MULTICRITERIA OPTIMIZATION

Best Simultaneous Approximation of Functions

Alex Bacopoulos[a] and Dimitra Kouloumpou[b]

[a]*Department of Applied Mathematics and Physical Sciences, National Technical University of Athens, Zografou Campus, 157 80, Athens, Greece, (abacop@central.ntua.gr)*

[b] *Section of Mathematics, Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, Piraeus 18539, Greece, (dimkouloumpou@hna.gr)*

Abstract. The problem we consider is to find (the) **best approximation**(s) to a given function **simultaneously** with respect to more than one criterion of proximity. Questions of existence, characterization, unicity and computation are examined. Examples are given.

Keywords: Best vectorial approximation(s), minimal projection norms, computational schemata

1. INTRODUCTION

Among other formulations of simultaneous approximation, the notion of a "Vectorially Minimal Approximation" is introduced, which is shown to be the **natural setting** for problems of simultaneity, both theoretically as well as computationally. For the above formulations of Multicriteria Optimization we propose 3 types of "models" and show their interrelationships in each "primal" and "dual" spaces. In particular, attention has been given to effective models suitable for **numerical computation.** A related problem situated in the "dual space" of approximation operators is to approximate the (non-linear) **best approximation operator** by projection operators. This approach, as a tool of "good" approximation of functions (in situations to be specified), is motivated by the following inequality, where the role of minimal projections, i.e. $min\|P\|$ is self-explanatory.

$$\|f - Pf\| \le \|I - P\| dist(f, Y) \le (1 + \|P\|) dist(f, Y).$$

Here again the approximation in the operator space is done **simultaneously** with respect to several norms. As just indicated, this reduces to finding "simultaneously" **minimal projection norms**. Examples are given and a "Zero in the Convex Hull" as well as a "Kolmogorov-type" characterization theorems are presented.

The tools used in this presentation are Elementary Optimization Theory, Computational Numerical Analysis and Elementary Functional Analysis.

2. VECTORIAL APPROXIMATION

Let $\|\cdot\|_{\boldsymbol{a}}, \|\cdot\|_{\boldsymbol{b}}$ be two norms defined on a linear space S and let $f \in S \sim K$ be a given function to be approximated by approximation $p \in K \subset S$. K is assumed to be a closed, convex, proper subset of S. Let $G(p) = (|| f - p ||_a , || f - p||_b)$ and define the partial ordering $\trianglerighteq$ on $G(K)$ by

$$G(p) \trianglelefteq G(q) \Leftrightarrow \begin{cases} \|f-p\|_a \leq \|f-q\|_a \\ \quad and \\ \|f-p\|_b \leq \|f-q\|_b \end{cases}$$

We shall write $G(p) \triangleleft G(q)$ if and only if $G(p) \trianglelefteq G(q)$ and $G(p) \neq G(q)$.

Definition 2.1
We say that p is a best vec approximation if there does not exist a $q \in K$ such that $G(q) \triangleleft G(p)$.

Definition 2.2
The minimal set M is given by

$$M = \{G(p) : p \in K \textit{ is a best vec approximation} \}.$$

There are some general geometric facts that are easy to verify. We cite some of them here:

- $\pi_1\big(G(K)\big)$ has zero homotopy group.
- M is a convex, decreasing arc.

Let Λ is the $45°$ bisector of the $\|\cdot\|_{\boldsymbol{a}}, \|\cdot\|_{\boldsymbol{b}}$ orthogonal axes. L is the supporting line to $G(K)$ which makes $135°$ angle with $\|\cdot\|_{\boldsymbol{a}}$ axes.
The proof of the following theorem is a consequence of the definitions, convexity and, in the case of $(P_m) = M \cup L$, the continuity of the best approximation operator. Sum here denotes the sum of two norms. Max means the maximum of two norms.

Theorem 2.1
Let p_s be a best sum approximation. Then $G(P_s) \subseteq M \cup L$. Similarly, if p_m denotes the best max approximation then $G(P_m) = M \cup L$ (assuming $M \cup L \neq \emptyset$).

Furthermore, we define the set D by

$$D = \left\{d \colon \inf_{q \in K} \|f-q\|_a \leq d \leq \inf_{q \in B} \|f-q\|_a \right\}$$

where,

$$B = \left\{r \in K \colon \|f-r\|_b = \inf_{q \in K} \|f-q\|_b \right\}.$$

Theorem 2.2
An element $p \in K$ is a best vectorial approximation if and only if there exists $d \in D$ and $\Phi \in S^*$ satisfying

$$\|\Phi\|_b = 1$$
$$\Phi(f-b) = \|f-q\|_a \leq d$$

and

$$Re\Phi(p-q) \leq 0 \text{ for all } q \in K \text{ satisfying } \|f-q\|_a \leq d.$$

3. VECTORIALY MINIMAL PROJECTIONS

Let $\Lambda = \Lambda(X, V)$ be the space of all linear operators from a real or complex space X into a finite-dimensional subspace V, and let Π be the family of all operators in Λ with a given fixed action on V (e.g., the identity action corresponds to the family of projections onto V). Let X be equipped with norms $\|\cdot\|_i, i = 1,2, \dots, k$. Let X_i denote the normed space given by X with the norm $\|\cdot\|_i$, and define

$$\|x\| := (\|x\|_1, \|x\|_2, \dots, \|x\|_k).$$

Define the partial ordering " $\trianglelefteq$ " on X by

$$\|x\| \trianglelefteq \|z\| \Leftrightarrow \|x\|_i \le \|z\|_i \; for\ every\ i = 1,2, \dots, k.$$

We write $\|x\| \vartriangleleft \|z\|$ if and only if $\|x\| \trianglelefteq \|z\|$ and $\|x\| \neq \|z\|$.

Definition 3.1

For $Q \in \Lambda$, let $\|Q\|_i$ denote the operator norm on X_i, let $\|Q\| := \left(\|Q\|_1, \|Q\|_2, \dots, \|Q\|_k\right)$ and define the partial ordering " $\trianglelefteq$ " on Λ by

$$\|P\| \trianglelefteq \|Q\| \Leftrightarrow \|P\|_i \le \|Q\|_i \; for\ every\ i = 1,2, \dots, k.$$

We write $\|P\| \vartriangleleft \|Q\|$ if and only if $\|P\| \trianglelefteq \|Q\|$ and $\|P\| \neq \|Q\|$.
P is a vectorially minimal operator in Π if there no exist $Q \in \Pi$ such that $\|Q\| \vartriangleleft \|P\|$.

Notation

The minimal set M is given by

$$M := \{\|P\| : P \in \Pi\ is\ a\ vectorially\ minimal\ operator\ in\ \Pi\}.$$

Definition 3.2

For $i = 1,2, \dots, k$ $(x, y) \in S(X_i^{**}) \times S(X_i^{*})$ will be called an extremal pair for $Q \in \Lambda$, if $\langle Q_i^{**} x, y\rangle = \|Q\|_i$, where $Q_i^{**} : X_i^{**} \longrightarrow V$ is the second adjoint extension of Q to X_i^{**}.
(S denotes the unit sphere).

Notation

Let $E(Q)$ be the set of all extremal pairs for Q. To each $(x, y) \in Q$ associate the rank-one operator $y \otimes x$ from X_i to X_i^{**} given by $(y \otimes x)(z) = \langle z, y\rangle x$ for $\in X_i$, where i is the subscript associated with (x, y).

Theorem 3.1 (Characterization)

P has vectorially minimal norm in Π **if and only if** the closed convex hull of $\{y \otimes x\}_{(x,y)\in E(P)}$ contains an operator E_P for which V is an invariant subspace, i.e.

$$\boldsymbol{E_P} = \int_{\boldsymbol{E}(\boldsymbol{P})} y \otimes x \, d\mu(x, y) : V \longrightarrow V.$$

Theorem 3.2

P has vectorially minimal norm in Π **if and only if** there does not exist
$D \in \Delta = \{D \in \Lambda : D = 0\ in\ V\}$ such that

$$\sup_{(x,y)\in E(P)} Re\langle P_i^{**} x, y\rangle \overline{\langle D^{**} x, y\rangle} < 0.$$

4. SOME SPECIAL CASES

We give some examples of **Theorem 2.2**. In the notation of this theorem, let $S = C[a,b]$, $K = \Pi_n[a,b]$ (the set of polynomials on $[a,b]$ of degree less than or equal to n), $\|\cdot\|$ is the supremum norm on $[a,b]$ and $w_1, w_2 \in C[a,b]$ two (weight) functions, positive and continuous on $[a,b]$.
We introduce extreme points, for a given $f \in C[a,b]$ to be approximated, in connection with the next theorem, as follows:

$$\begin{aligned}
\overline{\underline{X}}_{+1} &= \{x \in [a,b]: w_1(x)\big(f(x) - p(x)\big) = +\|w_1(f-p)\|\} \\
\overline{\underline{X}}_{+2} &= \{x \in [a,b]: w_2(x)\big(f(x) - p(x)\big) = +\|w_2(f-p)\|\} \\
\overline{\underline{X}}_{-1} &= \{x \in [a,b]: w_1(x)\big(f(x) - p(x)\big) = -\|w_1(f-p)\|\} \\
\overline{\underline{X}}_{-2} &= \{x \in [a,b]: w_2(x)\big(f(x) - p(x)\big) = -\|w_2(f-p)\|\}.
\end{aligned}$$

$$\overline{\underline{X}}_p = \overline{\underline{X}}_{+1} \cup \overline{\underline{X}}_{+2} \cup \overline{\underline{X}}_{-1} \cup \overline{\underline{X}}_{-2}$$

The sign function $\sigma(x)$ on $\overline{\underline{X}}_p$ is defined by

$$\sigma(x) = -1 \; when \; x \in \overline{\underline{X}}_{-1} \cup \overline{\underline{X}}_{-2}$$

and

$$\sigma(x) = +1 \; when \; x \in \overline{\underline{X}}_{+1} \cup \overline{\underline{X}}_{+2}.$$

Theorem 4.1 (Application)
Consider the Vectorial Chebyshev optimization, with w_1 and w_2 as defined above. Then p is a best vec approximation to f if and only if there exist $n+2$ points $x_1 < x_2 < \cdots < x_{n+2} \in \overline{\underline{X}}_p \subset [a,b]$ satisfying

$$\sigma(x_i) = (-1)^{i+1}\sigma(x_1) \;\; for \; every \; i = 1,2,\dots,n+2.$$

Theorem 4.2
Each best vec approximation is unique; i.e. given $\mu \in M$ there is only one $p \in \Pi_n[a,b]$ such that $G(p) = \mu$.

Note that this uniqueness does not contradict the fact that the minimal set M has, in general, an infinite number of points, each of which corresponds to a (unique) best vectorial approximation. Likewise, the easily shown existence of M proves the existence of best solutions.

Theorem 4.3 (Application)
Let $X = C[a,b], K = \Pi_n[a,b], \|\cdot\|_a, \|\cdot\|_b$ the sup and L_2 norms on $C[a,b]$ which we denote by $\|\cdot\|_\infty$ and $\|\cdot\|_2$ respectively.
Find the best vectorial approximation p_d whose error in Chebyshev norm equals a prescribed value $d \in P^+$, $\| f - p_1 \|_\infty \le d \le \| f - p_2 \|_\infty$. It is clear that the desired polynomial p_d is the unique solution to the problem

$$\min_{p\in\Pi_n} \|f - p\|_2$$

subject to

$$\| f - p\|_\infty \le d.$$

Since the number of constraints here is infinite, we proceed by solving a sequence of quadratic programming problems, each with a finite number of constraints. The sequence of solutions $\{p_k\}$ is shown to converge to the theoretical solution p_d.

Algorithm Corresponding to Theorem 4.3
At the $k - th$ step we have from the preceding steps a finite set of points $X^k \subset [a,b]$. We solve the quadratic program

$$\min_{p\in\Pi_n} \|f - p\|_2$$

subject to

$$\| f(x) - p(x)\|_\infty \le d, x \in X_k.$$

Denoting by p_k the solution of this problem, we calculate a point $x_k \in [a,b]$ such that

$$|f(x_k) - p_k(x_k)| = \| f - p_k \|_\infty.$$

We form $X^{k+1} = X^k \cap \{x_k\}$ and proceed to the next cycle. At the beginning X^1 may be an arbitrary finite set, containing a maximum of $|f(x) - p_L(x)|$.

REFERENCES

1. S.M. Lozinski, On a class of linear operators, Dokl. Akad. Nauk SSSR 64**,** 1948, pp. 193-196.
1. E.W. Cheney, C.R. Hobby, P.D. Morris, F. Schurer and D.E. Wulbert, On the minimal property of the Fourier projection, Trans. Am. Math. Soc. 143**,** 1969, pp. 249-258.
2. E.W. Cheney and K.H Price, Minimal projections, in Approximation Theory, Proceedings of Symposium, Lancaster 1969, New York 1971, pp. 261-289.
3. W. Odyniec and G. Lewicki, Minimal projections in Banach spaces, Springer-Verlag Lecture Notes in Mathematics, 1449.
5. D.H Hyers, G. Isac and Th.M. Rassias, Topics in nonlinear analysis and applications, World Scientific, 1997.
6. B. Brosowski and A.Da Silva, Scalarization of vectorial relations applied to certain optimization problems, Note Mat. 11 ,1991, pp. 69-91.
7. B.L. Chalmers and F.T. Metcalf, A characterization and equations for minimal projections and extensions, J. Operator Theory 32 ,1994, pp. 31-46.
8. B.L. Chalmers and F.T. Metcalf, minimal projections and extensions for compact Abelian groups, in sl Approximation Theory VI, C.K. Chui et al., eds., Academic Press, 1989, pp. 129-132.
9. A. Bacopoulos and B.L. Chalmers, Vectorially minimal projections, in Approximation Theory *III* Vol. 1: Approximation and Interpolation, World Scientific, C.K. Chui end L.L. Schumaker, eds. 1995, pp. 15-22.
10. A. Bacopoulos and C.A Botsaris, An efficient blend of the first Remes algorithm with quadratic programming, J. Math. Anal. Appls. 174, No. 2 ,1993, pp. 342-358.
11. A. Bacopoulos, Computational n-D Optimization, BOOK in progress. Suggested to A. Bacopoulos by Richard Bellman (available R.B.'s letter)

The Ramanujan Journal
https://doi.org/10.1007/s11139-021-00542-4

Limit behavior of the q-Pólya urn

Dimitris Cheliotis[1] · **Dimitra Kouloumpou**[2]

Received: 1 September 2021 / Accepted: 23 November 2021

Abstract

The q-Pólya urn is a q-analog of the Pólya urn and is a model of ball extraction from an urn with balls of two colors, A and B. Balls of color B have priority to be picked over those of color A. We prove that, in an infinite sequence of extractions, almost surely, the number of balls of color A that are picked has a finite limit and we identify its distribution. Then we prove functional limit theorems for the number of balls of color A extracted. The limit is either a pure birth process or a diffusion, depending on the initial composition of the urn. Finally, we discuss basic results for the q-Pólya urn with more than two colors.

Keywords Pólya urn · q-Pólya urn · q-Calculus · Functional limit theorems

Mathematics Subject Classification Primary 60F17 · Secondary 60K99 · 60C05

1 Introduction and results

The Pólya urn This is the model where in an urn that has initially a finite number of white and black balls we draw, successively and uniformly at random, a ball from it and then we return the ball back together with k balls of the same color as the one drawn. The number $k \in \mathbb{N}^+$ is fixed.

The first author was supported by the Hellenic Foundation for Research and Innovation (H. F. R. I.) under the "First Call for H. F. R. I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 1034).

✉ Dimitris Cheliotis
dcheliotis@math.uoa.gr

Dimitra Kouloumpou
dimkouloumpou@hna.gr

[1] Department of Mathematics, National and Kapodistrian University of Athens, 15784 Panepistimiopolis, Athens, Greece

[2] Section of Mathematics, Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, 18539 Piraeus, Greece

Published online: 04 February 2022

Standard references for the theory and the applications of Pólya urn and related models are [14,17].

The q-Pólya urn This is a q-analog of the Pólya urn (see [10,15] for more on q-analogs) introduced in [16] and studied further in [4] (see also [5]).

A q-analog of a mathematical object A is another object $A(q)$ so that when $q \to 1$, $A(q)$ "tends" to A. Take $q \in (0, \infty)\backslash\{1\}$. The q-analog of any $x \in \mathbb{C}$ is defined as

$$[x]_q := \frac{q^x - 1}{q - 1}. \tag{1}$$

Note that $\lim_{q\to 1}[x]_q = x$.

Now consider an urn that contains a finite number of white and black balls. We perform a sequence of additions of balls to the urn according to the following rule. If at a given time the urn contains A_1 white and A_2 black balls ($A_1, A_2 \in \mathbb{N}$, $A_1 + A_2 > 0$), then we add k white balls with probability

$$\mathbf{P}_q(\text{white}) = \frac{[A_1]_q}{[A_1 + A_2]_q}. \tag{2}$$

Otherwise, we add k black balls, and this has probability

$$\mathbf{P}_q(\text{black}) = 1 - \mathbf{P}_q(\text{white}) = q^{A_1}\frac{[A_2]_q}{[A_1 + A_2]_q}. \tag{3}$$

This stochastic process we call q-Pólya urn. To understand how it works, it helps to realize the probabilities $\mathbf{P}_q(\text{white})$, $\mathbf{P}_q(\text{black})$ through the following experiment.

If $q \in (0, 1)$, then we put the balls in a line with the A_1 white coming first and the A_2 black following. To pick a ball, we go through the line, starting from the beginning and picking each ball with probability $1-q$ independently of what happened with the previous balls. If we finish the line without picking a ball, we start from the beginning. Once we pick a ball, we return it to its position together with k balls of the same color. Given these rules, the probability of picking a white ball is

$$(1 - q^{A_1})\sum_{j=0}^{\infty}(q^{A_1+A_2})^j = \frac{1 - q^{A_1}}{1 - q^{A_1+A_2}} = \frac{[A_1]_q}{[A_1 + A_2]_q}, \tag{4}$$

which is (2), because before picking a white ball, we will go through the entire list a random number of times, say j, without picking any ball and then, going through the white balls, we pick one (probability $1 - q^{A_1}$).

If $q > 1$, we place in the line first the black balls and we go through the list picking each ball with probability $1 - q^{-1}$. According to the above computation, the probability of picking a black ball is

$$\frac{[A_2]_{q^{-1}}}{[A_1 + A_2]_{q^{-1}}} = q^{A_1} \frac{[A_2]_q}{[A_1 + A_2]_q},$$

which is (3).

