v_0 = w_{k+1} = 0$. Moreover, since the formula found in (A.3) also ensures uniform integrability, we also have

$$
\begin{aligned}
&\varphi_{Y_n^U(u_1),\ldots,Y_n^U(u_k),-Y_n^L(u_1),\ldots,-Y_n^L(u_k)}(v_1, \ldots, v_k, w_1, \ldots, w_k) \\
&= \mathbb{E}\left\{ \prod_{j=1}^{k+1} \left(1 + I_{j1} + I_{j2} + I_{j3}\right) \right\} \to \prod_{j=1}^{k} \frac{1}{1 - v_j} \times \frac{1}{1 - w_j},
\end{aligned}
$$

which concludes the proof of Theorem 1. ■

We can recognize Theorem 4 to be a consequence of Theorem 3, by means of a suitable application of the continuous mapping theorem. The detailed proof follows below.

Proof of Theorem 4. For every t, there are $nF_n(t)$ observations in our sample of size n which are smaller than t, and the value of the last observation is exactly equal to

$$
R_n^L(t) = U_{n,n\mathbb{F}_n(t)} = \mathbb{Q}_n(\mathbb{F}_n(t)).
$$

By (A.3) we have $I_n(t) = \frac{n+1}{2} \sum_{i=1}^{nF_n(t)} D_{n,i}^2 + \frac{n+1}{2}\{t - R_n^L(t)\}^2$. Then

$$
\frac{n+1}{2} \sum_{i=1}^{nF_n(t)} D_{n,i}^2 \le I_n(t) \le \frac{n+1}{2} \sum_{i=1}^{(n+1)F_n(t)} D_{n,i}^2.
$$

Since $\lfloor (n+1)F_n(t) \rfloor = nF_n(t)$ for all $0 \le t < U_{n,n}$ and $\lfloor (n+1)F_n(t) \rfloor = nF_n(t) + 1$ for all $U_{n,n} \le t \le 1$, it follows that the difference between $2\sqrt{n+1}\ \{\ I_n(t) - \mathbb{F}_n(t)\ \}$ and $E_n\{\mathbb{F}_n(t)\}$ is uniformly bounded by $(n+1)(\max_i D_{n,i})/2$. By Slud (1978) we have that $\max_i D_{n,i} = O\left(\frac{\log n}{n}\right)$ almost surely as $n \to \infty$, and

$$
2\sqrt{n}\ \{\ I_n^L(t) - \mathbb{F}_n(t)\ \} = E_{n+1}\{\mathbb{F}_n(t)\} + O\left[\left\{ \frac{(\log n)^2}{\sqrt{n}} \right\}\right]
$$

with probability one. Accordingly, $E_n\{\mathbb{F}_n(t)\}$ and $2\sqrt{n+1}\ \{\ I_n(t) - \mathbb{F}_n(t)\ \}$ have the same limit distribution. Now $\mathbb{F}_n(\cdot)$ converges to the identity function id$(\cdot)$ (which of course is deterministic) and $E_n(\cdot)$ converges weakly to $V(\cdot)$. As a result, the pair $(E_n(\cdot), \mathbb{F}_n(\cdot))$ converges weakly in $D^2(0,1)$ to $(V(\cdot), \mathrm{id}(\cdot))$, and by the continuity of the composition map, applying the continuous mapping theorem (Billingsley 1999, p. 151), we can conclude that

$$2\sqrt{n+1}\ \{\ I_n(t) - \mathbb{F}_n(t)\ \} \Rightarrow V(t), \quad t \in [0,1].$$

■

Proof of Theorem 4. We can derive the two-sided maximal probability distribution for $V(t)$ from that of the pinned Brownian motion. Clearly

$$\mathbb{P}(\sup_{0\le t\le 1} |V(t)| < b) = \mathbb{E}\left\{ \mathbb{P}\left(\sup_{0\le t\le 1} \left| B(t) + t\frac{W(1)}{\sqrt{5}} \right| < \frac{b}{2\sqrt{5}} \right) |W(1) \right\}.$$

Based on either Equation 4.12 in Beghin & Orsingher (1999) or Part II, Chapter 1, Equation 1.15.8(1) on p. 174 in Borodin & Salminen (2002), we know that

$$\mathbb{P}(\sup_{0\le t\le 1} |B(t) + ty| < a) = \sum_{h=-\infty}^{\infty} (-1)^h e^{-2ha(ha-y)}.$$

It follows that

$$\begin{aligned}
\mathbb{P}(\sup_{0\le t\le 1} |V(t)| < b) &= \mathbb{E}\left[\sum_{h=-\infty}^{\infty} (-1)^h e^{-2h\frac{b}{2\sqrt{5}}\left\{ h\frac{b}{2\sqrt{5}} - \frac{W(1)}{\sqrt{5}} \right\}} \right] \\
&= \sum_{h=-\infty}^{\infty} (-1)^h e^{-\frac{h^2b^2}{10}} \mathbb{E}\left\{ e^{\frac{hb}{5}W(1)} \right\} \\
&= \sum_{h=-\infty}^{\infty} (-1)^h e^{-\frac{4}{50}h^2b^2}.
\end{aligned}$$

■

ACKNOWLEDGEMENTS

G. Pistone acknowledges the support of Collegio Carlo Alberto, Torino. He is a member of INdAM-GNAMPA. This project received funding from MIUR, the Italian Ministry of Education, University, and Research project DISMA—Dipartimento di Eccellenza. We thank Ivo Stoepker for carefully reading the manuscript and eliminating several important errors.

BIBLIOGRAPHY

Aly, E. (1983). Some limit theorems for uniform and exponential spacings. *The Canadian Journal of Statistics*, 11, 211–219.

Aly, E. (1988). Strong approximations of quadratic sums of uniform spacings. *The Canadian Journal of Statistics*, 16, 201–207.

Beghin, L. & Orsingher, E. (1999). On the maximum of the generalized Brownian bridge. *Lietuvos Matematikos Rinkinys*, 39, 200–213.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. John Wiley & Sons, New York.

Borodin, A. N. & Salminen, P. (2002). *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. Springer-Verlag, Basel.

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

Csörgő, M. (1983). *Quantile Processes with Statistical Applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of Multivariate Analysis*, 28, 20–68.

Moran, P. A. P. (1947). The random division of an interval. *Supplement to the Journal of the Royal Statistical Society*, 9, 92–98.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revuz, D. & Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer-Verlag, Berlin.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Slud, E. (1978). Entropy and maximal spacings for random partitions. *Probability Theory and Related Fields*, 41, 341–352.

Zaitsev, A. Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *European Series in Applied and Industrial Mathematics: Probability and Statistics*, 2, 41–108.

Received 28 April 2017
Accepted 19 July 2018