

Numerical approximations to weighted Kolmogorov-Smirnov Distributions via Integral Equations

Dan Wu, Lin Yee Hin, Nino Kordzakhia and Alexander Novikov

Abstract We show that the distribution of two-sided weighted Kolmogorov-Smirnov (wK-S) statistics can be obtained via the solution of the system of two Volterra type integral equations for corresponding boundary crossing probabilities for a diffusion process. Based on this result we propose a numerical approximation method for evaluating the distribution of wK-S statistics. We provide the numerical solutions to the system of the integral equations which were also verified via Monte Carlo simulations.

1 Introduction

The applications of one-sided and two-sided weighted Kolmogorov-Smirnov (wK-S) statistical tests are ubiquitous in diverse areas of applications, including physics, finance, computational biology and Gene Set Enrichment Analysis (GSEA), see e.g. [8, 14, 22]. In some cases there exist modifications of the wK-S whose limit distributions (for large sample sizes) can be represented as the following random variable

$$D_{g,f} := \sup_{t \in T} \frac{|B_t - g(t)\xi|}{f(t)}, \quad (1.1)$$

Dan Wu (e-mail: dan.wu@uts.edu.au), Lin Hin (e-mail: LinYee.Hin@uts.edu.au)
School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123,
Broadway, NSW 2007, Australia.

Nino Kordzakhia (e-mail: nino.kordzakhia@mq.edu.au)
Department of Statistics, Macquarie University, Balaclava Road, North Ryde, NSW 2109, Australia.

Alexander Novikov (e-mail: Alex.Novikov@uts.edu.au)
School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123,
Broadway, NSW 2007, Australia and Steklov Institute of Mathematics, Gubkina str. 8, 119991
Moscow, Russia.

where $g(t)$ and $f(t)$ are some deterministic functions of t , $B = \{B_t, t \in [0, 1]\}$ is a standard Brownian bridge, the random variable ξ is independent of B and has the standard normal distribution, $\xi \sim N(0, 1)$, $T \subseteq [0, 1]$. Note that analytical expressions for the distribution function $P\{D_{g,f} < x\}$, $x > 0$, are not available in closed form besides the classical case when $g(t) = 0, f(t) = 1$ (see Kolmogorov [15]).

Recent applications of wK-S in GSEA (see e.g. [6]) require the development of fast and accurate numerical approximations for the cumulative distribution functions (cdf) of $D_{g,f}$ for specific functions f and g .

This paper addresses the issues of approximating cdf of the $D_{g,f}$ under the following two important settings:

1. $f(t) = 1, g = \{g(t) = t^\alpha - t, 1/2 < \alpha < 1, t \in T = [0, 1]\}$,
2. $f(t) = \sqrt{t(1-t)}, t \in T = [a, b], g(t) = 0, 0 < a \leq b < 1$.

Our goal is to find accurate numerical approximations for the following corresponding tail distributions

$$P_1(x) := P\{\sup_{t \in T} |B_t - g(t)\xi| > x\}; P_2(x) := P\left\{\sup_{t \in [a,b]} \frac{|B_t|}{\sqrt{t(1-t)}} > x\right\}.$$

Setting 1 was recently discussed in the context of GSEA, see e.g. [6],[16], [17]). The family of functions g is of special relevance there. In particular, the case $\alpha = 2/3$ in Setting 1 corresponds to GSEA analysis where the weights of the genes in question are replaced by their respective ranks obtained based on their expressions in typical experiments. Examples of such gene expression profiles are accessible from the Gene Expression Omnibus repository [12]. Note that $g = 0$ corresponds to the classical Kolmogorov-Smirnov test statistic where the closed-form expression for $P\{D_{0,1} < x\}$ is well known [15]. See also [10] and [21] for a historical account.

Setting 2 corresponds to the wK-S test suggested by Anderson and Darling [1]. It is designed with the purpose to increase the sensitivity for the tails of empirical distributions compared to the classical K-S test. Some asymptotic result for the tail probabilities under this setting have been derived recently in [7].