We extend the notion of drawing a ball from a q-Pólya urn to the case where exactly one of A_1, A_2 is infinity. Then the probability to pick a white (resp. black) ball is determined again by (2) (resp. (3)), where this is understood as the limit of the right hand side as A_1 or A_2 goes to ∞. For example, assuming that $A_1 = \infty$ and $A_2 \in \mathbb{N}$, we have $\mathbf{P}_q(\text{white}) = 1$ if $q < 1$ and $\mathbf{P}_q(\text{white}) = q^{-A_2}$ if $q > 1$. Again these probabilities are realized through the experiment described above. Thus, we can run the process even if we start with an infinite number of balls from one color and finite from the other.

Consider now a q-Pólya urn having $A_1(0), A_2(0)$ white and black balls, respectively, and start an infinite sequence of drawings. For $n \in \mathbb{N}^+$, denote by $A_1(n), A_2(n)$ the numbers of white and black balls, respectively, after n drawings.
We want to study two aspects of the asymptotic behavior of the sequence $\{A_1(n)\}_{n\in\mathbb{N}}$.

(1) The first concerns the limit, in any sense, of $A_1(n)$ properly normalized. In the Pólya urn, if we keep the same notation, the following convergence in distribution is a well-known fact: $\frac{A_1(n)}{A_1(n)+A_2(n)} \xrightarrow{d} \text{Beta}\,(A_1(0)/k, A_2(0)/k)$ as $n \to \infty$. For the q–Pólya urn, things are less exciting. If $q > 1$, after some point, we will be drawing only black balls, and consequently $A_1(n)$ becomes eventually (a random) constant $A_1(\infty)$. We identify the distribution of $A_1(\infty)$. By the above discussion, this answers the case $q \in (0, 1)$ too. Then, it is $A_2(n)$ that becomes eventually constant.

(2) The second concerns the entire path $\{A_1(n)\}_{n\in\mathbb{N}}$. Is it possible, by applying appropriate, natural transformations, to get convergence to a stochastic process? That is, an analogous result to Donsker's theorem for simple symmetric random walk in $\mathbb{Z}$. For the Pólya urn, this question has been investigated in the works [3,7] .

The results concerning these two points are exhibited in the following two subsections.

1.1 Basic results for the *q*-Pólya urn

We recall some notation from q-calculus (see [5,15]). For $q \in (0,\infty)\backslash\{1\}$, $x \in \mathbb{C}$, $k \in \mathbb{N}^+$, we define

$$[x]_q := \frac{q^x - 1}{q-1} \qquad \text{the } q\text{-number of } x, \tag{5}$$

$$[k]_q! := [k]_q [k-1]_q \cdots [1]_q \qquad \text{the } q\text{-factorial,} \tag{6}$$

$$[x]_{k,q} := [x]_q [x-1]_q \cdots [x-k+1]_q \qquad \text{the } q\text{-factorial of order } k, \tag{7}$$

$$\begin{bmatrix} x \\ k \end{bmatrix}_q := \frac{[x]_{k,q}}{[k]_q!} \qquad \text{the } q\text{-binomial coefficient,} \tag{8}$$

$$(x;q)_\infty := \prod_{i=0}^{\infty}(1-xq^i) \text{ when } q \in [0,1) \qquad \text{the } q\text{-Pochhammer symbol.} \tag{9}$$

We extend these definitions in the case $k=0$ by letting $[0]_q! = 1$, $[x]_{0,q} = 1$.

Now consider a q-Pólya urn that has initially A_1 white and A_2 black balls, where $A_1 \in \mathbb{N} \cup \{\infty\}$ and $A_2 \in \mathbb{N}$. Call $H_1(n)$ the number of drawings that give white ball in the first n drawings. Its distribution is specified by the following.

Fact 1 Let $\hat{A}_1 := A_1/k$ and $\hat{A}_2 := A_2/k$.

(i) If $A_1 \in \mathbb{N}$, then the probability mass function of $H_1(n)$ is

$$\mathbf{P}(H_1(n)=x) = q^{k(n-x)(\hat{A}_1+x)} \frac{\begin{bmatrix}-\hat{A}_1\\ x\end{bmatrix}_{q^{-k}} \begin{bmatrix}-\hat{A}_2\\ n-x\end{bmatrix}_{q^{-k}}}{\begin{bmatrix}-\hat{A}_1-\hat{A}_2\\ n\end{bmatrix}_{q^{-k}}} \tag{10}$$

$$= q^{-A_2 x} \frac{\begin{bmatrix}\hat{A}_1+x-1\\ x\end{bmatrix}_{q^{-k}} \begin{bmatrix}\hat{A}_2+n-x-1\\ n-x\end{bmatrix}_{q^{-k}}}{\begin{bmatrix}\hat{A}_1+\hat{A}_2+n-1\\ n\end{bmatrix}_{q^{-k}}} \tag{11}$$

$$= q^{-kx(\hat{A}_2+n-x)} \frac{\begin{bmatrix}-\hat{A}_1\\ x\end{bmatrix}_{q^{k}} \begin{bmatrix}-\hat{A}_2\\ n-x\end{bmatrix}_{q^{k}}}{\begin{bmatrix}-\hat{A}_1-\hat{A}_2\\ n\end{bmatrix}_{q^{k}}} \tag{12}$$

for all $x \in \{0,1,\ldots,n\}$.

(ii) If $A_1 = \infty$ and $q > 1$, then the probability mass function of $H_1(n)$ is

$$\mathbf{P}(H_1(n)=x) = \begin{bmatrix}n\\ x\end{bmatrix}_{q^{-k}} q^{-A_2 x} \prod_{j=1}^{n-x}(1-q^{-A_2}(q^{-k})^{j-1}) \tag{13}$$

for all $x \in \{0,1,\ldots,n\}$. This is the probability mass function of the q-binomial distribution of the second kind with parameters n, q^{-A_2}, q^{-k} (see Theorem 3.2 in [5]). If $A_1 = \infty$ and $q \in (0,1)$, then $\mathbf{P}(H_1(n)=n) = 1$ obviously.

Relation (10) is (3.2) in [4], where it is proved through recursion. In Sect. 2 we give an alternative proof.

According to the experiment described in Sect. 1, the balls that are placed first in the line have an advantage to be picked (the white if $q \in (0,1)$, the black if $q > 1$). In fact, this leads to the extinction of drawings from the balls of the other color; there is a point after which the number of balls in the urn of that color stays fixed to a random number. In the next theorem, we identify the distribution of this number. We treat the case $q > 1$.

Theorem 1 (Extinction of the second color) *Assume that* $q > 1$, $A_1 \in \mathbb{N} \cup \{\infty\}$, $A_2 \in \mathbb{N}$. *With probability one, as* $n \to \infty$, $\{H_1(n)\}_{n \geq 1}$ *converges to a random variable* $H_1(\infty)$ *with values in* $\mathbb{N}$ *and probability mass function*

(i)

$$f(x)=\begin{bmatrix}\frac{A_1}{k}+x-1\\ x\end{bmatrix}_{q^{-k}} q^{-A_2x}\frac{(q^{-A_2};q^{-k})_\infty}{(q^{-A_1-A_2};q^{-k})_\infty} \tag{14}$$

for all $x\in\mathbb{N}$ in the case $A_1\in\mathbb{N}$ and

(ii)

$$f(x)=\left(\frac{q^{-A_2}}{1-q^{-k}}\right)^x\frac{1}{[x]_{q^{-k}}!}(q^{-A_2};q^{-k})_\infty \tag{15}$$

for all $x\in\mathbb{N}$ in the case $A_1=\infty$.

When $A_1\in\mathbb{N}$, the random variable $H_1(\infty)$ has the negative q-binomial distribution of the second kind with parameters $A_1/k, q^{-A_2}, q^{-k}$. We recall here that for $\nu\in(0,\infty)$, $\theta\in(0,1)$, and $q\in(0,1)$, the function $f:\mathbb{R}\to[0,\infty)$ with

$$f(x)=\begin{bmatrix}\nu+x-1\\ x\end{bmatrix}_q\theta^x\frac{(\theta;q)_\infty}{(\theta q^\nu;q)_\infty} \tag{16}$$

for $x\in\mathbb{N}$ and $f(x)=0$ for $x\in\mathbb{R}\backslash\mathbb{N}$ defines the probability mass function of a distribution with support $\mathbb{N}$ (f sums to 1 due to the q-binomial theorem, relation (1.3.2) in [10]). We call this distribution negative q-binomial of the second kind with parameters ν,θ,q (see §3.1 of [5]). When $\nu\in\mathbb{N}^+$, formula (16) simplifies to

$$f(x)=\begin{bmatrix}\nu+x-1\\ x\end{bmatrix}_q\theta^x\prod_{j=1}^{\nu}(1-\theta q^{j-1}). \tag{17}$$

When $A_1=\infty$, $H_1(\infty)$ has the Euler distribution with parameters $q^{-A_2}/(1-q^{-k}), q^{-k}$ (see §3.3 in [5] again).

1.2 Functional scaling limits

Consider a q-Pólya urn whose initial composition depends on $m\in\mathbb{N}^+$. That is, it has $A_1^{(m)}(0)$, $A_2^{(m)}(0)$ white and black balls, respectively. Start an infinite sequence of drawings and for $n\in\mathbb{N}^+$, denote by $A_1^{(m)}(n)$, $A_2^{(m)}(n)$ the numbers of white and black balls, respectively, after n drawings.

To see a new process arising out of the path of $\{A_1^{(m)}(n)\}_{n\geq 0}$, we start with an initial number of balls that tends to infinity as $m\to\infty$. We assume that $A_2^{(m)}(0)$ grows linearly with m. Regarding $A_1^{(m)}(0)$, we study three regimes:

(a) $A_1^{(m)}(0)$ stays fixed with m.
(b) $A_1^{(m)}(0)$ grows to infinity but sublinearly with m.
(c) $A_1^{(m)}(0)$ grows linearly with m.

The regime where $A_1^{(m)}(0)$ grows superlinearly with m follows from regime b) by changing the roles of the two colors. We remark on this after Theorem 3.

The other parameter that we have to tune is q. If q is kept fixed, then:

(i) if $q > 1$, then nothing interesting happens because the assumption $\lim_{m\to\infty} A_2^{(m)}(0) = \infty$ implies that the process $\{A_1^{(m)}(n)\}_{n\geq 0}$ converges (as $m \to \infty$) to the one that never increases (we always pick a black ball) and

(ii) if $q < 1$, then in the scenario $\lim_{m\to\infty} A_1^{(m)}(0) = \infty$ the situation is analogous to (i) while in the scenario that $A_1^{(m)}(0)$ stays fixed with m the process $\{A_1^{(m)}(n)\}_{n\geq 0}$ converges (as $m \to \infty$) to the q-Polya urn with $A_2 = \infty$.

Interesting limits appear once we take $q = q_m$ to depend on m and approach 1 as $m \to \infty$. We study the case that $q_m > 1$ and the distance of q_m from 1 is $\Theta(1/m)$ and remark on the case that the distance is $o(1/m)$.

In the regimes (a) and (b), the scarcity of white balls has as a result that the time between two consecutive drawings of a white ball is large. We expect then that speeding up time by an appropriate factor we will see a birth process. And indeed this is the case as our first two theorems show.

All processes appearing in this work with index set $[0, \infty)$ and values in some Euclidean space $\mathbb{R}^d$ are elements of $D_{\mathbb{R}^d}[0, \infty)$, the space of functions $f : [0, \infty) \to \mathbb{R}^d$ that are right continuous and have limits from the left at each point of $[0, \infty)$. This space is endowed with the Skorokhod topology (defined in §5 of Chapter 3 of [9]), and convergence in distribution of processes with values on that space is defined through that topology.

We remind the reader that the negative binomial distribution with parameters $\nu \in (0, \infty)$ and $p \in (0, 1)$ is the distribution with support in $\mathbb{N}$ and probability mass function

$$f(x) = \binom{\nu + x - 1}{x} p^{\nu}(1-p)^x \tag{18}$$

for all $x \in \mathbb{N}$. When $\nu \in \mathbb{N}^+$, this is the distribution of the number of failures until we obtain the ν-th success in a sequence of independent trials, each having probability of success p. For a random variable X with this distribution, we write $X \sim NB(\nu, p)$.

In all results of this subsection we assume that the parameter of the urn is $q_m = c^{1/m}$ with $c > 1$.

Theorem 2 *Fix $w_0 \in \mathbb{N}^+$ and $b > 0$. If $A_1^{(m)}(0) = w_0$ for all $m \in \mathbb{N}^+$ and $\lim_{m\to\infty} A_2^{(m)}(0)/m = b$, then, as $m \to \infty$, the process $(k^{-1}\{A_1^{(m)}([mt]) - A_1^{(m)}(0)\})_{t\geq 0}$ converges in distribution to an inhomogeneous in time pure birth process Z with $Z(0) = 0$ and such that for all $0 \leq t_1 < t_2$, $j \in \mathbb{N}$, the random variable*

$$Z(t_2) - Z(t_1)|Z(t_1) = j \text{ has distribution } NB\Big(\frac{w_0}{k} + j, \frac{1 - c^{-b-kt_1}}{1 - c^{-b-kt_1}}\Big).$$

In particular, Z has rates

$$\lambda_{t,j} = \frac{w_0 + jk}{c^{b+kt} - 1} \log c \tag{19}$$

for all $(t, j) \in [0, \infty) \times \mathbb{N}$.

Theorem 3 *Assume that $A_1^{(m)}(0) = g_m$ and $\lim_{m\to\infty} A_2^{(m)}(0)/m = b$, where $b \in (0, \infty)$ and $g_m \in \mathbb{N}^+$, $g_m \to \infty$, $g_m = o(m)$ as $m \to \infty$. Then, as $m \to \infty$, the process*

$$(k^{-1}\{A_1^{(m)}([tm/g_m]) - A_1^{(m)}(0)\})_{t\geq 0}$$

converges in distribution to the Poisson process on $[0, \infty)$ with rate

$$\frac{\log c}{c^b - 1}. \tag{20}$$

We return to the discussion at the beginning of the subsection. The regime where $\lim_{m\to\infty} A_2^{(m)}(0)/m = b > 0$ and $A_1^{(m)}(0)/m \to \infty$ is covered by the previous theorem. We need to change the roles of the colors and remark that the role of m as a scaling parameter is played now by $a_m := A_1^{(m)}(0)$. The result that we obtain is that in the q-Pólya urn with $q_m := c^{1/a_m}$ and $c > 1$, the process

$$\frac{1}{k}\left(A_2^{(m)}([ta_m/(bm)]) - A_2^{(m)}(0)\right)_{t\geq 0}$$

converges in distribution, as $m \to \infty$, to the Poisson process on $[0, \infty)$ with rate $(\log c)/(c - 1)$.

Theorem 4 *Assume that $A_1^{(m)}(0), A_2^{(m)}(0)$ are such that $\lim_{m\to\infty} A_1^{(m)}(0)/m = a$, $\lim_{m\to\infty} A_2^{(m)}(0)/m = b$, where $a, b \in [0, \infty)$ are not both zero. Then, as $m \to +\infty$, the process $\left(A_1^{(m)}([mt])/m\right)_{t\geq 0}$ converges in distribution to the unique solution of the differential equation*

$$X_0 = a, \tag{21}$$

$$\mathrm{d}X_t = k\frac{1 - c^{X_t}}{1 - c^{a+b+kt}}\,\mathrm{d}t, \tag{22}$$

which is

$$X_t := a - \frac{1}{\log c} \log\left(\frac{c^b - 1 + c^{-kt}(1 - c^{-a})}{c^b - c^{-a}}\right) \tag{23}$$

for all $t \geq 0$.

Next, we determine the fluctuations of the process $(A_1^{(m)}([mt])/m)_{t\geq 0}$ around its $m \to \infty$ limit, X. Let

$$C_t^{(m)} = \sqrt{m}\left(\frac{A_1^{(m)}([mt])}{m} - X_t\right) \tag{24}$$

for all $m \in \mathbb{N}^+$ and $t \geq 0$.

Theorem 5 *Let $a, b \in [0, \infty)$, not both zero, $\theta_1, \theta_2 \in \mathbb{R}$, and assume that $A_1^{(m)}(0) := [am + \theta_1\sqrt{m}]$, $A_2^{(m)}(0) = [bm + \theta_2\sqrt{m}]$ for all large $m \in \mathbb{N}$. Then, as $m \to +\infty$, the process $(C_t^{(m)})_{t\geq 0}$ converges in distribution to the unique solution of the stochastic differential equation*

$$\begin{aligned} Y_0 &= \theta_1, \\ dY_t &= \frac{k \log c}{c^{a+b+kt} - 1}\left\{\frac{(c^{a+b} - 1)Y_t - c^b(c^a - 1)(\theta_1 + \theta_2)}{c^b - 1 + c^{-kt}(1 - c^{-a})}\right\} \mathrm{d}t \\ &\quad + k\sqrt{(c^a - 1)(c^b - 1)}\frac{c^{(a+kt)/2}}{c^{a+b+kt} - c^{a+kt} + c^a - 1}\, \mathrm{d}W_t, \end{aligned} \tag{25}$$

which is

$$\begin{aligned} Y_t =& \frac{c^{a+b+kt} - 1}{c^{a+b+kt} - c^{a+kt} + c^a - 1}\Bigg(\theta_1 - (\theta_1 + \theta_2)\frac{c^{a+b}(c^a - 1)}{c^{a+b} - 1}\frac{c^{kt} - 1}{c^{a+b+kt} - 1} \\ &+ k\sqrt{(c^a - 1)(c^b - 1)}\int_0^t \frac{c^{(a+kt)/2}}{c^{a+b+kt} - 1}\, \mathrm{d}W_s\Bigg) \end{aligned} \tag{26}$$

for all $t \geq 0$. W is a standard Brownian motion

Remark If we assume that $q = q(m) := c^{\varepsilon_m/m}$ where $c > 1$ and $\varepsilon_m \to 0^+$ as $m \to \infty$, then $q = 1 + o(m^{-1})$. With computations analogous to those of the results of the previous subsection, it is easy to see that the limits of the processes considered in all theorems of this subsection coincide with those in the case of the plain Pólya urn (i.e., when $q = 1$), which are described in the work [7]. Of course, in (24), the role of X_t will be played by the limit one gets from the analogous to Theorem 4.

1.3 *q*-Pólya urn with many colors

In this paragraph, we give a q-analog for the Pólya urn with more than two colors. The way to do the generalization is inspired by the experiment we used in order to explain relation (2).

Let $l \in \mathbb{N}, l \geq 2$, and $q \in (0, 1)$. Assume that we have an urn containing A_i balls of color i for each $i \in \{1, 2, \ldots, l\}$. To draw a ball from the urn, we do the following. We order the balls in a line, first those of color 1, then those of color 2, and so on. Then we visit the balls, one after the other, in the order that they have been placed, and we select each with probability $1 - q$ independently of what happened with the previous balls. If we go through all balls without picking any, we repeat the same procedure starting from the beginning of the line. Once a ball is selected, the drawing is completed. We return the ball to its position together with another k of the same color. For each $i = 0, 1, \dots, l$, let $s_i = \sum_{1 \le j \le i} A_j$. Notice that s_l is the total number of balls in the urn. Then, working as for (4), we see that

$$\mathbf{P}(\text{color } i \text{ is drawn}) = q^{s_{i-1}} \frac{1-q^{A_i}}{1-q^{s_l}} = \frac{q^{s_{i-1}} - q^{s_i}}{1-q^{s_l}} = q^{s_{i-1}} \frac{[A_i]_q}{[s_l]_q}. \tag{27}$$

Call p_i the number in the last display for all $i = 1, 2, \dots, l$. Note that when $q \to 1$, p_i converges to A_i/s_l, which is the probability for the usual Pólya urn with l colors. It is clear that for any given $q \in (0, \infty)\backslash\{1\}$, the numbers $p_1, p_2, \dots, p_l$ are non-negative and add to 1 (the second fraction in (27) shows this). We define then for this q the q-Pólya urn with colors $1, 2, \dots, l$ to be the sequential procedure in which, at each step, we add k balls of a color picked randomly among $\{1, 2, \dots, l\}$ so that the probability that this color is i is p_i .

When $q > 1$, these probabilities come out of the experiment described above but in which we place the balls in reverse order (that is, first those of color l, then those of color $l - 1$, and so on) and we go through the list selecting each ball with probability $1 - q^{-1}$. It is then easy to see that the probability to pick a ball of color i is p_i.

Theorem 6 *Assume that $q \in (0, 1)$ and that we start with $A_1, A_2, \dots, A_l$ balls from colors 1, 2, $\dots$, l respectively, where $A_1, A_2, \dots, A_l \in \mathbb{N}$ are not all zero. Call $H_i(n)$ the number of times in the first n drawings that we picked color i. The probability mass function for the vector $(H_2(n), H_3(n), \dots, H_l(n))$ is*

$$\mathbf{P}\left(H_2(n) = x_2, \dots, H_l(n) = x_l\right) = q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1}(A_j + k x_j)} \frac{\prod_{i=1}^{l} \left[\begin{smallmatrix} -\frac{A_i}{k} \\ x_i \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} -\frac{A_1+A_2 \dots + A_l}{k} \\ n \end{smallmatrix}\right]_{q^{-k}}} \tag{28}$$

$$= \begin{bmatrix} n \\ x_1, x_2, \dots, x_l \end{bmatrix}_{q^{-k}} \frac{q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1}(A_j + k x_j)} \prod_{i=1}^{l} \left[-\frac{A_i}{k}\right]_{x_i, q^{-k}}}{\left[-\frac{A_1+A_2+\dots+A_l}{k}\right]_{n, q^{-k}}} \tag{29}$$

for all $x_2, \dots, x_l \in \{0, 1, 2, \dots, n\}$ with $x_2 + \cdot + x_l \le n$, where $x_1 := n - \sum_{i=2}^{l} x_i$ and $\left[\begin{smallmatrix} n \\ x_1, x_2, \dots, x_l \end{smallmatrix}\right]_{q^{-k}} := \frac{[n]_{q^{-k}}!}{[x_1]_{q^{-k}}! \cdot \dots \cdot [x_l]_{q^{-k}}!}$ is the q-multinomial coefficient.

This theorem has also been derived in [6] (Theorem 3.1 of that work) with a different proof than ours, based on a recursion relation.

It follows from Theorem 1 that when $q \in (0, 1)$, after some random time, we will be picking only balls of color 1. So that the number of times, say H_i, that we pick color i, where $i = 2, 3, \dots, l$, is finite. The next theorem identifies the joint distribution of $H_2, H_3, \dots, H_l$.

Theorem 7 *Under the assumptions of Theorem 6, with probability one, as $n \to +\infty$, the vector $(H_2(n), H_3(n), \dots, H_l(n))$ converges to a random vector $(H_2(\infty), H_3(\infty),$*

$\ldots, H_l(\infty))$ *with values in* $\mathbb{N}^{l-1}$ *and probability mass function*

$$f\,(x_2, x_3, \ldots, x_l) = q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1} A_j} \prod_{i=2}^{l} \begin{bmatrix} x_i + \frac{A_i}{k} - 1 \\ x_i \end{bmatrix}_{q^k} \frac{(q^{A_1}; q^k)_\infty}{(q^{A_1+\cdots+A_l}; q^k)_\infty} \tag{30}$$

for all $x_2, \ldots, x_l \in \mathbb{N}$.

Note that the random variables $H_2(\infty), \ldots, H_l(\infty)$ are independent although $(H_2(n), H_3(n), \ldots, H_l(n))$ are dependent.

Next, we look for a scaling limit for the path of the process. For each $m \in \mathbb{N}^+$, we consider a q-Pólya urn with initial composition $(A_1^{(m)}(0), A_2^{(m)}(0), \ldots, A_l^{(m)}(0))$ and q_m that will be specified below. Let $A_i^{(m)}(j)$ be the number of balls of color i in this urn after j drawings.

Theorem 8 *Assume that* $c \in (0, 1)$, $q_m = c^{1/m}$ *for all* $m \in \mathbb{N}^+$, *and*

$$\frac{1}{m}\left(A_1^{(m)}(0), A_2^{(m)}(0), \ldots, A_l^{(m)}(0)\right) \overset{m\to\infty}{\to} (a_1, a_2, \ldots, a_l)\,,$$

where $a_1, \ldots, a_l \in [0, \infty)$ *are not all zero. Set* $\sigma_0 = 0$ *and* $\sigma_i := \sum_{j\le i} a_j$ *for all* $i = 1, 2, \ldots, l$. *Then the process* $\frac{1}{m}\left(A_1^{(m)}([mt]), A_2^{(m)}([mt]), \ldots, A_l^{(m)}([mt])\right)_{t\ge 0}$ *converges in distribution, as* $m \to +\infty$, *to* $(X_{t,1}, X_{t,2}, \ldots, X_{t,l})_{t\ge 0}$ *with*

$$X_{t,i} = a_i + \frac{1}{\log c} \log \frac{(1 - c^{\sigma_l + kt}) - c^{\sigma_{i-1}}(1 - c^{kt})}{(1 - c^{\sigma_l + kt}) - c^{\sigma_i}(1 - c^{kt})} \tag{31}$$

for all $i = 1, 2, \ldots, l$.

Theorem 9 *Assume that* $c \in (0, 1)$, $q_m = c^{\varepsilon_m/m}$ *for all* $m \in \mathbb{N}^+$ *with* $\lim_{m\to\infty} \varepsilon_m = 0$, *and*

$$\frac{1}{m}\left(A_{0,1}^{(m)}, A_{0,2}^{(m)}, \ldots, A_{0,l}^{(m)}\right) \overset{m\to\infty}{\to} (a_1, a_2, \ldots, a_l)\,,$$

where $a_1, \ldots, a_l \in [0, \infty)$ *are not all zero. Then the process* $\frac{1}{m}\big(A_{[mt],1}^{(m)}, A_{[mt],2}^{(m)}, \ldots, A_{[mt],l}^{(m)}\big)_{t\ge 0}$ *converges in distribution, as* $m \to +\infty$, *to* $(X_t)_{t\ge 0}$ *with*

$$X_t = \left(1 + \frac{kt}{a_1 + \cdots + a_l}\right)(a_1, a_2, \ldots, a_l) \tag{32}$$

for all $t \ge 0$.

Remark Discussing this preprint with Prof. Ch. Charalambides, we were informed that he considered this q-Pólya urn with many colors in a work that was then in progress and now has appeared [6]. That work studies other aspects of the urn, and the only common result with the present work is Theorem 6.

Orientation. In Sect. 2, we prove Fact 1 and Theorem 1, which are basic results for the q-Pólya urn. Section 3 (Sect. 4) contains the proofs of the theorems that give convergence to a jump process (to a continuous process). Finally, Sect. 5 contains the proofs for the results that refer to the q-Pólya urn with arbitrary, finite number of colors.

2 Prevalence of a single color

In this section, we prove the claims of Sect. 1.1. Before doing so, we mention three properties of the q-binomial coefficient. For all $q \in (0,\infty)\setminus\{1\}$, $x \in \mathbb{C}$, $n, k \in \mathbb{N}$ with $k \le n$ it holds

$$[-x]_q = -q^{-x}[x]_q, \tag{33}$$

$$\begin{bmatrix} -x \\ k \end{bmatrix}_q = (-1)^k q^{-k(k+2x-1)/2} \begin{bmatrix} x+k-1 \\ k \end{bmatrix}_q, \tag{34}$$

$$\begin{bmatrix} x \\ k \end{bmatrix}_{q^{-1}} = q^{-k(x-k)} \begin{bmatrix} x \\ k \end{bmatrix}_q, \tag{35}$$

$$\sum_{1\le i_1<i_2<\cdots<i_k\le n} q^{i_1+i_2+\cdots+i_k} = q^{\binom{k+1}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_q. \tag{36}$$

The first is trivial, the second follows from the first, the third is easily shown, while the last is Theorem 6.1 in [15].

Proof of Fact 1 (i) The probability to get black balls exactly at the drawings $i_1 < i_2 < \cdots < i_{n-x}$ is

$$g(i_1, i_2, \ldots, i_{n-x}) = \frac{\prod_{j=0}^{x-1}[A_1 + jk]_q \prod_{j=0}^{n-x-1}[A_2 + jk]_q}{\prod_{j=0}^{n-1}[A_1 + A_2 + jk]_q} q^{\sum_{\nu=1}^{n-x}\{A_1+(i_\nu-\nu)k\}}. \tag{37}$$

To see this, note that, due to (2) and (3), the required probability would be equal to the above fraction if in (3) the term q^{A_1} were absent. This term appears whenever we draw a black ball. Now, when we draw the ν-th black ball, there are $A_1 + (i_\nu - \nu)k$ white balls in the urn, and this explains the exponent of q in (37).

Since $[x + jk]_q = \frac{1-q^{x+jk}}{1-q} = [-\frac{x}{k} - j]_{q^{-k}}[-k]_q$ for all $x, j \in \mathbb{R}$, the fraction in (37) equals

$$\frac{[-\hat{A}_1]_{x,q^{-k}}[-\hat{A}_2]_{n-x,q^{-k}}}{[-\hat{A}_1 - \hat{A}_2]_{n,q^{-k}}}. \tag{38}$$

Then

$$\sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} q^{\sum_{\nu=1}^{n-x} A_1+(i_\nu-\nu)k} \tag{39}$$

$$= q^{(n-x)A_1-k(n-x)(n-x+1)/2} \sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} (q^k)^{i_1+i_2+\cdots+i_{n-x}} \tag{40}$$

$$= q^{(n-x)A_1-k(n-x)(n-x+1)/2} q^{k\binom{n-x+1}{2}} \begin{bmatrix} n \\ x \end{bmatrix}_{q^k} \tag{41}$$

$$= q^{(n-x)A_1} q^{kx(n-x)} \begin{bmatrix} n \\ x \end{bmatrix}_{q^{-k}} = q^{k(n-x)(\hat{A}_1+x)} \begin{bmatrix} n \\ x \end{bmatrix}_{q^{-k}}. \tag{42}$$

The second equality follows from (36) and the equality $\left[\begin{smallmatrix} n \\ x \end{smallmatrix}\right]_{q^k} = \left[\begin{smallmatrix} n \\ n-x \end{smallmatrix}\right]_{q^k}$. The third, from (35). Thus, employing (8) too, we obtain that the sum $\sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} g(i_1, i_2, \dots, i_{n-x})$ equals the right hand side of (10). Then (11) and (12) follow by using (34) and (35) respectively.

(ii) In this scenario, we take $A_1 \to \infty$ in (11). We will explain shortly why this gives the probability we want. Since $q^{-k} \in (0, 1)$, we have $\lim_{t\to\infty}[t]_{q^{-k}} = (1-q^{-k})^{-1}$ and thus, for each $\nu \in \mathbb{N}$, it holds

$$\lim_{t\to\infty} \begin{bmatrix} t+\nu-1 \\ \nu \end{bmatrix}_{q^{-k}} = \frac{1}{[\nu]_{q^{-k}}!} \frac{1}{(1-q^{-k})^{\nu}}. \tag{43}$$

Applying this twice in (11) (there $\hat{A}_1 = A_1/k \to \infty$), we get as limit

$$q^{-A_2 x} \begin{bmatrix} \hat{A}_2+n-x-1 \\ n-x \end{bmatrix}_{q^{-k}} \frac{[n]_{q^{-k}}!(1-q^{-k})^{n-x}}{[x]_{q^{-k}}!} \tag{44}$$

$$= \begin{bmatrix} n \\ x \end{bmatrix}_{q^{-k}} q^{-A_2 x}(1-q^{-k})^{n-x}[\hat{A}_2+n-x-1]_{n-x,q^{-k}}, \tag{45}$$

which equals the right hand side of (13).

Now, to justify that passage to the limit $A_1 \to \infty$ in (11) gives the required result, we argue as follows. For clarity, denote the probability $\mathbf{P}_q(\text{white})$ when there are w white and b black balls in the urn by $\mathbf{P}_q^{w,b}(\text{white})$. And when there are A_1 white and A_2 black balls in the urn in the beginning of the procedure, denote the probability of the event $H_1(n) = x$ by $\mathbf{P}^{A_1,A_2}(H_1(n) = x)$. It is clear that the probability $\mathbf{P}^{A_1,A_2}(H_1(n) = x)$ is a continuous function (in fact, a polynomial) of the quantities

$$\mathbf{P}_q^{A_1+ki,A_2+kj}(\text{white}) : i = 0, 1, \dots, x-1, \quad j = 0, 1, \dots, n-x-1,$$

for all values of $A_1 \in \mathbb{N}\cup\{\infty\}$, $A_2 \in \mathbb{N}$. In $\mathbf{P}^{\infty,A_1}(H_1(n) = x)$, each such quantity, $\mathbf{P}_q^{\infty,m}(\text{white})$, equals $\lim_{A_1\to\infty} \mathbf{P}^{A_1,m}(\text{white})$.

Thus $\mathbf{P}^{\infty,A_2}(H_1(n) = x) = \lim_{A_1\to\infty} \mathbf{P}^{A_1,A_2}(H_1(n) = x)$. □

Before proving Theorem 1, we give a simple argument that shows that eventually we will be picking only black balls. That is, the number $H_1(\infty) := \lim_{n\to\infty} H_1(n)$ of white balls drawn in an infinite sequence of drawings is finite. It is enough to show it in the case that $A_1 = \infty$ and $A_2 = 1$ since, by the experiment that realizes the q-Pólya urn, we have (using the notation from the proof of Fact 1 (ii))

$$\mathbf{P}^{A_1,A_2}(H_1(\infty) = \infty) \leq \mathbf{P}^{\infty,1}(H_1(\infty) = \infty).$$

For each $n \in \mathbb{N}^+$, call E_n the event that at the n-th drawing we pick a white ball, B_n the number of black balls present in the urn after that drawing (also, $B_0 := 1$), and write $\hat{q} := 1/q$. Then $\mathbf{P}(E_n) = \mathbf{E}(\mathbf{P}(E_n|B_{n-1})) = \mathbf{E}(\hat{q}^{B_{n-1}})$. We will show that this decays exponentially with n. Indeed, since at every drawing there is probability at least $1-\hat{q}$ to pick a black ball, we can construct in a common probability space the random variables $(B_n)_{n\geq 1}$ and $(Y_i)_{i\geq 1}$ so that the Y_i are i.i.d. with $Y_1 \sim \text{Bernoulli}(1-\hat{q})$ and $B_n \geq 1 + k(Y_1 + \cdots + Y_n)$ for all $n \in \mathbb{N}^+$. Consequently,

$$\mathbf{P}(E_n) \leq \mathbf{E}(\hat{q}^{1+k(Y_1+\cdots+X_{n-1})}) = \hat{q}\{\mathbf{E}(\hat{q}^{kY_1})\}^{n-1}.$$

This implies that $\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty$, and the first Borel–Cantelli lemma gives that $\mathbf{P}^{\infty,1}(H_1(\infty) = \infty) = 0$.

Proof of Theorem 1 Since $\{H_1(n)\}_{n\geq 1}$ is increasing, it converges to a random variable $H_1(\infty)$ with values in $\mathbb{N}\cup\{\infty\}$. In particular, it converges to this variable in distribution. Our aim is to take the limit as $n \to \infty$ in (11) and in (13) in order to determine the distribution of H_1. Note that for $a \in \mathbb{R}$ and $\theta \in [0, 1)$ it is immediate that (recall (9) for the notation)

$$\lim_{n\to\infty} \begin{bmatrix} a+n \\ n \end{bmatrix}_{\theta} = \frac{(\theta^{a+1};\theta)_\infty}{(\theta;\theta)_\infty}. \tag{46}$$

(i) Taking $n \to \infty$ in (11) and using (46), we get the required expression, (14), for f. Then relation (2.2) in [4] (or (8.1) in [15]) shows that $\sum_{x\in\mathbb{N}} f(x) = 1$, so that it is a probability mass function of a random variable H_1 with values in $\mathbb{N}$.

(ii) This follows after taking limit in (13) and using (46) and $\lim_{n\to\infty}(1 - q^{-k})^n [n]_{q^{-k}}! = (q^{-k}; q^{-k})_\infty$. □

3 Jump process limits. Proof of Theorems 2, 3

In the case of Theorem 2, we let $g_m := 1$ for all $m \in \mathbb{N}^+$, and for both theorems we let $v := v_m := m/g_m$. Our interest is in the sequence of the processes $(Z^{(m)})_{m\geq 1}$ with

$$Z^{(m)}(t) = \frac{1}{k}\left\{A_1^{(m)}([vt]) - A_1^{(m)}(0)\right\} \tag{47}$$

for all $t \geq 0$.