In general, finding the distribution of $D_{g,f}$ is a computationally intensive numerical problem, a subject of pursuit for many different approaches, each with its own merits and shortcomings, that are devised specifically to address this problem from different perspectives. In [9] and [2], among others, the authors reduce the problem of approximating the extrema of modified Brownian bridges to finding boundary crossing probabilities (BCP) with respect to Brownian motion. In line with this approach, piecewise linear boundaries were used to replace nonlinear boundaries and approximate the desired distribution by an n -dimensional integrals in a similar way as used in [19], [4], and [23]. The convenient feature of this approach, as demonstrated in [16] and [17] for the case of the one-sided version of wK-S, is a possibility to obtain analytical upper and lower bounds for the tail distributions of statistics which lead to fast and reasonably accurate approximations. However, for the case of the two-sided wK-S this approach requires substantial computational cost in situations when highly accurate approximations are needed.

Historically it was Kolmogorov [15] who first found the distribution of $D_{0,1}$ as a solution of a partial differential equation (PDE). Subsequently, Anderson and Darling in [1], had applied this approach for solving related problems in the construction of goodness-of-fit tests. Under Settings 1 and 2 it is possible to use the finite-difference schemes or finite element method to obtain numerical approximations. However, the PDE approach seems to be not computationally efficient due to the fact that not only function evaluations are required at larger number of discretised points, but substantially higher computational burden is incurred due to the element assembly process [5].

The technique discussed in our paper is inspired by the work of Peskir (see Theorem 2.2 in [20]) who derived an integral equation of Volterra type for BCP with one-sided boundaries. In this paper we are expanding this technique for BCP with two-sided boundaries deriving a system of two integral equations of Volterra type. Note that this system of integral equations was derived in a different way by Buonocore et al [3] using a different approach. As a matter of fact our technique is applicable to all regular diffusion processes where transition probabilities are available in a closed form. The advantage of this approach is that a system of integral equations are rather straight forward to obtain for all one-dimensional diffusion processes and efficient numerical techniques can be easily developed to solve the equations. The complete results including the case of general diffusion processes will be presented in another publication.

The paper is organised as follows. In Section 2 we formulate the results which provide a system of two Volterra integral equations for $P\{D_{g,f} < x\}$. In Section 3 we describe the numerical algorithm and results as well as comparisons with Monte Carlo simulation.

2 Construction of a system of integral equations of Volterra type

In this section, we formulate a general result on BCP for two-sided boundaries which will be used for deriving a system of integral equations to evaluate the distribution of $D_{g,f}$.

Note that in Setting 1 we need to find BCP of the following form

$$P_1(x, y) := P\{\sup_{t \in T} |B_t - g(t)y| < x\} = P\{L(t) \leq B_t \leq U(t), t \in [0, 1]\}$$

where

$$L(t) = -x + g(t)y, \quad U(t) = x + g(t)y, \quad x > 0, y \in R.$$

Having the function $P_1(x, y)$ the distribution of two-sided wK-S statistics can be evaluated via integration with respect the density function $\phi(y, \sigma) := \frac{1}{\sigma\sqrt{2\pi}} e^{-(y/\sigma)^2/2}$:

$$P_1(x) = 1 - P\{D_{g,1} < x\} = \int_{-\infty}^{\infty} \phi(y, 1)(1 - P_1(x, y))dy. \quad (2.1)$$

In Setting 2 a similar integral representation holds due to the following relations:

$$\begin{aligned}
P_2(x) &= P\left(\sup_{t \in [a,b]} \frac{|B_t|}{\sqrt{t(1-t)}} > x\right) = P(|B_a| > x\sqrt{a(1-a)}) + \\
&P\left(\{|B_a| \leq x\sqrt{a(1-a)}\} \cap \left\{\sup_{t \in [a,b]} \frac{|B_t|}{\sqrt{t(1-t)}} > x\right\}\right) \\
&= 1 - \operatorname{erf}(x/\sqrt{2}) + \int_{-x\sqrt{1-a}}^{x\sqrt{1-a}} \phi(y, \sqrt{t(1-t)})(1 - P_2(x, y)) dy \quad (2.2)
\end{aligned}$$

with

$$\begin{aligned}
P_2(x, y) &= P\{L(t) \leq B_t \leq U(t), t \in [a, b] | B_a = y\}, y \in \mathbb{R}, \\
L(t) &= -x\sqrt{t(1-t)}, \quad U(t) = x\sqrt{t(1-t)}, \quad x > 0.
\end{aligned}$$