To show convergence in distribution, according to Theorem 7.8 of Chapter 3 of [9], it is enough to show that the sequence $(Z^{(m)})_{m\geq 1}$ is tight and its finite dimensional distributions converge. The description of the limiting process is obtained on the way.

An easy argument shows that tightness follows from the convergence of the finite dimensional distributions because each $Z^{(m)}$ has non decreasing paths. It thus remains to establish the convergence of the finite dimensional distributions.

Notation For sequences $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ with values in $\mathbb{R}$, we will say that they are asymptotically equivalent, and will write $a_n \sim b_n$ as $n \to \infty$, if $\lim_{n\to\infty} a_n/b_n = 1$. We use the same expressions for functions f, g defined in a neighborhood of ∞ and satisfy $\lim_{x\to\infty} f(x)/g(x) = 1$.

3.1 Convergence of finite dimensional distributions

By definition, $Z^{(m)}(0) = 0 = Z(0)$ for all $m \in \mathbb{N}^+$.

Since for each $m \geq 1$ the process $Z^{(m)}$ is Markov taking values in $\mathbb{N}$ and non decreasing in time, it is enough to show that the conditional probability

$$\mathbf{P}(Z^{(m)}(t_2) = k_2 | Z^{(m)}(t_1) = k_1) \tag{48}$$

converges as $m \to \infty$ for each $0 \leq t_1 < t_2$ and non-negative integers $k_1 \leq k_2$.
Define

$$n := [vt_2] - [vt_1], \tag{49}$$
$$x := k_2 - k_1, \tag{50}$$
$$\sigma := \frac{A_1^{(m)}(0) + kk_1}{k}, \tag{51}$$
$$\tau := \frac{k[vt_1] - kk_1 + A_2^{(m)}(0)}{k}, \tag{52}$$
$$r := q_m^{-k} = c^{-k/m}. \tag{53}$$

Then, the probability in (48), with the help of (11), is computed as

$$r^{\tau x}\begin{bmatrix}\sigma + x - 1\\ x\end{bmatrix}_r \frac{\begin{bmatrix}\tau+n-x-1\\ n-x\end{bmatrix}_r}{\begin{bmatrix}\sigma+\tau+n-1\\ n\end{bmatrix}_r}$$
$$= r^{\tau x}\begin{bmatrix}\sigma + x - 1\\ x\end{bmatrix}_r \left(\prod_{i=n-x+1}^{n}(1-r^i)\right)\frac{1}{\prod_{i=n-x}^{n-1}(1-r^{\tau+i})}\frac{[\tau+n-1]_{n,r}}{[\sigma+\tau+n-1]_{n,r}}. \tag{54}$$

The last ratio is

$$\prod_{i=0}^{n-1}\frac{1-r^{\tau+i}}{1-r^{\sigma+\tau+i}} = \prod_{i=0}^{n-1}\left(1-(1-r^{\sigma})r^{\tau}\frac{r^i}{1-r^{\sigma+\tau+i}}\right). \tag{55}$$

Denote by $1-a_{m,i}$ the i-th term of the product. The logarithm of the product equals

$$-(1-r^{\sigma})r^{\tau}\sum_{i=0}^{n-1}\frac{r^i}{1-r^{\sigma+\tau+i}}+o(1) \tag{56}$$

as $m\to\infty$. To justify this, note that $1-r^{\sigma}\sim\frac{1}{m}(A_0^{(m)}+kk_1)\log c$ and $r^{\tau+i}/(1-r^{\sigma+\tau+i})\leq 1/(1-c^{-b})$ for all $i\in\mathbb{N}$. Thus, for all large m, $|a_{m,i}|<1/2$ for all $i=0,1,\ldots,n-1$, and the error in approximating the logarithm of $1-a_{m,i}$ by $-a_{m,i}$ is at most $|a_{m,i}|^2$ (by Taylor's expansion, we have $|\log(1-y)+y|\leq|y|^2$ for all y with $|y|\leq 1/2$). The sum of all errors is at most $n\max_{0\leq i<n}|a_{m,i}|^2$, which goes to zero as $m\to\infty$ because $1-r^{\sigma}\sim C/n$ for some appropriate constant $C>0$.

We will compute the limit of (54) as $m\to\infty$ under the assumptions of Theorems 2, 3.

Proof (**The computation for Theorem 2**) As $m\to\infty$, the first term of the product in (54) converges to $c^{-x(b+kt_1)}$. The q-binomial coefficient converges to $\binom{k^{-1}w_0+k_2-1}{k_2-k_1}$. The third term converges to $(1-c^{-k(t_2-t_1)})^x$, while the denominator of the fourth term converges to $(1-\rho_2)^x$, where we set $\rho_i:=c^{-b-kt_i}$ for $i=1,2$. The expression preceding $o(1)$ in (56) is asymptotically equivalent to

$$-\frac{k}{m}\sigma(\log c)\rho_1\sum_{i=0}^{n-1}\frac{c^{-ki/m}}{1-r^{\sigma+\tau}c^{-ki/m}} \tag{57}$$

$$=-\rho_1 k\sigma(\log c)\frac{1}{m}\sum_{i=0}^{n-1}\frac{c^{-ki/m}}{1-\rho_1 c^{-ki/m}}+o(1) \tag{58}$$

$$=-\rho_1 k\sigma\log c\int_0^{t_2-t_1}\frac{1}{c^{ky}-\rho_1}\,\mathrm{d}y+o(1)=\sigma\log\frac{1-\rho_1}{1-\rho_2}+o(1). \tag{59}$$

The first equality is true because $\lim_{m\to\infty}r^{\sigma+\tau}=\rho_1$ and the function $x\mapsto c^{-ki/m}/(1-xc^{-ki/m})$ has derivative bounded uniformly in i,m when x is confined to a compact subset of $[0,1)$. Thus, the limit of (54), as $m\to\infty$, is

$$\binom{\sigma+x-1}{x}\left(\frac{\rho_1-\rho_2}{1-\rho_2}\right)^x\left(\frac{1-\rho_1}{1-\rho_2}\right)^{\sigma}, \tag{60}$$

which means that, as $m\to\infty$, the distribution of $\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1)=k_1$ converges to the negative binomial distribution with parameters σ, $(1-\rho_1)/(1-\rho_2)$. $\square$

Proof (**The computation for Theorem 3**) Now the term $r^{\tau x}$ converges to c^{-xb}, while

$$\begin{bmatrix}\sigma+x-1\\ x\end{bmatrix}_r\left(\prod_{i=n-x+1}^{n}(1-r^i)\right)=\frac{\prod_{i=0}^{x-1}(1-r^{\sigma+i})}{\prod_{i=1}^{x}(1-r^i)}\left(\prod_{i=n-x+1}^{n}(1-r^i)\right) \tag{61}$$

$$\sim \frac{\prod_{i=0}^{x-1}(\sigma+i)}{\prod_{i=1}^{x} i} \frac{((t_2-t_1)k\log c)^x}{g_m^x} \sim \frac{1}{x!}((t_2-t_1)\log c)^x. \tag{62}$$

The denominator of the fourth term in (54) converges to $(1-c^{-b})^x$. The expression in (56) is asymptotically equivalent to

$$\begin{aligned}-r^{\tau}(1-r^{\sigma})\sum_{i=0}^{n-1}\frac{r^i}{1-r^{\sigma+\tau+i}} &\sim -c^{-b}\frac{g_m}{m}\log c\frac{n}{1-c^{-b}}\\ &\sim -\frac{\log c}{c^b-1}(t_2-t_1).\end{aligned} \tag{63}$$

In the first $\sim$, we used the fact that the terms of the sum, as $m \to \infty$, converge uniformly in i to $(1-c^{-b})^{-1}$. Thus, the limit of (54), as $m \to \infty$, is

$$\frac{1}{x!}\left(\frac{\log c}{c^b-1}(t_2-t_1)\right)^x \mathrm{e}^{-\frac{\log c}{c^b-1}(t_2-t_1)}, \tag{64}$$

which means that, as $m \to \infty$, the distribution of $\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1)=k_1$ converges to the Poisson distribution with parameter $\frac{t_2-t_1}{c^b-1}\log c$. $\square$

3.2 Conclusion

It is clear from the form of the finite dimensional distributions that in both Theorems 2, 3 the limiting process Z is a pure birth process that does not explode in finite time. Its rate at the point $(t,j) \in [0,\infty)\times\mathbb{N}$ is

$$\lambda_{t,j} = \lim_{h\to 0^+}\frac{1}{h}\mathbf{P}(Z(t+h)=j+1|Z(t)=j)$$

and is found as stated in the statement of each theorem.

4 Deterministic and diffusion limits. Proof of Theorems 4, 5

These theorems are proved with the use of Theorem 7.1 in Chapter 8 of [8], which is concerned with convergence of time-homogeneous Markov chains to diffusions. The chains whose convergence is of interest to us are time inhomogeneous, but we reduce their study to the time-homogeneous setting by considering for each such chain $\{Z_n\}_{n\in\mathbb{N}}$ the time-homogeneous chain $\{(Z_n,n)\}_{n\in\mathbb{N}}$. The following consequence of the aforementioned theorem suffices for our purposes.

Corollary 1 *Assume that for each $m \in \mathbb{N}^+$, $(Z_n^{(m)})_{n\in\mathbb{N}}$ is a Markov chain in $\mathbb{R}$. For each $m \in \mathbb{N}^+$ and $n \in \mathbb{N}$, let $\Delta Z_n^{(m)} := Z_{n+1}^{(m)} - Z_n^{(m)}$ and*

$$\mu^{(m)}(x,n) := m\boldsymbol{E}(\Delta Z_n^{(m)}\boldsymbol{1}_{|\Delta Z_n^{(m)}|\le 1}|Z_n^{(m)}=x), \tag{65}$$

$$a^{(m)}(x,n) := m\boldsymbol{E}((\Delta Z_n^{(m)})^2 \boldsymbol{1}_{|\Delta Z_n^{(m)}|\le 1} | Z_n^{(m)} = x) \tag{66}$$

for all $x \in \mathbb{R}$ *with* $\mathbf{P}(Z_n^{(m)} = x) > 0$. *Also, for* $R > 0$ *and for the same* m, n *as above, let* $A(m,n,R) := \{(x,n) : |x| \le R, n/m \le R, \mathbf{P}(Z_n^{(m)} = x) > 0\}$.

Assume that there are continuous functions $\mu, a : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ *and* $x_0 \in \mathbb{R}$ *so that:*
For every $R, \varepsilon > 0$, *it holds*

(i) $\sup_{(x,n)\in A(m,n,R)} |\mu^{(m)}(x,n) - \mu(x,n/m)| \to 0$ *as* $m \to \infty$.
(ii) $\sup_{(x,n)\in A(m,n,R)} |a^{(m)}(x,n) - a(x,n/m)| \to 0$ *as* $m \to \infty$.
(iii) $\sup_{(x,n)\in A(m,n,R)} m\mathbf{P}(|\Delta Z_n^{(m)}| \ge \varepsilon | Z_n^{(m)} = x) \to 0$ *as* $m \to \infty$.

And also

(iv) $Z_0^{(m)} \to x_0$ *as* $m \to \infty$ *with probability 1.*
(v) *For each* $x \in \mathbb{R}$, *the stochastic differential equation*

$$\begin{aligned} \mathrm{d}Z_t &= \mu(Z_t,t)\,\mathrm{d}t + \sqrt{a(Z_t,t)}\,\mathrm{d}B_t, \\ Z_0 &= x, \end{aligned} \tag{67}$$

where B *is a one dimensional Brownian motion, has a weak solution which is unique in distribution.*

Then, as $m \to \infty$, *the process* $(Z_{[mt]}^{(m)})_{t\ge 0}$ *converges in distribution to the weak solution of* (67) *with* $x = x_0$.

Proof For each $m \in \mathbb{N}^+$, we consider the process $Y_n^{(m)} := (Z_n^{(m)}, n/m)$, $n \in \mathbb{N}$, which is a time-homogeneous Markov chain with values in $\mathbb{R}^2$, and we apply Theorem 7.1 in Chapter 8 of [8] Conditions (i), (ii), (iii) of that theorem follow from our conditions (ii), (i), (iii), respectively, while condition (A) there translates to the requirement that the martingale problem for the functions μ and $\sqrt{a}$ is well posed, and this follows from condition (v). $\square$

The tool we will use in checking that condition (v) of the corollary is satisfied is the well known existence and uniqueness theorem for strong solutions of SDEs which requires that for all $T > 0$, the coefficients $\mu(x,t)$, $\sqrt{a(x,t)}$ are Lipschitz in x uniformly for $t \in [0,T]$ and $\sup_{t\in[0,T]}\{|\mu(0,t)| + a(0,t)\} < \infty$ (e.g., Theorem 2.9 of Chapter 5 or [8]). The same conditions imply uniqueness in distribution.

4.1 Proof of Theorem 4

We will apply Corollary 1. For each $m \in \mathbb{N}^+$, consider the Markov chain $Z_n^{(m)} = \frac{A_1^{(m)}(n)}{m}$, $n \in \mathbb{N}$. From any given state x of $Z_n^{(m)}$, the chain moves to either of $x + km^{-1}$, x with corresponding probabilities $p(x,n,m)$, $1 - p(x,n,m)$, where

$$p(x,n,m) := \frac{1 - q_m^{mx}}{1 - q_m^{A_1^{(m)}(0) + A_2^{(m)}(0) + kn}}.$$

In particular, for any $\varepsilon > 0$, is holds $|\Delta Z_n^{(m)}| < 1 \wedge \varepsilon$ for m large enough. Thus, condition (iii) of the corollary is satisfied trivially. Also, for large m, with the notation of the corollary, we have

$$\mu^{(m)}(x, n) = kp(x, n, m), \tag{68}$$

$$a^{(m)}(x, n) = \frac{k}{m} p(x, n, m). \tag{69}$$

And it is easy to see that conditions (i), (ii) are satisfied by the functions a, μ with $a(x, t) = 0$ and $\mu(x, t) = kp(x, t)$ where

$$p(x, t) := \frac{1 - c^x}{1 - c^{a+b+kt}}. \tag{70}$$

Now, for each $x \in \mathbb{R}$, the equation

$$\begin{aligned} \mathrm{d}Z_t &= kp(Z_t, t)\,\mathrm{d}t, \\ Z_0 &= x \end{aligned} \tag{71}$$

has a unique solution. Thus, Corollary 1 applies. In fact, (71) is a separable ordinary differential equation and its unique solution is the one given in the statement of the theorem.

4.2 Proof of Theorem 5

For each $m \in \mathbb{N}^+$, consider the Markov chain

$$Z_n^{(m)} = \sqrt{m}\Big(\frac{A_1^{(m)}(n)}{m} - X_{n/m}\Big), \quad n \in \mathbb{N}.$$

From any given state x of $Z_n^{(m)}$, the chain moves to either of

$$x + km^{-1/2} + \sqrt{m}(X_{n/m} - X_{(n+1)/m}), \tag{72}$$

$$x + \sqrt{m}(X_{n/m} - X_{(n+1)/m}) \tag{73}$$

with corresponding probabilities $p(x, n, m)$, $1 - p(x, n, m)$, where

$$p(x, n, m) = \frac{[A_1^{(m)}(n)]_{q_m}}{[A_1^{(m)}(0) + A_2^{(m)}(0) + kn]_{q_m}} \tag{74}$$

and

$$A_1^{(m)}(n) = mX_{n/m} + x\sqrt{m}, \tag{75}$$

$$A_2^{(m)}(n) = A_1^{(m)}(0) + A_2^{(m)}(0) + kn - A_1^{(m)}(n). \tag{76}$$

For convenience, let $\Delta X_{n/m} = X_{(n+1)/m} - X_{n/m}$. We compute

$$\mathbf{E}\left[\Delta Z_n^{(m)} | Z_n^{(m)} = x\right] = \frac{k}{\sqrt{m}} p(x,n,m) - \sqrt{m}\Delta X_{n/m}, \tag{77}$$

$$\mathbf{E}\left[(\Delta Z_n^{(m)})^2 | Z_n^{(m)} = x\right] = \left(\frac{k^2}{m} - 2k\Delta X_{n/m}\right) p(x,n,m) + m(\Delta X_{n/m})^2. \tag{78}$$

The asymptotics of these expectations are as follows.

Claim Fix $R > 0$. For n such that $\tau := n/m \le R$ and as $m \to \infty$, we have

(a) $$\mathbf{E}\left[\Delta Z_n^{(m)} | Z_n^{(m)} = x\right] = \frac{1}{m} \frac{k \log c}{c^{a+b+k\tau} - 1} \left(c^{X_\tau} x - \frac{(c^{X_\tau} - 1)c^{a+b+k\tau}}{c^{a+b+k\tau} - 1}(\theta_1 + \theta_2)\right) + O\left(\frac{1}{m^{3/2}}\right) \tag{79}$$

(b) $$\mathbf{E}\left[(\Delta Z_n^{(m)})^2 | Z_n^{(m)} = x\right] = \frac{1}{m} k^2 g(\tau)\{1 - g(\tau)\} + O\left(\frac{1}{m^{3/2}}\right) \tag{80}$$

where $g(t) := \frac{c^{X_t} - 1}{c^{a+b+kt} - 1}$ for all $t \ge 0$.