To derive equations for BCP under the general setting for a general diffusion process $X = \{X_t, t \geq 0\}$ (defined on a suitable probability space) we set

$$\tau_L = \inf_{t \geq t_0} \{t : X_t \leq L(t); B_s < U(s), \forall s \in (t_0, t) | X_{t_0} = x_0\}, \quad (2.3)$$

$$\tau_U = \inf_{t \geq t_0} \{t : X_t \geq U(t); X_s > L(s), \forall s \in (t_0, t) | X_{t_0} = x_0\}, \quad (2.4)$$

$$\tau = \inf_{t \geq t_0} \{t : X_t \notin (L(t), U(t)) | X_{t_0} = x_0\} = \inf\{\tau^-, \tau^+\}. \quad (2.5)$$

Let $f_L(t|x_0, t_0)$, $f_U(t|x_0, t_0)$ and $f(t|x_0, t_0)$ be the densities of τ_L , τ_U and τ respectively (subject their existence) and $X_{t_0} = x_0$.

Now we state a modified version of Theorem 2.2 in [20] for the two-sided BCP.

Theorem 1. *Let X be a one-dimensional diffusion process with boundaries U and L being continuously differentiable functions satisfying inequalities $L(s) < x_0 < U(s)$ and $L(t) < U(t)$ for all $t > s$. Let F_U and F_L be the cumulative distribution functions of τ_U and τ_L respectively. The following system of integral equations*

$$P(G_1, t|x_0, s) = \int_s^t P(G_1, t|U(s), s)F_U(ds|x_0, s) + \int_s^t P(G_1, t|L(s), s)F_L(ds|x_0, s), \quad (2.6)$$

$$P(G_2, t|x_0, s) = \int_s^t P(G_2, t|U(s), s)F_U(ds|x_0, s) + \int_s^t P(G_2, t|L(s), s)F_L(ds|x_0, s), \quad (2.7)$$

hold for any measurable sets $G_1 \subseteq [U(t), \infty)$ and $G_2 \subseteq (-\infty, L(t)]$.

The proof of this result will be presented in a full version of this paper; we just mention that it is based on the use of the Chapman-Kolmogorov equation as a starting point.

In both Settings 1 and 2 we need to derive equations for the case when $X = B$ is a standard Brownian bridge. Since B is a Gauss-Markov process we have

$$P(y, t|x, s) \sim N\left(\frac{R(s, t)}{R(s, s)}x, R(t, t) - \frac{R^2(s, t)}{R(s, s)}\right) \quad (2.8)$$

where R is the covariance function of B . Using this representation, upon the substitution of the initial condition $B_{t_0} = x_0$ into the equations (2.6) and (2.7). Letting

$$\Psi(y|x, s) = \Psi\left(\frac{y - \frac{1-t}{1-s}x}{\sqrt{\frac{(t-s)(1-t)}{(1-s)}}}\right), \quad \Phi(y|x, s) = \Phi\left(\frac{y - \frac{1-t}{1-s}x}{\sqrt{\frac{(t-s)(1-t)}{(1-s)}}}\right),$$

we have

$$\Psi(U(t)|x_0, t_0) = \int_{t_0}^t \Psi(U(t)|U(s), s) f_U(s|x_0, t_0) ds + \int_{t_0}^t \Psi(U(t)|L(s), s) f_L(s|x_0, t_0) ds, \quad (2.9)$$

$$\Phi(L(t)|x_0, t_0) = \int_{t_0}^t \Phi(L(t)|U(s), s) f_U(s|x_0, t_0) ds + \int_{t_0}^t \Phi(L(t)|L(s), s) f_L(s|x_0, t_0) ds \quad (2.10)$$

respectively, where $\Phi(x) = \int_{-\infty}^x \phi(z) dz$ and $\Psi(x) = 1 - \Phi(x)$.

Since by assumptions $U(t)$ and $L(t)$ are differentiable, we have

$$\lim_{t \rightarrow s} \Psi(U(t), t|U(s), s) = \lim_{t \rightarrow s} \Phi(L(t), t|L(s), s) = \Psi(0) = \frac{1}{2}, \quad (2.11)$$

$$\lim_{t \rightarrow s} \Psi(U(t), t|L(s), s) = \lim_{t \rightarrow s} \Phi(L(t), t|U(s), s) = \Psi(-\infty) = 0. \quad (2.12)$$

Hence, the kernels $\Psi(\cdot)$ and $\Phi(\cdot)$ are non-singular and are differentiable with respect to t for the case $X = B$.

The system of integral equations (2.6) and (2.7) (and the corresponding equations (2.9) and (2.10 for the case of Brownian bridge) are Volterra equations of the first kind; they can be reduced to Volterra integral equations of the second kind which are numerically more suitable.

Theorem 2. Let $f_U(t|x_0, t_0)$ and $f_L(t|x_0, t_0)$ be the probability density functions of τ_U and τ_L respectively and let $p(y, t|x, s) = \frac{\partial}{\partial t} P(y, t|x, s)$, then

$$\begin{aligned} f_U(t|x_0, t_0) &= 2p(y_1, t|x_0, t_0) - 2 \int_{t_0}^t p(y_1, t|U(s), s) f_U(s|x_0, t_0) ds \\ &\quad - 2 \int_{t_0}^t p(y_1, t|L(s), s) f_L(s|x_0, t_0) ds, \end{aligned} \quad (2.13)$$

$$\begin{aligned} f_L(t|x_0, t_0) &= 2p(y_2, t|x_0, t_0) - 2 \int_{t_0}^t p(y_2, t|U(s), s) f_U(s|x_0, t_0) ds \\ &\quad - 2 \int_{t_0}^t p(y_2, t|L(s), s) f_L(s|x_0, t_0) ds \end{aligned} \quad (2.14)$$

hold for any $y_1 \subseteq [U(t), \infty)$ and $y_2 \subseteq (-\infty, L(t)]$.

Equations (2.13) and (2.14) are obtained by differentiating (2.6) and (2.7) with respect to t and then using the relation

$$\lim_{s \rightarrow t} P(U(t), t | U(s), s) = \lim_{s \rightarrow t} P(L(t), t | L(s), s) = \frac{1}{2}.$$

Note that a similar approach for the one-sided case was used by Fortet [13].

It follows that for a standard Brownian bridge with the initial condition $B_{t_0} = x_0$, we have

$$\begin{aligned} f_U(t|x_0, t_0) &= 2 \frac{\partial \Psi(U(t)|x_0, t_0)}{\partial t} - 2 \int_{t_0}^t \frac{\partial \Psi(U(t)|U(s), s)}{\partial t} f_U(s|x_0, t_0) ds \\ &\quad - 2 \int_{t_0}^t \frac{\partial \Psi(U(t)|L(s), s)}{\partial t} f_L(s|x_0, t_0) ds, \end{aligned} \quad (2.15)$$

and

$$\begin{aligned} f_L(t|x_0, t_0) &= 2 \frac{\partial \Phi(L(t)|x_0, t_0)}{\partial t} - 2 \int_{t_0}^t \frac{\partial \Phi(L(t)|U(s), s)}{\partial t} f_U(s|x_0, t_0) ds \\ &\quad - 2 \int_{t_0}^t \frac{\partial \Phi(L(t)|L(s), s)}{\partial t} f_L(s|x_0, t_0) ds. \end{aligned} \quad (2.16)$$

Due to properties (2.11) and (2.12), singularities in the denominator of the kernels can be removed.

3 Numerical integration procedure for approximation BCP

For Settings 1 and 2 we use (2.1) and (2.2)

$$P_1(x) = \int_{-\infty}^{\infty} \int_0^1 \phi(y, 1) (f_U(t|0, 0) + f_L(t|0, 0)) dy$$

and

$$P_2(x) = \operatorname{erfc}(x/\sqrt{2}) + \int_{-x\sqrt{t_a}}^{x\sqrt{t_a}} \int_a^b \phi(y, \sqrt{t(1-t)}) (f_U(t|y, a) + f_L(t|y, a)) dy$$

with boundaries L and U shown in Section 1 respectively. To calculate f_U and f_L we use (2.15) and (2.16).

Let $t_i = t_0 + ih$, $i = 1, \dots, m$, where h is the time step size of uniform discretisations. We use the Euler approximation to obtain $f_L(t_i) = f_L(t_i|x, s)$ and $f_U(t_i) = f_U(t_i|x, s)$ at an increasing sequence of knots t_i and the appropriate Gaussian quadrature for numerical integrations.

For each of the aforementioned cases, we perform N simulations and use m equally-spaced discretization in the time interval T , and we estimate the tail proba-

bilities

$$p_{f,g;m}(x) := P \left[D_{f,g;m}^{(i)} \geq x \right] \approx \frac{1}{N} \sum_{i=1}^N I(D_{f,g;m}^{(i)} \geq x),$$

where $I(\cdot)$ is an indicator function.