Proof of the claim We examine the asymptotics of $p(x,n,m)$ and $\Delta X_{n/m}$. As $\tau \le R$ and $m \to \infty$, we have

$$p(x,n,m) \tag{81}$$

$$= \frac{c^{X_\tau + \frac{1}{\sqrt{m}x}} - 1}{c^{\frac{A_1^{(m)}(0) + A_2^{(m)}(0)}{m} + k\tau} - 1} = \frac{c^{X_\tau + \frac{1}{\sqrt{m}x}} - 1}{c^{a+b+k\tau + \frac{\theta_1 + \theta_2}{\sqrt{m}} + O(\frac{1}{m})} - 1} \tag{82}$$

$$= g(\tau) + \frac{\log c}{c^{a+b+k\tau} - 1}\left(c^{X_\tau} x - \frac{(c^{X_\tau} - 1)c^{a+b+k\tau}}{c^{a+b+k\tau} - 1}(\theta_1 + \theta_2)\right)\frac{1}{\sqrt{m}} + O\left(\frac{1}{m}\right).$$

The third equality follows from a Taylor's development. Also

$$\Delta X_{n/m} = X'_{n/m}\frac{1}{m} + O(m^{-2}) = kg(\tau)\frac{1}{m} + O(m^{-2}). \tag{83}$$

For X' we used the differential equation, (22), that X satisfies instead of the explicit expression for it. Substituting these expressions in (77), (78), we get the claim. $\square$

Relation (23) implies that $c^{X_\tau} = (c^{a+b} - 1)/\{c^b - 1 + c^{-k\tau}(1 - c^{-a})\}$, and this gives that the parenthesis following $\frac{1}{m}$ in equation (a) of the claim above equals

$$\frac{(c^{a+b} - 1)x - c^b(c^a - 1)(\theta_1 + \theta_2)}{c^b - 1 + c^{-k\tau}(1 - c^{-a})} \tag{84}$$

and also that

$$g(\tau)\{1-g(\tau)\} = \frac{(c^a-1)(c^b-1)c^{a+k\tau}}{(c^{a+b+k\tau}-c^{a+k\tau}+c^a-1)^2}. \tag{85}$$

Thus, the claim implies that conditions (i), (ii) of Corollary 1 are satisfied by the functions

$$\mu(x,t) = \frac{k\log c}{c^{a+b+kt}-1}\left\{\frac{(c^{a+b}-1)x-c^b(c^a-1)(\theta_1+\theta_2)}{c^b-1+c^{-kt}(1-c^{-a})}\right\}, \tag{86}$$

$$a(x,t) = k^2(c^a-1)(c^b-1)\frac{c^{a+kt}}{(c^{a+b+kt}-c^{a+kt}+c^a-1)^2}. \tag{87}$$

As in the proof of Theorem 4, condition (iii) of the corollary holds trivially, while $\lim_{m\to\infty} Z_0^{(m)} = \theta_1$ (condition (iv)). Finally, for each $x\in\mathbb{R}$ and for the choice of μ, a above, Eq. (67) has a strong solution and uniqueness in distribution holds. Thus, the process $(Z_{[mt]}^{(m)})_{t\geq 0}$ converges, as $m\to\infty$, to the unique solution of the stochastic differential equation (25).

The same is true for the process $(C_t^{(m)})_{t\geq 0}$ because $\sup_{t\geq 0}|Z_{[mt]}^{(m)}-C_t^{(m)}|\leq k/\sqrt{m}$ for all $m\in\mathbb{N}^+$ (we use the fact that $0<X_t'\leq k$ for all $t\geq 0$). To solve (25), we remark that a solution of an equation of the form

$$\mathrm{d}Y_t = (\alpha(t)Y_t+\beta(t))\,\mathrm{d}t+\gamma(t)\,\mathrm{d}W_t \tag{88}$$

with $\alpha,\beta,\gamma:[0,\infty)\to\mathbb{R}$ continuous functions is given by

$$Y_t = \mathrm{e}^{\int_0^t\alpha(s)\,\mathrm{d}s}\left(Y_0+\int_0^t\beta(s)\mathrm{e}^{-\int_0^s\alpha(r)\,\mathrm{d}r}\,\mathrm{d}s+\int_0^t\gamma(s)\mathrm{e}^{-\int_0^s\alpha(r)\,\mathrm{d}r}\,\mathrm{d}W_s\right). \tag{89}$$

[To discover the formula, we apply Itó's rule to $Y_t\exp\{-\int_0^t\alpha(s)\,\mathrm{d}s\}$ and use (88).] Applying this formula for the values of α,β,γ dictated by (25) we arrive at (26).

5 Proofs for the *q*-Pólya urn with many colors

Proof of Theorem 6 First, the equality of the expressions in (28), (29) follows from the definition of the q-multinomial coefficient.

We will prove (28) by induction on l. When $l=2$, (28) holds because of (10). In that relation, we have $x_1=x, x_2=n-x$. Assuming that (28) holds for $l\geq 2$ we will prove the case $l+1$. The probability

$$\mathbf{P}((H_2(n)=x_2,\ldots,H_{l+1}(n)=x_{l+1})$$

equals

$$\begin{aligned}
&\mathbf{P}\left(H_3(n)=x_3,\ldots,H_{l+1}(n)=x_{l+1}\right)\mathbf{P}(H_2(n)=x_2 \mid H_3(n) \\
&\quad = x_3,\ldots,H_{l+1}(n)=x_{l+1}) \\
&\quad = q^{\sum_{i=3}^{l+1} x_i \sum_{j=1}^{i-1}(w_j+kx_j)} \frac{\left[{-\frac{w_1+w_2}{k} \atop x_1+x_2}\right]_{q^{-k}} \prod_{i=3}^{l+1}\left[{-\frac{w_i}{k} \atop x_i}\right]_{q^{-k}}}{\left[{-\frac{w_1+\ldots w_{l+1}}{k} \atop n}\right]_{q^{-k}}} \\
&\qquad \times\ q^{x_2(w_1+kx_1)} \frac{\left[{-\frac{w_1}{k} \atop x_1}\right]_{q^{-k}}\left[{-\frac{w_2}{k} \atop x_2}\right]_{q^{-k}}}{\left[{-\frac{w_1+w_2}{k} \atop x_1+x_2}\right]_{q^{-k}}} \\
&\quad = q^{\sum_{i=2}^{l+1} x_i \sum_{j=1}^{i-1}(w_j+kx_j)} \frac{\prod_{i=1}^{l+1}\left[{-\frac{w_i}{k} \atop x_i}\right]_{q^{-k}}}{\left[{-\frac{w_1+\ldots w_{l+1}}{k} \atop n}\right]_{q^{-k}}}.
\end{aligned} \tag{90}$$

This finishes the induction provided that we can justify these two equalities. The second is obvious, so we turn to the first. The first probability in (90) is specified by the inductive hypothesis. That is, given the description of the experiment, in computing this probability it is as if we merge colors 1 and 2 into one color which is placed in the line before the remaining $l-1$ colors. This color has initially a_1+a_2 balls and we require that in the first n drawings we choose it x_1+x_2 times. The second probability in (90) is specified by the $l=2$ case of (28), which we know. More specifically, since the number of drawings from colors $3,4,\ldots,l+1$ is given, it is as if we have an urn with just two colors 1, 2 that have initially w_1 and w_2 balls, respectively. We do x_1+x_2 drawings with the usual rules for a q-Pólya urn, placing in a line all balls of color 1 before all balls of color 2, and we want to pick x_1 times color 1 and x_2 times color 2. □

Proof of Theorem 7 The components of $(H_2(n), H_3(n),\ldots,H_l(n))$ are increasing in n, and from Theorem 1 we have that each of them has finite limit (we treat all colors $2,\ldots,l$ as one color). Thus the convergence of the vector with probability one to a random vector with values in $\mathbb{N}^{l-1}$ follows. In particular, we also have convergence in distribution, and it remains to compute the distribution of the limit. Let $x_1 := n-(x_2+\cdots+x_l)$. Then the probability in (28) equals

$$\mathbf{P}\left(H_2(n)=x_2,\ldots,H_l(n)=x_l\right) = q^{-\sum_{1\le i<j\le l} w_j x_i} \frac{\prod_{i=1}^{l}\left[{\frac{w_i}{k}+x_i-1 \atop x_i}\right]_{q^{-k}}}{\left[{\frac{\sum_{i=1}^{l} w_i}{k}+n-1 \atop n}\right]_{q^{-k}}} \tag{91}$$

$$= q^{\sum_{1\le j<i\le l} x_i w_j} \frac{\prod_{i=1}^{l}\left[{\frac{w_i}{k}+x_i-1 \atop x_i}\right]_{q^{k}}}{\left[{n+\frac{\sum_{i=1}^{l} w_i}{k}-1 \atop n}\right]_{q^{k}}} \tag{92}$$

$$= q^{\sum_{i=2}^{l}\left(x_i \sum_{j=1}^{i-1} w_j\right)} \left\{ \prod_{i=2}^{l} \begin{bmatrix} \frac{w_i}{k} + x_i - 1 \\ x_i \end{bmatrix}_{q^k} \right\} \frac{\begin{bmatrix} x_1 + \frac{w_1}{k} - 1 \\ x_1 \end{bmatrix}_{q^k}}{\begin{bmatrix} n + \frac{\sum_{i=1}^{l} w_i}{k} - 1 \\ n \end{bmatrix}_{q^k}}. \tag{93}$$

In the first equality, we used (34) while in the second we used (35). When we take $n \to \infty$ in (93), the only terms involving n are those of the last fraction, and (46) determines their limit. Thus, the limit of (93) is found to be the function $f(x_2, \dots, x_l)$ in the statement of the theorem. $\square$

Proof of Theorem 8 For each $m \in \mathbb{N}^+$, we consider the discrete time-homogeneous Markov chain

$$Z_n^{(m)} := \left(\frac{n}{m}, \frac{A_2^{(m)}(n)}{m}, \frac{A_3^{(m)}(n)}{m}, \dots, \frac{A_l^{(m)}(n)}{m} \right), \quad n \in \mathbb{N}.$$

From any given state $(t, x) := (t, x_2, x_3, \dots, x_l)$ that $Z^{(m)}$ finds itself it moves to one of

$$\left(t + \frac{1}{m}, x_2, \dots, x_i + \frac{1}{m}, \dots, x_l\right), \quad i = 2, \dots, l,$$
$$\left(t + \frac{1}{m}, x_2, \dots, x_i, \dots, x_l\right)$$

with corresponding probabilities

$$p_i(x_2, \dots, x_l, t, m) = q^{m s_{i-1}(t)} \frac{[m x_i]_q}{[m s_l(t)]_q}, \quad i = 2, \dots, l, \tag{94}$$

$$p_1(x_2, \dots, x_l, t, m) = \frac{[m x_1(t)]_q}{[m s_l(t)]_q}, \tag{95}$$

where

$$s_i(t) = x_1(t) + \sum_{1 < j \le i} x_j \tag{96}$$

for $i \in \{1, 2, \dots, l\}$ and

$$x_1(t) := m^{-1} \sum_{j=1}^{l} A_j^{(m)}(0) + kt - \sum_{2 \le j \le l} x_i. \tag{97}$$

These follow from (27) once we count the number of balls of each color present at the state (t, x). To do this, we note that $Z_n^{(m)} = (t, x)$ implies that $n = mt$ drawings have taken place so far, the total number of balls is $A_{0,1}^{(m)} + \cdots + A_{0,l}^{(m)} + kmt$, and the number of balls of color i, for $2 \le i \le l$, is $m x_i$. Thus, the number of balls of color 1 is $A_1^{(m)}(0) + \cdots + A_l^{(m)}(0) + kmt - m\sum_{2\le j\le l} x_i = mx_1(t)$. The required relations follow.

Let $x_1 := \lim_{m\to\infty} x_1(t) = \sigma_l + kt - \sum_{2\le j\le l} x_i$ and $s_i := \lim_{m\to\infty} s_i(t) = \sum_{1\le j\le i} x_i$ for all $i \in \{1, 2, \ldots, l\}$. Then, since $q = c^{1/m}$, for fixed $(t, x_2, \ldots, x_l) \in [0,\infty)^l$ with $(x_2, \ldots, x_l) \neq 0$, we have

$$\lim_{m\to\infty} p_i(x_2, \ldots, x_l, t, m) = c^{s_{i-1}} \frac{[x_i]_c}{[s_l]_c} \tag{98}$$

for all $i = 2, \ldots, l$. We also note the following.

$$Z_{n+1,1}^{(m)} - Z_{n,1}^{(m)} = \frac{1}{m}, \tag{99}$$

$$\mathbf{E}\Big[Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)} | Z_n^{(m)} = (t, x_2, \ldots, x_l)\Big] = \frac{k}{m} p_i(x_2, \ldots, x_l, t, m), \tag{100}$$

$$\mathbf{E}\Big[(Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)})^2 | Z_n^{(m)} = (t, x_2, \ldots, x_l)\Big] = \frac{k^2}{m^2} p_i(x_2, \ldots, x_l, t, m), \tag{101}$$

$$\mathbf{E}\Big[(Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)})(Z_{n+1,j}^{(m)} - Z_{n,j}^{(m)}) | Z_n^{(m)} = (t, x_2, \ldots, x_l)\Big] = 0 \tag{102}$$

for $i, j = 2, 3, \ldots, l$ with $i \neq j$.

Therefore, with similar arguments as in the proof of Theorem 4, as $m \to +\infty$, $(Z_{[mt]}^{(m)})_{t\ge 0}$ converges in distribution to Y, the solution of the ordinary differential equation

$$\begin{aligned} \mathrm{d}Y_t &= b(Y_t)\,\mathrm{d}t, \\ Y_0 &= (0, a_2, \ldots, a_l), \end{aligned} \tag{103}$$

where $b(t, x_2, \ldots, x_l) = \big(1, b^{(2)}(t, x), b^{(3)}(t, x), \ldots, b^{(l)}(t, x)\big)$ with

$$b^{(i)}(t, x) = kc^{s_{i-1}} \frac{[x_i]_c}{[s_l]_c}$$

for $i = 2, 3, \ldots, l$. Note that $s_l = \sigma_l + kt$ does not depend on x.

Since $A_1^{(m)}([mt]) + A_2^{(m)}([mt]) + \cdots + A_l^{(m)}([mt]) = kmt + A_1^{(m)}(0) + A_2^{(m)}(0) + \cdots + A_l^{(m)}(0)$, we get that the process $(A_{[mt],1}^{(m)}/m, A_{[mt],2}^{(m)}/m + \cdots + A_{[mt],l}^{(m)}/m)_{t\ge 0}$ converges in distribution to a process $(X_{t,1}, X_{t,2}, \ldots, X_{t,l})_{t\ge 0}$ so that $X_{t,1} + \cdots + X_{t,l} = a_1 + a_2 + \cdots + a_l + kt$, while the $X_{t,i}$, $i = 2, \ldots, l$, satisfy the system

$$X_{t,i}' = kc^{\sigma_l + kt - \sum_{j=i}^{l} X_{t,i}} \frac{1 - c^{X_{t,i}}}{1 - c^{\sigma_l + kt}} \qquad \text{for all } t > 0, \tag{104}$$

$$X_{0,i} = a_i, \tag{105}$$

with $i = 2, 3, \ldots, l$. Letting $Z_{r,i} = c^{X_{\frac{\log r}{k \log c}, i}}$ for all $r \in (0, 1]$ and $i \in \{1, 2, \ldots, l\}$, we have for the $Z_{r,i}$, $i \in \{2, 3, \ldots, l\}$ the system

$$\frac{Z'_{r,i}}{1 - Z_{r,i}} = \frac{\sigma_l}{1 - \sigma_l r} \frac{1}{\prod_{i<j\leq l} Z_{r,j}}, \tag{106}$$

$$Z_{1,i} = c^{a_i}. \tag{107}$$

In the case $i = l$, the empty product equals 1. It is now easy to prove by induction (starting from $i = l$ and going down to $i = 2$) that

$$Z_{r,i} = \frac{c^{\sigma_l - \sigma_{i-1}}(1 - c^{\sigma_l} r) - c^{\sigma_l}(1 - r)}{c^{\sigma_l - \sigma_i}(1 - c^{\sigma_l} r) - c^{\sigma_l}(1 - r)} \tag{108}$$

for all $r \in (0, 1]$. Since $Z_{r,1} Z_{r,2} \cdots Z_{r,l} = c^{\sigma_l} r$, we can check that (108) holds for $i = 1$ too. The fraction in (108) equals

$$c^{a_i} \frac{(1 - c^{\sigma_l} r) - c^{\sigma_{i-1}}(1 - r)}{(1 - c^{\sigma_l} r) - c^{\sigma_i}(1 - r)}. \tag{109}$$

Recalling that $X_{t,i} = (\log c)^{-1} \log Z_{c^{kt}}$, we get (31) for all $i \in \{1, 2, \ldots, l\}$. □

Proof of Theorem 9 This is proved in the same way as Theorem 8. We keep the same notation as there. The only difference now is that $\lim_{m\to\infty} p_i(t, x_2, ..., x_l, m) = x_i/s_l$. As a consequence, the system of ordinary differential equations for the limit process $Y_t := (t, X_{t,2}, \ldots, X_{t,l})$ is (103) but with

$$b^{(i)}(t, x) = \frac{k x_i}{s_l}.$$

Recall that $s_l = \sigma_l + kt$. Thus, for $i = 2, 3, \ldots, l$, the process $X_{t,i}$ satisfies $X'_{t,i} = k X_{t,i}/(\sigma_l + kt)$, $X_{0,i} = a_i$, which give immediately the last $l - 1$ coordinates of (32). The formula for the first coordinate follows from $X_{t,1} + X_{t,2} + \cdots + X_{t,l} = kt + \sigma_l$. □

Acknowledgements We thank the anonymous referee for comments that improved the exposition of the article.

Data availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Consent to participate Not applicable.

References

1. Bagchi, A., Pal, A.K.: Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. SIAM J. Algebr. Discrete Methods **6**(3), 394–405 (1985)
2. Bai, Z.-D., Feifang, H., Zhang, L.-X.: Gaussian approximation theorems for urn models and their applications. Ann. Appl. Prob. **12**(4), 1149–1173 (2002)
3. Borovkov, K.: Gaussian process approximations for multicolor Pólya urn models. J. Appl. Prob. **58**(1), 274–286 (2021)
4. Charalambides, C.A.: A q-Pólya urn model and the q-Pólya and inverse q-Pólya distributions. J. Stat. Planning Inference **142**(1), 276–288 (2012)
5. Charalambides, C.A.: Discrete q-Distributions. Wiley, New York (2016)
6. Charalambides, C.A.: Multivariate q-Pólya and inverse q-Pólya distributions. Commun. Stat. Theory Methods (2020). https://doi.org/10.1080/03610926.2020.1825740
7. Cheliotis, D., Kouloumpou, D.: Functional limit theorems for the Pólya urn. J. Theor. Prob. (2021). https://doi.org/10.1007/s10959-021-01123-3
8. Durrett, R.: Stochastic Calculus: A Practical Introduction. CRC Press, Boca Raton (1996)
9. Ethier, S.N., Thomas, G.K.: Markov Processes: Characterization and Convergence, vol. 282. Wiley, New York (2009)
10. Gasper, G., Rahman, M.: Basic Hypergeometric Series. Cambridge University Press, Cambridge (1990)
11. Gouet, R.: Martingale functional central limit theorems for a generalized Pólya urn. Ann. Prob. **21**, 1624–1639 (1993)
12. Heyde, C.C.: On central limit and iterated logarithm supplements to the martingale convergence theorem. J. Appl. Prob. **14**(4), 758–775 (1977)
13. Janson, S.: Functional limit theorems for multitype branching processes and generalized Pólya urns. Stoch. Process. Appl. **110**(2), 177–245 (2004)
14. Johnson, N.L., Samuel, K.: Urn Models and Their Application: An Approach to Modern Discrete Probability Theory. Wiley, New York (1977)
15. Kac, V., Pokman, C.: Quantum Calculus. Springer, New York (2001)
16. Kupershmidt, B.A.: q-Probability: I. Basic discrete distributions. J. Nonlinear Math. Phys. **7**(1), 73–93 (2000)
17. Mahmoud, H.: Pólya Urn Models. Chapman and Hall/CRC, Boca Raton (2008)

Journal of Theoretical Probability
https://doi.org/10.1007/s10959-021-01123-3

Functional Limit Theorems for the Pólya Urn

Dimitris Cheliotis[1] · **Dimitra Kouloumpou**[2]

Received: 7 March 2021 / Revised: 21 July 2021 / Accepted: 1 August 2021

Abstract
For the plain Pólya urn with two colors, black and white, we prove a functional central limit theorem for the number of white balls, assuming that the initial number of black balls is large. Depending on the initial number of white balls, the limit is either a pure birth process or a diffusion.

Keywords Pólya urn · Functional limit theorems · Birth processes · Diffusion processes

Mathematics Subject Classification (2020) 60F17 · 60K99 · 60C05

1 Introduction and Results

The model. The Pólya urn is the model where in an urn that has initially A_0 white and B_0 black balls we draw, successively, and uniformly at random, a ball from it and then we return the ball back together with k balls of the same color as the one drawn. The number $k \in \mathbb{N}^+$ is fixed. Call A_n and B_n the number of white and black balls, respectively, after n drawings. The most notable result regarding the asymptotic behavior of the urn is that the proportion of white balls in the urn after n drawings, $A_n/(A_n + B_n)$, converges almost surely as $n \to \infty$ to a random variable with distribution Beta$(A_0/k, B_0/k)$.

Our aim in this work is to examine whether the entire path $(A_n)_{n\in\mathbb{N}}$, after appropriate natural transformations, converges in distribution to a nontrivial stochastic process.

✉ Dimitris Cheliotis
dcheliotis@math.uoa.gr

Dimitra Kouloumpou
dimkouloumpou@hna.gr

1 Department of Mathematics, National and Kapodistrian University of Athens, Panepistimiopolis, Athens 15784, Greece

2 Section of Mathematics, Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, Piraeus 18539, Greece

Published online: 18 August 2021

Standard references for the theory and the applications of the Pólya urn and related models are [9] and [10].

The setting. We consider an urn whose initial composition depends on $m \in \mathbb{N}^+$. It is $A_0^{(m)}$ and $B_0^{(m)}$ white and black balls, respectively. After n drawings, the composition is $A_n^{(m)}, B_n^{(m)}$.

To see a new process arising out of the path $(A_n^{(m)})_{n\in\mathbb{N}}$, we start with an initial number of balls that tends to infinity as $m \to \infty$. More specifically, we assume then that $B_0^{(m)}$ grows linearly with m. Regarding $A_0^{(m)}$, we study three regimes:

a) $A_0^{(m)}$ stays fixed with m.
b) $A_0^{(m)}$ grows to infinity but sublinearly with m.
c) $A_0^{(m)}$ grows linearly with m.

The regime where $A_0^{(m)}$ grows superlinearly with m follows from regime b) by changing the roles of the two colors. We remark on this after Theorem 2.

In the regimes a) and b), the scarcity of white balls has as a result that the time between two consecutive drawings of a white ball is of order $m/A_0^{(m)}$ (the probability of picking a white ball in the first few drawings is approximately $A_0^{(m)}/m$, which is small). We expect then that speeding up time by this factor we will see a birth process. And indeed this is the case as our first two theorems show.

In this work, all processes appearing with index set $[0, \infty)$ and values in some Euclidean space $\mathbb{R}^d$ are elements of $D_{\mathbb{R}^d}[0, \infty)$, the space of functions $f : [0, \infty) \to \mathbb{R}^d$ that are right continuous and have limits from the left at each point of $[0, \infty)$. This space is endowed with the Skorokhod topology (defined in section 5 of Chapter 3 of [5]), and convergence in distribution of processes with values on that space is defined through that topology.

We remind the reader that the negative binomial distribution with parameters $\nu \in (0, \infty)$ and $p \in (0, 1)$ is the distribution with support in $\mathbb{N}$ and probability mass function

$$f(x) = \binom{x+\nu-1}{x} p^\nu (1-p)^x \tag{1}$$

for all $x \in \mathbb{N}$. When $\nu \in \mathbb{N}^+$, this is the distribution of the number of failures until we obtain the ν-th success in a sequence of independent trials, each having probability of success p. For a random variable X with this distribution, we write $X \sim NB(\nu, p)$.

Since in each drawing we add k balls in the urn, the quantity $k^{-1}\{A_n^{(m)} - A_0^{(m)}\}$, appearing in our first two theorems, counts the number of times in the first n drawings that we selected a white ball.

Theorem 1 *Fix $a_0 \in \mathbb{N}^+$ and $b > 0$. If $A_0^{(m)} = a_0$ for all $m \in \mathbb{N}^+$ and $\lim_{m\to\infty} B_0^{(m)}/m = b$, then the process $(k^{-1}\{A_{[mt]}^{(m)} - A_0^{(m)}\})_{t\geq 0}$ converges in distribution, as $m \to \infty$, to an inhomogeneous in time pure birth process $Z = \{Z(t)\}_{t\geq 0}$ with $Z(0) = 0$ and such that for all $0 \leq t_1 < t_2$, $j \in \mathbb{N}$,*

$$Z(t_2) - Z(t_1) | Z(t_1) = j \text{ has distribution } NB\Big(\frac{a_0}{k} + j, \frac{t_1 + (b/k)}{t_2 + (b/k)}\Big).$$

In particular, Z has rates $\lambda_{t,j} = (kj + a_0)/(kt + b)$ *for all* $(t, j) \in [0, \infty) \times \mathbb{N}$.

Theorem 2 *If* $A_0^{(m)} =: g_m$ *with* $g_m \to \infty$, $g_m = o(m)$ *and* $\lim_{m\to\infty} B_0^{(m)}/m = b$ *with* $b > 0$ *constant, then the process* $(k^{-1}\{A^{(m)}_{[tm/g_m]} - A_0^{(m)}\})_{t\geq 0}$, *as* $m \to \infty$, *converges in distribution to the Poisson process on* $[0, \infty)$ *with rate* $1/b$.

We return to the discussion at the beginning of the subsection. The regime where $\lim_{m\to\infty} B_0^{(m)}(0)/m = b > 0$ and $A_0^{(m)}/m \to \infty$ is covered by the previous theorem. We need to change the roles of the colors and remark that the role of m as a scaling parameter is played now by $A_0^{(m)}$. The result that we obtain is that the process

$$\frac{1}{k}\left(B^{(m)}_{[tA_0^{(m)}/(bm)]} - B_0^{(m)}\right)_{t\geq 0}$$

converges in distribution, as $m \to \infty$, to the Poisson process on $[0, \infty)$ with rate 1.

Next, we look at regime c), i.e., in the case that at time 0 both black and white balls are of order m. In this case, the normalized process of the number of white balls has a non-random limit, which we determine, and then we study the fluctuations of the process around this limit.

Theorem 3 *Assume that* $A_0^{(m)}, B_0^{(m)}$ *are such that*

$$\lim_{m\to\infty} \frac{A_0^{(m)}}{m} = a, \lim_{m\to\infty} \frac{B_0^{(m)}}{m} = b,$$

where $a, b \in [0, \infty)$ *are not both zero. Then the process* $(A^{(m)}_{[mt]}/m)_{t\geq 0}$, *as* $m \to \infty$, *converges in distribution to the deterministic process* $X_t = \frac{a}{a+b}(a + b + kt)$, $t \geq 0$.

The limit X is the same as in an urn in which we add at each step k white or black balls with corresponding probabilities $a/(a+b)$, $b/(a+b)$, that is, irrespective of the composition of the urn at that time.

To determine the fluctuations of the process $(A^{(m)}_{[mt]}/m)_{t\geq 0}$ around its $m \to \infty$ limit, X, we let

$$C_t^{(m)} = \sqrt{m}\left(\frac{A^{(m)}_{[mt]}}{m} - X_t\right) \tag{2}$$

for all $m \in \mathbb{N}^+$ and $t \geq 0$.

Theorem 4 *Let* $a, b \in [0, \infty)$, *not both zero,* $\theta_1, \theta_2 \in \mathbb{R}$, *and assume that* $A_0^{(m)} := [am + \theta_1\sqrt{m}]$, $B_0^{(m)} = [bm + \theta_2\sqrt{m}]$ *for all large* $m \in \mathbb{N}$. *Then the process* $(C_t^{(m)})_{t\geq 0}$ *converges in distribution, as* $m \to \infty$, *to the unique strong solution of the stochastic differential equation*

$$Y_0 = \theta_1, \tag{3}$$

$$\mathrm{d}Y_t = \frac{k}{a+b+kt}\left\{Y_t - \frac{a}{a+b}(\theta_1+\theta_2)\right\}\mathrm{d}t + k\frac{\sqrt{ab}}{a+b}\,\mathrm{d}W_t, \tag{4}$$

which is

$$Y_t = \theta_1 + \frac{b\theta_1 - a\theta_2}{(a+b)^2}kt + k\frac{\sqrt{ab}}{a+b}(a+b+kt)\int_0^t \frac{1}{a+b+ks}\,\mathrm{d}W_s. \tag{5}$$

W is a standard Brownian motion

In the previous theorem, it is possible to allow other kinds of deviations away from linearity (and not only of order $\sqrt{m}$) for the values of $A_0^{(m)}$, $B_0^{(m)}$. And then we get a diffusion limit if instead of (2) we look at the process

$$D_t^{(m)} = \sqrt{m}\left(\frac{A_{[mt]}^{(m)}}{m} - \frac{A_0^{(m)}}{m} - kt\frac{A_0^{(m)}}{A_0^{(m)}+B_0^{(m)}}\right) \tag{6}$$

for all $m \in \mathbb{N}^+$ and $t \geq 0$. More specifically, we have the following.

Theorem 5 *Assume that* $\lim_{m\to\infty}\frac{A_0^{(m)}}{m} = a$, $\frac{B_0^{(m)}}{m} = b$ *where* $a, b \in [0,\infty)$ *are not both zero. Then the process* $(D_t^{(m)})_{t\geq 0}$ *converges in distribution, as* $m \to \infty$*, to the unique strong solution of the stochastic differential equation*

$$V_0 = 0, \tag{7}$$

$$dV_t = \frac{kV_t}{a+b+kt}\,\mathrm{d}t + k\frac{\sqrt{ab}}{a+b}\,\mathrm{d}W_t, \tag{8}$$

which is

$$V_t = k\frac{\sqrt{ab}}{a+b}(a+b+kt)\int_0^t \frac{1}{a+b+ks}\,\mathrm{d}W_s. \tag{9}$$

W is a standard Brownian motion

Remark Functional central limit theorems for Pólya type urns have been proven with increasing generality in the works [2,6,8]. The major difference with our results is that in theirs the initial number of balls, $A_0^{(m)}$, $B_0^{(m)}$, is fixed (see however the last point in the list, concerning the recent work [3]). More specifically:

(1) Gouet ([6]) studies urns with two colors (black and white) in the setting of Bagchi and Pal ([1]). According to that, when a white ball is drawn, we return it in the urn together with a white and b black balls, while if a black ball is drawn, we return it together with c white and d black. The numbers a, b, c, d are fixed integers (possibly negative), the number of balls added to the urn is fixed (that is $a + b = c + d$), and balls are drawn uniformly form the urn. The plain Pólya urn is not studied in that work because, according to the author, it has been studied by Heyde in [7]. However,

for the Pólya urn, [7] discusses the central limit theorem and the law of the iterated logarithm. In any case, following the techniques of Heyde and Gouet one can prove the following. Assume for simplicity that $k = 1$ and let $L =: \lim_{n\to\infty} \frac{A_n}{n}$. The limit exists with probability one because of the martingale convergence theorem. Then

$$\left\{\sqrt{n}\left(t\frac{A_{[n/t]}}{n} - L\right)\right\}_{t\geq 0} \xrightarrow{d} \{W_{L'(1-L')t}\}_{t\geq 0}$$

as $n \to \infty$. W is a standard Brownian motion and L' is a random variable independent of W and having the same distribution as L. On the other hand, de-Finetti's theorem gives easily the more or less equivalent statement that, as $n \to \infty$,

$$\left\{\sqrt{n}\left(\frac{A_{[nt]}}{nt} - L\right)\right\}_{t\geq 0} \xrightarrow{d} \{W_{L'(1-L')/t}\}_{t\geq 0}$$

with W, L' as before.

(2) Bai, Hu, and Zhang ([2]) work again in the setting of Bagchi and Pal, but now the numbers a, b, c, d depend on the order of the drawing and are random. The requirement that each time we add the same number of balls is relaxed.

(3) Janson ([8]) considers urns with many colors, labeled $1, 2, \dots, l$, where after each drawing, if we pick a ball of color i, we place in the urn balls of every color according to a random vector $(\xi_{i,1}, \dots, \xi_{i,l})$ whose distribution depends on i ($\xi_{i,j}$ is the number of balls of color j that we add in the urn). Also, each ball is assigned a certain nonrandom activity that depends only on its color, and then the probability to pick a certain color at a drawing equals the ratio of the total of the activities of all balls of that color to the total of the activities of all balls present in the urn at that time. A restriction in that work is that there is a color i_0 so that starting the urn with just one ball and this ball has this color, there is positive probability to see in the future every other color. This excludes the classical Pólya urn that we study.

(4) In [3], K. Borovkov studies a Pólya urn with $d + 1$ colors, $1, 2, \dots, d + 1$, and proves convergence after appropriate scaling for the path $\{M([nt])\}_{t\in[0,1]}$, as $n \to \infty$, where

$$M(j) := (\xi_1(j), \xi_1(j) + \xi_2(j), \dots, \sum_{i=1}^{d} \xi_i(j)) \in \mathbb{N}^d$$

and $\xi_i(j)$ is the number of balls of color i present in the urn at time j. The initial total number of balls in the urn is N and the author considers limits as $N, n \to \infty$ with $n/N \to c$ under the regimes $c = 0$, $c \in (0, \infty)$, $c = \infty$. It assumes that at each drawing we add one ball, i.e., $k = 1$ in our notation.

Its relation to the present work is the following. We study only the case $d = 1$, and then $M(j) = \xi_1(j) = A_j^{(m)}$.

a) Theorems 1 and 2 are not covered by [3] because in Corollary 1 of [3] the changes $A_{[mt]}^{(m)} - A_0^{(m)}$, $A_{[tm/g_m]}^{(m)} - A_0^{(m)}$ are divided by $\sqrt{m}$ and $\sqrt{m/g_m}$, respectively (and then m is sent to infinity), while in Theorem 1 of [3], these changes are related to certain processes but with an error term of the order of $\log^2 m$. That is, in the scenarios of Theorems 1 and 2, the results of [3] are too rough to capture the birth process that we identify.

b) Theorems 4, 5 follow from Corollary 1(ii) in [3]. For example, under the assumptions of Theorem 4, the Corollary gives that

$$C_t^{(m)} - \theta_1 - \frac{b\theta_1 - a\theta_2}{(a+b)^2} t = H_t + o_{\mathbf{P}}(1)$$

for all $t \in [0, 1]$, where the supremum of the error term, $o_{\mathbf{P}}(1)$, over $t \in [0, 1]$ goes to zero in probability as $m \to \infty$, while the process H is Gaussian with continuous paths, mean function zero, and covariance function

$$\mathrm{Cov}(H_s, H_t) = \frac{ab}{(a+b)^3} s(a+b+t)$$

for all $0 \le s \le t$. The term involving the stochastic integral in (5) also defines a Gaussian process with continuous paths and the same mean and covariance function as H. The justification for Theorem 5 is similar.

A preprint of the present work appeared in the arxiv on May 30, 2019, a few months before the preprint of [3].

2 Jump Process Limits. Proof of Theorems 1, 2

In the case of Theorem 1, we let $g_m := 1$ for all $m \in \mathbb{N}^+$, and for both theorems we let $v := v_m := m/g_m$ (we suppress the dependence of v on m). Our interest is in the sequence of the processes $(Z^{(m)})_{m \in \mathbb{N}^+}$ with

$$Z^{(m)}(t) = \frac{1}{k}(A_{[vt]}^{(m)} - A_0^{(m)}) \tag{10}$$

for all $t \ge 0$.

To show convergence in distribution, according to Theorem 7.8 of Chapter 3 of [5], it is enough to show that the sequence $(Z^{(m)})_{m \ge 1}$ is tight and its finite dimensional distributions converge. The description of the limiting process is determined on the way.

An easy argument shows that tightness follows from the convergence of the finite dimensional distributions because each $Z^{(m)}$ has non-decreasing paths. It thus remains to establish the convergence of the finite dimensional distributions.

Notation: (i) For sequences $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ with values in $\mathbb{R}$, we will say that they are asymptotically equivalent, and will write $a_n \sim b_n$ as $n \to \infty$, if $\lim_{n \to \infty} a_n/b_n = 1$. We use the same expressions for functions f, g defined in a neighborhood of ∞ and satisfy $\lim_{x \to \infty} f(x)/g(x) = 1$.

(ii) For $a \in \mathbb{C}$ and $k \in \mathbb{N}^+$, let

$$(a)_k := a(a-1)\cdots(a-k+1), \tag{11}$$

$$a^{(k)} := a(a+1)\cdots(a+k-1), \tag{12}$$

the falling and rising factorial, respectively. Also let $(a)_0 := a^{(0)} := 1$.

2.1 Convergence of Finite Dimensional Distributions

By definition, $Z^{(m)}(0) = 0 = Z(0)$ for all $m \in \mathbb{N}^+$.