For Setting 1 we calculate $\hat{P}_1(x)$, the tail probabilities approximated using the aforementioned approximation for $p_{f,g;m}(x)$ with $m = 2^{10}$ discretised time steps and the usage of $n = 20$ Gauss-Hermite nodes in the numerical integrations. Table 1 contains the comparisons of $\hat{P}_1(x)$ and $p_{f,g;m}$.

For Setting 2 we calculate $\hat{P}_2(x)$, the tail probabilities approximated using the aforementioned approximation for $p_{f,g;m}(x)$ in this setting, with $m = 2^{10}$ discretised time steps and the usage of $n = 20$ Gauss-Legendre nodes in the numerical integrations. In Table 2 we compare $\hat{P}_2(x)$ to simulation results, and asymptotic estimators $\tilde{P}_2(x)$ generated from [7].

The numerical approximations using integral equations are performed on a Macintosh laptop computer running OS X with 8 GB RAM and 1600MHz CPU, and computer programs are implemented in C++98. Each point of $\hat{p}_g(x)$ and $\hat{p}_t(x)$ in the tables takes 1.92s and 1.87s respectively. The Monte Carlo simulations for all cases are carried out on a cluster computer with 28 parallel CPUs using $N = 10^7$ simulation runs and $m = 2^{19}$ equally-spaced discretization time intervals.

In both settings, tail probabilities estimated using integral equations and simulation approach differ only at the third decimal place and beyond, suggesting that the numerical integration approach delivers a level of accuracy comparable to those of the simulation approach despite the use of a comparatively coarser grain discretization time step, i.e., $m = 2^{10}$ in the former as opposed to $m = 2^{19}$ in the latter.

The major advantage of our method is the relative simplicity and fast calculations compared to other techniques. For approximations using integral equations, since discretization time steps of the order $m = 2^{10}$ is sufficient to deliver tail probability estimates comparable to those of simulation at $N = 10^7$ and $m = 2^{19}$, they can be evaluated in a modest computational framework to reduce costs.

| x | $\hat{P}_1(x)$ | $p_{f,g;m}(x)$ | $Var[p_{f,g;m}(x)] \times 10^7$ | $ \hat{P}_1(x) - p_{f,g;m}(x) $ |
|-----|----------------|----------------|---------------------------------|---------------------------------|
| 0.4 | 0.997467 | 0.997395 | 0.025982 | 0.000072 |
| 1.2 | 0.128036 | 0.127972 | 1.115952 | 0.000064 |
| 2.0 | 0.001056 | 0.001040 | 0.010389 | 0.000016 |
| 2.2 | 0.000219 | 0.000233 | 0.002329 | 0.000014 |

Table 1a Setting 1: estimated tail probabilities $\hat{P}_1(x)$ using integral equations, compared to simulations $p_{f,g;m}(x)$.

| x | $\hat{P}_2(x)$ | $p_{f,gm}(x)$ | $\text{Var}[p_{f,gm}(x)] \times 10^7$ | $\tilde{P}_2(x)$ | $ \hat{P}_2(x) - p_{f,gm}(x) $ | $ \tilde{P}_2(x) - p_{f,gm}(x) $ |
|-----|----------------|---------------|---------------------------------------|------------------|--------------------------------|----------------------------------|
| 1.2 | 0.997822 | 0.997820 | 0.021752 | 0.930700 | 0.000002 | 0.067120 |
| 2.0 | 0.695515 | 0.698015 | 2.107901 | 0.663005 | 0.002500 | 0.035010 |
| 2.8 | 0.175139 | 0.177285 | 1.458550 | 0.192957 | 0.002146 | 0.015672 |
| 3.6 | 0.018395 | 0.018750 | 0.183984 | 0.020708 | 0.000355 | 0.001958 |
| 4.4 | 0.000925 | 0.000969 | 0.009681 | 0.001032 | 0.000044 | 0.000063 |

Table 2b Setting 2: estimated tail probabilities $\hat{P}_2(x)$ using integral equations, compared to simulations $p_{f,gm}(x)$; $\tilde{P}_2(x) = 1 - \tilde{A}(x) (1/\underline{u})^{-\theta_0(x)}$ is the asymptotic tail probability estimator in equations (9), (10) and (13) of Chicheportiche & Bouchaud [7] where $\tilde{A}(x) = (\text{erf}(x/\sqrt{2}))^2$, $\theta_0(x) = \sqrt{2/\pi} x e^{-x^2/2}$, and $1/\underline{u}$ is the sample size which is set to be 100 in this numerical example.