Since, for each $m \in \mathbb{N}^+$, the process $Z^{(m)}$ is Markov taking values in $\mathbb{N}$ and non-decreasing in time, our objective will have been accomplished if we show that the conditional probability

$$\mathbf{P}(Z^{(m)}(t_2) = k_2 | Z^{(m)}(t_1) = k_1) \tag{13}$$

converges as $m \to \infty$ for each $0 \le t_1 < t_2$ and nonnegative integers $k_1 \le k_2$. Define

$$n := [vt_2] - [vt_1], \tag{14}$$

$$x := k_2 - k_1, \tag{15}$$

$$\sigma := \frac{A_0^{(m)} + kk_1}{k}, \tag{16}$$

$$\tau := \frac{k[vt_1] - kk_1 + B_0^{(m)}}{k}. \tag{17}$$

Then, the above probability equals

$$\begin{aligned} &\mathbf{P}(A^{(m)}_{[vt_2]} = kk_2 + a_0 | A^{(m)}_{[vt_1]} = kk_1 + a_0) \\ &\quad = \binom{n}{x} \frac{k\sigma(k\sigma+k)\cdots(k\sigma+(x-1)k)k\tau(k\tau+k)\cdots(k\tau+(n-x-1)k)}{(k\sigma+k\tau)(k\sigma+k\tau+k)\cdots(k\sigma+k\tau+(n-1)k)} \end{aligned} \tag{18}$$

$$= \frac{(n)_x}{x!} \frac{\sigma^{(x)}\tau^{(n-x)}}{(\sigma+\tau)^{(n)}} = \frac{(n)_x}{x!} \sigma^{(x)} \frac{\Gamma(\tau+n-x)}{\Gamma(\tau)} \frac{\Gamma(\sigma+\tau)}{\Gamma(\sigma+\tau+n)}. \tag{19}$$

To compute the limit as $m \to \infty$ of (19), we will use Stirling's approximation for the Gamma function,

$$\Gamma(y) \sim \left(\frac{y}{e}\right)^y \sqrt{\frac{2\pi}{y}} \tag{20}$$

as $y \to \infty$, and its consequence

$$\Gamma(y+a) \sim \Gamma(y) y^a \tag{21}$$

as $y \to \infty$ for all $a \in \mathbb{R}$.

Proof (**The computation for Theorem 1**) Recall that $v = m$ in this theorem. Using (21) twice, with the role of a played by $-x$ and σ, we see that the last quantity in (19), for $m \to \infty$, is asymptotically equivalent to

$$\begin{aligned}\frac{(m(t_2-t_1))^x}{x!}\sigma^{(x)}\tau^{\sigma}\frac{(\tau+n)^{-x}}{(\tau+n)^{\sigma}} &\sim \frac{(m(t_2-t_1))^x}{x!}\sigma^{(x)}\frac{\{m(t_1+(b/k))\}^{\sigma}}{\{m(t_2+(b/k))\}^{\sigma+x}}\\ &= \frac{(t_2-t_1)^x}{x!}\sigma^{(x)}\frac{\{t_1+(b/k)\}^{\sigma}}{\{t_2+(b/k)\}^{\sigma+x}}\\ &= \binom{\sigma+x-1}{x}\left(\frac{t_2-t_1}{t_2+(b/k)}\right)^x\left(1-\frac{t_2-t_1}{t_2+(b/k)}\right)^{\sigma}.\end{aligned} \tag{22}$$

[For reader's convenience, we remark that the asymptotics, as $m \to \infty$, of the relevant quantities are as follows: x, σ are constants while $n \sim (t_2-t_1)m$, $\tau \sim (t_1+(b/k))m$.]

Thus, as $m \to \infty$, the distribution of $\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1) = k_1$ converges to the negative binomial distribution with parameters σ, $\frac{t_1+(b/k)}{t_2+(b/k)}$ [recall (1)]. □

Proof (**The computation for Theorem 2**) Using (20), we see that the last quantity in (19), for $m \to \infty$, is asymptotically equivalent to

$$\begin{aligned}&\frac{(m(t_2-t_1))^x}{x!g_m^x}\frac{g_m^x}{k^x}e^x\frac{(\tau+n-x)^{\tau+n-x}}{\tau^{\tau}}\frac{(\sigma+\tau)^{\sigma+\tau}}{(\sigma+\tau+n)^{\sigma+\tau+n}}\\ &\sim \frac{m^x(t_2-t_1)^x}{x!k^x}e^x(\tau+n-x)^{-x}\left(\frac{\tau+n-x}{\sigma+\tau+n}\right)^n\\ &\times\left(\frac{\sigma+\tau}{\sigma+\tau+n}\right)^{\sigma}\left(\frac{(\tau+n-x)(\sigma+\tau)}{\tau(\sigma+\tau+n)}\right)^{\tau}\\ &\sim \frac{m^x(t_2-t_1)^x}{x!k^x}e^x\tau^{-x}e^{-(t_2-t_1)/b}e^{-(t_2-t_1)/b}e^{-x+(t_2-t_1)/b}\\ &\sim \frac{1}{x!}\left(\frac{t_2-t_1}{b}\right)^x e^{-(t_2-t_1)/b}.\end{aligned}$$

[Here, the asymptotics, as $m \to \infty$, of the relevant quantities are as follows: x is constant while $n \sim (t_2-t_1)m/g_m$, $\tau \sim (b/k)m$, $\sigma \sim g_m/k$.]

Thus, as $m \to \infty$, the distribution of

$$\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1) = k_1$$

converges to the Poisson distribution with parameter $(t_2-t_1)/b$. □

2.2 Conclusion

It is clear from the form of the finite dimensional distributions that in both Theorems 1, 2 the limiting process Z is a pure birth process that does not explode in finite time.

Its rate at the point $(t, j) \in [0, \infty) \times \mathbb{N}$ is

$$\lambda_{t,j} = \lim_{h \to 0^+} \frac{1}{h} \mathbf{P}(Z(t+h) = j+1 | Z(t) = j)$$

and is found as stated in the statement of each theorem.

3 Deterministic and Diffusion Limits. Proof of Theorems 3, 4, 5.

Theorems 3, 4, 5 are proved with the use of Theorem 7.1 in Chapter 8 of [4], which is concerned with convergence of time-homogeneous Markov chains to diffusions. The chains whose convergence is of interest to us are time inhomogeneous, but we reduce their study to the time-homogenous setting by considering for each such chain $\{Z_n\}_{n\in\mathbb{N}}$ the time homogeneous chain $\{(Z_n, n)\}_{n\in\mathbb{N}}$. The following consequence of the aforementioned theorem suffices for our purposes.

Corollary 1 *Assume that for each $m \in \mathbb{N}^+$, $(Z_n^{(m)})_{n\in\mathbb{N}}$ is a Markov chain in $\mathbb{R}$. For each $m \in \mathbb{N}^+$ and $n \in \mathbb{N}$, let $\Delta Z_n^{(m)} := Z_{n+1}^{(m)} - Z_n^{(m)}$ and*

$$\mu^{(m)}(x, n) := m\boldsymbol{E}(\Delta Z_n^{(m)} \boldsymbol{1}_{|\Delta Z_n^{(m)}| \le 1} | Z_n^{(m)} = x), \tag{23}$$

$$a^{(m)}(x, n) := m\boldsymbol{E}\{(\Delta Z_n^{(m)})^2 \boldsymbol{1}_{|\Delta Z_n^{(m)}| \le 1} | Z_n^{(m)} = x\} \tag{24}$$

for all $x \in \mathbb{R}$ with $\mathbf{P}(Z_n^{(m)} = x) > 0$. Also, for $R > 0$ and for the same m, n as above, let $A(m, n, R) := \{(x, n) : |x| \le R, n/m \le R, \mathbf{P}(Z_n^{(m)} = x) > 0\}$.

Assume that there are continuous functions $\mu : \mathbb{R} \times [0, \infty) \to \mathbb{R}$, $a : \mathbb{R} \times [0, \infty) \to [0, \infty)$, and $x_0 \in \mathbb{R}$ so that:
For every $R, \varepsilon > 0$, it holds

(i) $\sup_{(x,n)\in A(m,n,R)} |\mu^{(m)}(x, n) - \mu(x, n/m)| \to 0$ *as* $m \to \infty$,
(ii) $\sup_{(x,n)\in A(m,n,R)} |a^{(m)}(x, n) - a(x, n/m)| \to 0$ *as* $m \to \infty$,
(iii) $\sup_{(x,n)\in A(m,n,R)} m\mathbf{P}(|\Delta Z_n^{(m)}| \ge \varepsilon | Z_n^{(m)} = x) \to 0$ *as* $m \to \infty$,

and also

(iv) $Z_0^{(m)} \to x_0$ *as* $m \to \infty$ *with probability 1,*
(v) for each $x \in \mathbb{R}$, the stochastic differential equation

$$\begin{aligned} dZ_t &= \mu(Z_t, t)\,\mathrm{d}t + \sqrt{a(Z_t, t)}\,\mathrm{d}B_t, \\ Z_0 &= x, \end{aligned} \tag{25}$$

where B is a one-dimensional Brownian motion, has a weak solution which is unique in distribution.

Then, the process $(Z_{[mt]}^{(m)})_{t\ge0}$ converges in distribution to the weak solution of (25) with $x = x_0$.

Proof For each $m \in \mathbb{N}^+$, we consider the process $Y_n^{(m)} := (Z_n^{(m)}, n/m), n \in \mathbb{N}$, which is a time-homogeneous Markov chain with values in $\mathbb{R}^2$, and we apply Theorem 7.1 in Chapter 8 of [4] Conditions (i), (ii), (iii) of that theorem follow from our conditions (ii), (i), (iii) respectively, while condition (A) there translates to the requirement that the martingale problem for the functions μ and $\sqrt{a}$ is well posed, and this follows from condition (v). □

The tool we will use in checking that condition (v) of the corollary is satisfied is the well-known existence and uniqueness theorem for strong solutions of SDEs which requires that for all $T > 0$, the coefficients $\mu(x, t)$, $\sqrt{a(x, t)}$ are Lipschitz in x uniformly for $t \in [0, T]$ and $\sup_{t\in[0,T]}\{|\mu(0, t)| + a(0, t)\} < \infty$ (e.g., Theorem 2.9 of Chapter 5 or [4]). The same conditions imply uniqueness in distribution.

3.1 Proof of Theorem 3

We will apply Corollary 1. For each $m \in \mathbb{N}^+$, consider the Markov chain $Z_n^{(m)} = \frac{A_n^{(m)}}{m}, n \in \mathbb{N}$. From any given state x of $Z_n^{(m)}$, the chain moves to either of $x + km^{-1}$, x with corresponding probabilities $p(x, n, m)$, $1 - p(x, n, m)$, where

$$p(x, n, m) := \frac{mx}{A_0^{(m)} + B_0^{(m)} + kn}. \tag{26}$$

In particular, for any $\varepsilon > 0$, is holds $|\Delta Z_n^{(m)}| < 1 \wedge \varepsilon$ for m large enough. Thus, condition (iii) of the corollary is satisfied trivially. Also, for large m, with the notation of the corollary, we have

$$\mu^{(m)}(x, n) = kp(x, n, m), \tag{27}$$

$$a^{(m)}(x, n) = \frac{k}{m} p(x, n, m). \tag{28}$$

And it is easy to see that conditions (i), (ii) are satisfied by the functions a, μ with $a(x, t) = 0$ and $\mu(x, t) = kp(x, t)$ where

$$p(x, t) := \frac{x}{a + b + kt}. \tag{29}$$

Now for each $x \in \mathbb{R}$, the equation

$$\begin{aligned} dZ_t &= kp(Z_t, t)\,\mathrm{d}t, \\ Z_0 &= x, \end{aligned} \tag{30}$$

has a unique solution. Thus, Corollary 1 applies. In fact, (30) is a separable ordinary differential equation and its unique solution is the one given in the statement of the theorem.

3.2 Proof of Theorem 4

Call $\lambda := a/(a+b)$. For each $m \in \mathbb{N}^+$, consider the Markov chain

$$Z_n^{(m)} = \sqrt{m}\Big(\frac{A_n^{(m)}}{m} - X_{\frac{n}{m}}\Big), n \in \mathbb{N}.$$

From any given state x of $Z_n^{(m)}$, the chain moves to either of $x - km^{-1/2}\lambda$, $x + km^{-1/2}(1-\lambda)$ with corresponding probabilities

$$\frac{B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}}, \frac{A_n^{(m)}}{A_n^{(m)} + B_n^{(m)}},$$

where

$$A_n^{(m)} = ma + \lambda kn + x\sqrt{m}, \tag{31}$$

$$B_n^{(m)} = A_0^{(m)} + B_0^{(m)} + kn - A_n^{(m)}. \tag{32}$$

Note that

$$A_0^{(m)} + B_0^{(m)} = (a+b)m + (\theta_1+\theta_2)\sqrt{m} + \delta_m, \tag{33}$$

with $\delta_m \in [0, 2)$, and consequently

$$A_n^{(m)} = \lambda(A_n^{(m)} + B_n^{(m)}) + \sqrt{m}(x - \lambda(\theta_1+\theta_2)) - \lambda\delta_m. \tag{34}$$

Again, condition (iii) of Corollary 1 holds trivially, while $\lim_{m\to\infty} Z_0^{(m)} = \theta_1$ (condition (iv)). Then, for large m we have

$$\mu^{(m)}(x,n) = k\sqrt{m}\frac{(1-\lambda)A_n^{(m)} - \lambda B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}} = k\frac{x - \lambda(\theta_1+\theta_2) - \lambda\frac{\delta_m}{\sqrt{m}}}{\frac{A_0^{(m)}+B_0^{(m)}}{m} + k\frac{n}{m}}, \tag{35}$$

$$a^{(m)}(x,n) = k^2\left(\lambda^2\frac{B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}} + (1-\lambda)^2\frac{A_n^{(m)}}{A_n^{(m)} + B_n^{(m)}}\right) \tag{36}$$

$$= k^2\lambda(1-\lambda) + k^2(1-2\lambda)\frac{\sqrt{m}(x - \lambda(\theta_1+\theta_2)) - \lambda\delta_m}{A_n^{(m)} + B_n^{(m)}}. \tag{37}$$

It follows that conditions (i), (ii) are satisfied by the functions μ, a with

$$\mu(x,t) = \frac{k\{x - (\theta_1+\theta_2)\lambda\}}{a+b+kt}, \tag{38}$$

$$a(x,t) = \frac{k^2ab}{(a+b)^2}. \tag{39}$$

For each $x \in \mathbb{R}$, the stochastic differential equation

$$\mathrm{d}Y_t = \frac{k\{Y_t - (\theta_1 + \theta_2)\lambda\}}{a+b+kt}\,\mathrm{d}t + k\frac{\sqrt{ab}}{a+b}\mathrm{d}W_t, \tag{40}$$

$$Y_0 = x, \tag{41}$$

where W is a standard Brownian motion, has a unique strong solution as the drift and diffusion coefficients are Lipschitz in Y_t and grow at most linearly in Y_t at infinity (both conditions uniformly in t). Thus, Corollary 1 applies and gives that the process $(Z^{(m)}_{[mt]})_{t\geq 0}$ converges in distribution, as $m \to \infty$, to the solution of (40), (41) with $x = \theta_1$. The same is true for $(C^{(m)}_t)_{t\geq 0}$ because $\sup_{t\geq 0}|C^{(m)}_t - Z^{(m)}_{[mt]}| = \sup_{t\geq 0}\sqrt{m}\lambda k(t - [mt]/m) = \lambda k/\sqrt{m} \to 0$ as $m \to \infty$.

To solve the stochastic differential Eq. (40), (41), we set $U_t := \{Y_t - (\theta_1+\theta_2)\lambda\}/(a+b+kt)$. Itô's lemma gives that

$$\mathrm{d}U_t = k\frac{\sqrt{ab}}{(a+b)}\frac{1}{a+b+kt}\mathrm{d}W_t,$$

and since $U_0 = (b\theta_1 - a\theta_2)/(a+b)^2$, we get

$$U_t = \frac{b\theta_1 - a\theta_2}{(a+b)^2} + k\frac{\sqrt{ab}}{a+b}\int_0^t \frac{1}{a+b+ks}\mathrm{d}W_s.$$

This gives (5).

3.3 Proof of Theorem 5

The proof is analogous to that of Theorem 4. Call $\lambda_m := A^{(m)}_0/(A^{(m)}_0 + B^{(m)}_0)$. For each $m \in \mathbb{N}^+$, consider the Markov chain

$$Z^{(m)}_n = \sqrt{m}\left(\frac{A^{(m)}_n}{m} - \frac{A^{(m)}_0}{m} - \lambda_m k\frac{n}{m}\right), n \in \mathbb{N}.$$

From any given state x of $Z^{(m)}_n$, the chain moves to either of $x - km^{-1/2}\lambda_m, x + km^{-1/2}(1-\lambda_m)$ with corresponding probabilities

$$\frac{B^{(m)}_n}{A^{(m)}_n + B^{(m)}_n}, \frac{A^{(m)}_n}{A^{(m)}_n + B^{(m)}_n},$$

where

$$A^{(m)}_n = A^{(m)}_0 + \lambda_m kn + x\sqrt{m}, \tag{42}$$

$$B^{(m)}_n = A^{(m)}_0 + B^{(m)}_0 + kn - A^{(m)}_n. \tag{43}$$

Again, condition (iii) of Corollary 1 holds trivially, while $\lim_{m\to\infty} Z_0^{(m)} = 0$ (condition (iv)). Then, for large m we have

$$\mu^{(m)}(x,n) = k\sqrt{m}\frac{(1-\lambda_m)A_n^{(m)} - \lambda_m B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}} = \frac{kx}{\frac{A_0^{(m)}+B_0^{(m)}}{m} + k\frac{n}{m}}, \tag{44}$$

$$a^{(m)}(x,n) = k^2\left(\lambda_m^2 \frac{B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}} + (1-\lambda_m)^2 \frac{A_n^{(m)}}{A_n^{(m)} + B_n^{(m)}}\right) \tag{45}$$

$$= k^2\lambda_m(1-\lambda_m) + k^2(1-2\lambda_m)\frac{x\sqrt{m}}{A_n^{(m)} + B_n^{(m)}}. \tag{46}$$

Note now that $\lim_{m\to\infty}\lambda_m = a/(a+b)$ and $\lim_{m\to\infty}(A_n^{(m)} + B_n^{(m)})/m = a+b$. It follows that conditions (i), (ii) are satisfied by the functions μ, a with

$$\mu(x,t) = \frac{kx}{a+b+kt}, \tag{47}$$

$$a(x,t) = \frac{k^2ab}{(a+b)^2}. \tag{48}$$

For each $x \in \mathbb{R}$, the stochastic differential equation

$$\mathrm{d}V_t = \frac{kY_t}{a+b+kt}\,\mathrm{d}t + k\frac{\sqrt{ab}}{a+b}\mathrm{d}W_t, \tag{49}$$

$$V_0 = x, \tag{50}$$

where W is a standard Brownian motion, has a unique strong solution as the drift and diffusion coefficients are Lipschitz in V_t and grow at most linearly in V_t at infinity (both conditions uniformly in t). Thus, Corollary 1 applies and gives that the process $(Z_{[mt]}^{(m)})_{t\ge 0}$ converges in distribution, as $m\to\infty$, to the solution of (49), (50) with $x=0$. The same is true for $(D_t^{(m)})_{t\ge 0}$ because $\sup_{t\ge 0}|D_t^{(m)} - Z_{[mt]}^{(m)}| \le k/\sqrt{m} \to 0$ as $m\to\infty$.