Acknowledgement. The authors are grateful to the International Mathematical Research Institute MATRIX for hosting and supporting the Mathematics of Risk program during which they worked on the results presented in this note.

This work was partially supported by the ARC Discovery grant DP150102758.

References

1. Anderson, T.W., Darling, A.D.: Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193-212 (1952)
2. Bischoff, W., Hashorva, E., Hüslér, J., Miller, F.: Exact asymptotics for boundary crossings of the brownian bridge with trend with application to the Kolmogorov test. *Ann. Inst. Stat. Math.* **55**, 849-864 (2003)
3. Buonocore, A., Giorno, V., Nobile, A., Ricciardi, L.: On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability*, **27** (1), 102–114, 1990
4. Borovkov, K., Novikov, A.: Explicit bounds for approximation rates of boundary crossing probabilities for the Wiener process. *J. Appl. Prob.* **42**, 82-92 (2005)
5. Brebbia, C.A., Dominguez, J.: Boundary element methods for potential problems. *Appl. Math. Modelling*, **1**, 372-378 (1977)
6. Charmpi, K., Ycart, B.: Weighted Kolmogorov-Smirnov testing: an alternative for gene set enrichment analysis. *Stat. Appl. Genet. Mol. Biol.* **14**, 279-293 (2015)
7. Chicheportiche, R., Bouchaud, J.-P.: Weighted Kolmogorov-Smirnov test: Accounting for the tails. *Phys. Rev.* **86**, 041115 (2012).
8. Delande, D., Gay, J.C.: Quantum chaos and statistical properties of energy levels: Numerical study of the hydrogen atom in a magnetic field. *Phys. Rev. Lett.* **57**, (2006) “ Published 20 October 1986; Erratum *Phys. Rev. Lett.* **57**, 2877 (1986)
9. Durbin, J.: Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *J. Appl. Prob.* **8**, 431-453 (1971)
10. Durbin, J.: *Distribution theory for tests based on the sample distribution theory*. SIAM, Philadelphia (1973)
11. Durbin, J. The first-passage density of a continuous gaussian process to a general boundary. *J. Appl. Prob.* **22**, 99-122 (1985)

12. Edgar, R., Domrachev, M., Lash, A.E.: Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210 (2002)
13. Fortet, R. Les fonctions aléatoires du type de markoff associéesa certaines équations linéaires aux dérivées partielles du type parabolique. *J. Math. Pures. Appl.* **22**, 177-243 (1943)
14. Klein, J.P., Logan, B., Harhoff, M., Andersen, P.K.: Analyzing survival curves at a fixed point in time. *Statist. Med.* **26**, 4505-4519 (2007)
15. Kolmogorov, A.N.: Sulla Determinazione Empirica di una Legge di Distribuzione. *Giorn. Ist Ital. degli Attuari.* **4**, 83-91 (1933)
16. Kordzakhia, N., Novikov, A., Ycart, B.: Approximations for weighted Kolmogorov-Smirnov distributions via boundary crossing probabilities. *Stat. Comput.* **27**, 1-11 (2016)
17. Kordzakhia, N., Novikov, A.: Bounds and Approximations for Distributions of Weighted Kolmogorov-Smirnov Tests. *From Statistics to Mathematical Finance*, Springer, Cham, 235-250 (2017)
18. Krapivsky, P.L., Redner, S.: Life and death in an expanding cage and at the edge of a receding cliff. *Amer. J. Phy.* **64**, 546-552 (1996)
19. Novikov, A., Frishling, V., Kordzakhia, N.: Approximations of boundary crossing probabilities for a brownian motion. *J. Appl. Prob.* **36**, 1019-1030 (1999)
20. Peskir, G.: On integral equations arising in the first-passage problem for Brownian motion. *J Int. Equa. Appl.* **14**, 397-423 (2002)
21. Stephens, M.A.: Introduction to Kolmogorov (1933) on the empirical determination of a distribution. In *Breakthroughs in statistics*, Springer, 93-105 (1992)
22. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et.al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceed. Nat. Acad. Sci.* **102**, 15545-15550 (2005)
23. Wang, L., Pötzelberger, K.: Boundary crossing probability for Brownian motion and general boundaries. *J. Appl. Prob.* **34** 54-65 (1997)