Easily one finds that the solution of the stochastic differential Eq. (49), (50) with $x=0$ is (9)

Acknowledgements We thank an anonymous referee for comments that improved significantly the exposition and for suggesting the question of Theorem 5. The first author was supported by the Hellenic Foundation for Research and Innovation (H. F. R. I.) under the “First Call for H. F. R. I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment Grant” (Project Number: 1034).

References

1. Bagchi, A., Pal, A.K.: Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. SIAM J. Algebraic Discrete Methods **6**(3), 394–405 (1985)

2. Bai, Z.-D., Feifang, H., Zhang, L.-X.: Gaussian approximation theorems for urn models and their applications. Ann. Appl. Probab. **12**(4), 1149–1173 (2002)
3. Borovkov, K.: Gaussian process approximations for multicolor Pólya urn models. J. Appl. Probab. **58**(1), 274–286 (2021)
4. Durrett, R.: Stochastic Calculus: a Practical Introduction. CRC Press, Boca Raton (1996)
5. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence, vol. 282. Wiley, Hoboken (2009)
6. Gouet, R.: Martingale functional central limit theorems for a generalized Pólya urn. Ann. Probab. 1624–1639 (1993)
7. Heyde, C.C.: On central limit and iterated logarithm supplements to the martingale convergence theorem. J. Appl. Probab. **14**(4), 758–775 (1977)
8. Janson, S.: Functional limit theorems for multitype branching processes and generalized Pólya urns. Stoch. Process. Their Appl. **110**(2), 177–245 (2004)
9. Johnson, N.L., Samuel, K.: "Urn models and their application; an approach to modern discrete probability theory." (1977)
10. Mahmoud, H.: Pólya urn Models. Chapman and Hall/CRC, Boca Raton (2008)

Chaos, Solitons and Fractals 136 (2020) 109828

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

SBDiEM: A new mathematical model of infectious disease dynamics

Stelios Bekiros[a,b,*], Dimitra Kouloumpou[c]

[a] European University Institute, Via delle Fontanelle, 18, Florence I-50014, Italy
[b] RCEA, LH3079, Wilfrid Laurier University, 75 University Ave W., Waterloo, ON N2L3C5, Canada
[c] Hellenic Naval Academy, Section of Mathematics, Mathematical Modeling and Applications Laboratory, Piraeus 18539, Greece

ARTICLE INFO

Article history:
Received 28 March 2020
Accepted 16 April 2020
Available online 23 April 2020

JEL classification:
90.01 Social phenomena

Keywords:
Epidemiology
Outbreak analysis
COVID-19
Contagious dynamics
Stochastic models
Virus transmissibility

ABSTRACT

A worldwide multi-scale interplay among a plethora of factors, ranging from micro-pathogens and individual or population interactions to macro-scale environmental, socio-economic and demographic conditions, entails the development of highly sophisticated mathematical models for robust representation of the contagious disease dynamics that would lead to the improvement of current outbreak control strategies and vaccination and prevention policies. Due to the complexity of the underlying interactions, both deterministic and stochastic epidemiological models are built upon incomplete information regarding the infectious network. Hence, rigorous mathematical epidemiology models can be utilized to combat epidemic outbreaks. We introduce a new spatiotemporal approach (SBDiEM) for modeling, forecasting and nowcasting infectious dynamics, particularly in light of recent efforts to establish a global surveillance network for combating pandemics with the use of artificial intelligence. This model can be adjusted to describe past outbreaks as well as COVID-19. Our novel methodology may have important implications for national health systems, international stakeholders and policy makers.

1. Introduction

The World Health Organization (WHO) reported on December 31, 2019 cases of pneumonia of undetected etiology in the city of Wuhan, Hubei Province in China. A novel coronavirus (CoViD-19) was identified as the source of the disease by the Chinese authorities on January 7, 2020. Eventually, the International Committee on Taxonomy of Viruses on 11 February, 2020 named the Severe Acute Respiratory Syndrome Coronavirus as SARS-CoV-2 [1]. Concerns on public health were dispersed on a global scale about potentially infected countries. The virus might have been generated by animal populations and transmitted via the Huanan wholesale market [2–4] albeit not proven, while clinical findings demonstrated that international spread was caused mainly by commercial air travel [4–7]. The WHO declared SARS-CoV-2 a pandemic on March 11, 2020. Throughout the globe, huge efforts were in progress to limit the spread of the virus and find medications and vaccines. However, the scientific community could not fully comprehend the dynamics of the spread [8–10].

Several outbreaks of infectious diseases have occurred in the past with immense impact on public health. For instance, the Severe Acute Respiratory Syndrome (SARS) occurred in 2003, the swine flu in 2009 and the Middle East Respiratory Syndrome Coronavirus (MERS) in Saudi Arabia in 2012, which still survives at a sub-critical level causing some peaks [11–14]. Additionally, the Ebola epidemic emerged between 2014 and 2016 and caused over 28,000 cases in West Africa [15]. Its temporal decline coincided with the outbreak of Zika virus in Brazil [16]. Consequently, the outbreak of severe pathogens such as the aforementioned, require global interdisciplinary efforts in order to decode key epidemiological features and their transmission dynamics, and develop possible control policies.

Insights from mathematical modelling can be extremely beneficial. Indeed, dealing with infectious diseases from a mathematical angle could reveal inherent patterns and underlying structures that govern outbreaks. Stakeholders utilize available data from current and previous outbreaks in order to forecast infection rates, identify how to restrict the spread of diseases, and eventually introduce vaccination policies that will be most effective. Epidemiology is essentially a biology discipline concerned with public health and as such, it can be heavily influenced by mathematical theory. Most phenomena observed at population level are often very complex and difficult to decode just by observing the characteristics of isolated individuals [17]. Statistical analyses of epidemiological data help to characterize, quantify and summarize the way diseases spread in host populations. Interestingly, mathematical models appear as efficient ways to explore and test various epidemiologi-

* Corresponding author at: Department of Economics, Via delle Fontanelle, 18, I-50014 Florence, Italy.
E-mail addresses: stelios.bekiros@eui.eu (S. Bekiros), dimkouloumpou@hna.gr (D. Kouloumpou).

https://doi.org/10.1016/j.chaos.2020.109828
0960-0779/

cal hypotheses, mostly due to the existence of ethical and practical limitations when deducting experiments on living populations. Models provide conceptual results on e.g., the basic reproduction number, threshold effects or herd immunity. One additional element of epidemiological modeling is the link with data via statistical methods. Although simple epidemiological models are often used, viral and bacterial infections commonly require increased complexity. There are many models in the literature on single epidemics, endemic diseases and spatiotemporal disease dynamics. The aim is to develop robust public health policies in defining optimal vaccination strategies.

Our study presents for the first time a new stochastic mathematical model for describing infectious dynamics and tracking virus temporal transmissibility on 3-dimensional space (earth). This model can be adjusted to describe all past outbreaks as well as CoViD-19. As a matter of fact, it introduces a novel approach to mathematical modelling of infectious dynamics of any disease, and sets a starting point for conducting simulations, forecasting and nowcasting investigations based on real-world stereographic and spherical tracking on earth.

In short, a single epidemic outbreak as opposed to disease endemicity occurs in a time span short enough not to have the demographic changes perturbing the dynamics of contacts among individuals. The most popular mathematical model in this category is the Susceptible-Infected-Recovered (SIR) epidemic model, in which all individuals of a finite population interact in the same manner. Individuals at time t are susceptible (S), infected (I) or recovered (R). The final size of the epidemic will strongly depend upon the initial conditions of the number of susceptible and infected individuals as well as the infection parameter. The final size distribution of the simple SIR model in most cases is bimodal presenting two local maxima. This bimodal feature is caused by two likely scenarios; either the epidemic dies out quickly infecting few individuals, or it becomes long-lasting and substantial. However, stochasticity in the form of random walk transmission mechanisms related to spreading processes has never been explored in epidemiology widely [18–20]. For example, in computer science, some artificially created viruses propagate randomly by a plethora of online communication channels. To the best of our knowledge, we are the first to scrutinize extensively the role of random walks in epidemic spreading and provide the proper mathematical arsenal to model it robustly. Interestingly, random walk paths converge in distribution to Brownian motions [21]. In this work, we assume that a biological carrier of virus Y is at position $X(t)$ at any given time t. We call this the inaugural contamination focal point on earth.

The path defined by its motion is considered infectious. $X_t, t \geq 0$ is supposed to follow a Brownian motion on a 2-dimensional sphere S^2 of radius a, i.e the sphere in $\mathbb{R}$ of dimension 3. We consider this a proxy for earth, spreading via spherical and stereographic coordinates. Next, using the Laplace-Beltrami operator we construct the Brownian motion infectious process on the 2-dimensional sphere, using spherical and stereographic coordinates as local coordinates. We evaluate explicitly certain quantities related to generated diffusion processes. In what follows, we compute the transition and transmission density for the $X_t, t \geq 0$, and we derive the stochastic differential equations that govern the infectious disease dynamics for $X_t, t \geq 0$ in those local coordinates. We continue with the calculation of expectations of outbreak exit times in time and space of specific domains, possessing certain symmetries. Moreover, the moment generating functions are produced. In mathematical terms, we derive the stochastic reflection principle on S^2 for the infectious disease transmission process. Reflection points can be extremely useful to calculate the distribution functions of certain temporal quantities related to the dynamics. Additionally, we evaluate boundary local times of first hitting of the outbreak for an epidemic or a hybrid endemic-epidemic model. Hence, biological carrier(s) of a virus (infectious individuals) are tracked at any given time on earth coordinates, and the path(s) defined by each infectious dynamical motion. In the following two chapters we present a thorough literature review and a state-of-the-art analysis in order to pose clearly our novel approach optimally among the various methodologies followed thus far.

The rest of paper is organized as follows: Section 2 provides a brief literature review, past and recent, of mathematical epidemiology. Section 3 presents the state-of-the-art, and focal concepts and term definitions required to introduce our novel model. It also states which category the new model falls into, according to the official taxonomy of the various methodologies already utilized so far in the relevant literature. Next, section 4 exposes in detail the mathematical formulation of the model. Lastly, Section 5 discusses proposed policies and future paths of research, and concludes.

2. Literature review

The beginning of mathematical modeling in epidemiology dates back to 1766, when Bernoulli developed a mathematical model to analyze the mortality of smallpox in England [22]. Bernoulli used his model to show that inoculation against the virus would increase the life expectancy at birth by about three years. A revision of the main findings and a presentation of the criticism by D'Alembert, appears recently in Dietz and Heesterbeek [23]. Lambert in 1772 as well as Laplace in 1812 extended the Bernoulli model by incorporating age-dependent parameters [24,25]. However, further systematic research was absent until the beginning of the twentieth century with the pioneering work of Ross in 1911, which is considered the inaugural study of modern mathematical epidemiology [26]. Ross used a set of equations to approximate the discrete-time dynamics of malaria via a mosquito-based pathogen transmission [27]. Importantly, the past century has witnessed the rapid emergence and development of substantial theories in epidemics. In 1927, Kermack and McKendrick [28] derived the celebrated threshold theorem, which is one of the key results in epidemiology. It predicts – depending on the transmission potential of the infection – the critical fraction of susceptibles in the population that must be exceeded if an epidemic is to occur. Kermack and McKendrick published three seminal papers, establishing what is called the deterministic compartmental epidemic modelling [29–31], wherein they addressed the mass–action incident in disease transmission cycles, assuming that the probability of infection of a susceptible is analogous to the number of its contacts with infected individuals. This deterministic representation was in line with the Law of Mass Action [32] introduced by Guldberg and Waage in 1864 and renders the basic most commonly used SIR model, which assumes homogeneous mixing of the contacts and conservation of the total population and low rates of interaction. MacDonald extended Ross's model to explain in depth the transmission process of malaria. Utilizing modern computer power, the mathematical model for the dynamics and the control of mosquito-transmitted pathogens provided robust results in real-word applications. Overall, the family of models they introduced is known by now as Ross–MacDonald models [33]. Moreover, the classic work of Bartlett [34] examined models and data to explore the factors that determine disease persistence in large populations. Arguably, a landmark book on mathematical modelling of epidemiological systems was published by Bailey [35] and highlighted the importance of public health decision making [36]. Given the diversity of infectious diseases studied since the middle of the 1950s, an impressive variety of epidemiological models have been developed. In addition, we should highlight the 19th century works by Enko [37–39], who first published a probabilistic model for describing the epidemic of measles, yet in discrete time. This model is the precursor of the popular Reed-Frost chain binomial model introduced by

Frost in 1928 in biostatistics' lectures at Johns Hopkins University [40]. It assumes that the infection spreads from an infected to a susceptible individual via a discrete time Markov chain, and set the basis of contemporary stochastic epidemic modelling, on which we will also focus in our present work.

Moving to the 21st century, we mention some interesting works; Xing et al., [41] introduced a mathematical model on H7N9 influenza among migrant and resident birds, domestic poultry and humans in China. In this study they concluded that temperature seasonality might be a source of the disease, yet they suggested for the first time that controlling markets could help controlling outbreaks. Lee and Pietz [42] developed a mathematical model for Zika virus using logistic growth in human populations. Sun et al., [43] proposed a transmission model for cholera in China and observed that reducing the spread requires extensive immunization coverage of the population. Nishiura et al. [44] developed a Zika mathematical model which exhibited the same dynamics as dengue fever, and Khan et al. [45] introduced a model whereby a saturation function describes well the typhoid fever dynamics. Gui and Zhang [46], developed a modified SIR model demonstrating nonlinearities in recovery rates. Their model exhibited a backward bifurcation phenomenon, which in turn implied that a plain reduction of the reproduction number less than one, was not rendered sufficient to stop the disease spread. Li et al. [47] constructed a multi-group brucellosis model and found out that the best way to contain the disease is to avoid cross infection of animal populations. Moreover, Yu and Lin [48] identified complex dynamical behaviour in epidemiological models and particularly the existence of multiple limit cycle bifurcations using a predictor-prey model. Shi et al. [49] proposed an HIV model with a saturated reverse function to describe the dynamics of infected cells. Additionally, Bonyah et al. [50] developed a SIR model to study the dynamics of buruli ulcer and suggested policy measures to control the disease. Lastly, Zhang et al. [51] developed a model with a latent period of the disease wherein the person is not infectious with saturated incidence rates and treatment functions, called SEIR epidemic model.

3. State-of-the-art analysis and definitions

The SIR model is the basic one used for modelling epidemics. Kermack and McKendrick created the model in 1927 [29] in which they considered a fixed population with only three compartments, susceptible (S), infected (I) and recovered (R). There are a large number of modifications of the SIR model, including those that include births and deaths, the SIR without or with vital dynamics, a model where upon recovery there is no immunity called SIS and where immunity lasts for a short period of time, called SIRS model. Furthermore, a model where there is a latent period of the disease and where the person is not infectious is indentified as SEIS and SEIR respectively, or where infants can be born with immunity is named MSIR. Also, we mention the herd immunity model [52,53].

Overall, the transmission mechanism from infective populations to susceptibles is not well-comprehended for many infectious diseases. Interactions in a population are very complex, hence it is extremely difficult to capture the large scale dynamics of disease spread without formal mathematical modeling. An epidemiological model uses microscopic effects - the role of an infectious individual - to forecast the macroscopic behavior of disease spread via a population.

Deterministic models do not incorporate any form of uncertainty and as such, they can be thought to account for the mean trend of a process, alone. On the other hand, stochastic models describe the mean trend as well as the variance structure of the underlying processes. Two basic types of stochasticity are commonly used: demographic and environmental. Within the context of demographic stochasticity, all individuals are subject to the same potential events with the exact same probabilities but differences in the fates of population individuals. Disease propagation in large populations obeys to the weak law of large numbers, thus effects of demographic stochasticity can be decreased significantly, and many times a deterministic model becomes more suitable. However, random events cannot be neglected and a stochastic model can be equally appropriate. Environmental stochasticity involves variations in the probability associated with an exogenous event. Model parameters of stochastic models are characterized by probability distributions, whilst for fixed parameter values deterministic models will always produce the same results, except when chaotic behaviour emerges.

In the classic SIR model it is assumed that the individuals leave the infectious class at a constant rate and even if this assumption seems most intuitive, it is not always the most realistic, regarding the duration individuals stay infective [54–56]. Usually, random variables describe the time of recovery since infection. For discrete random variables (e.g., number of individuals) it is easy to define a probability distribution, whilst for continuous variables the time of recovery since infection is modelled. Often, in this last category it is not possible to fix a probability as there is infinity of such times. Hence, we first define a cumulative distribution and then express a probability density function from this cumulative distribution. Infectious periods are exponentially distributed with a mean infectious duration, however as frequently real data does not back up this assumption, we rather use constant duration. To account for such more realistic distributions, the assumption that the probability of recovery does not depend on the time since infection, is often relaxed. Then, a common method of stages can be used to replace the infective compartment by a series of successive ones, each with an exponential distribution of the same parameter, leading to a total duration of the infectious period modelled by a gamma distribution [17].

Epidemic models presented above describe rapid outbreaks during which normally the host population is assumed to be in a constant state. For longer periods, deaths and births feed the population with new susceptibles, possibly allowing the disease to persist at a constant prevalence. This state renders an endemic state in the population [17]. In this case, we account for birth and death rate of the host population, whereby a good approximation is that the population size $N = S + I + R$ is constant. When deterministic dynamics prevail a threshold on the value of the basic reproduction number exists. Conditions regarding this number guarantee the disease persistence, but in epidemic models such persistence can be dependent upon the magnitude of the stochastic fluctuations around the steady-state equilibrium. Furthermore, many times diseases are in an endemo-epidemic state. As endemic models exhibit damped oscillations which converge toward an endemic equilibrium, this equilibrium can be weakly stable with perturbations (intrinsic or extrinsic), which excite and sustain the inherent oscillation behaviour [57]. This behaviour is due to heterogeneity that is added temporally to the coefficient of transmission, spatially in the context of meta-populations, or by cohorts for age-structured models. Lastly, heterogeneity can be added statistically in case of stochastic versions. For example, a stochastic version of the endemic SIR model can utilize a Markov process, in which the future is independent of the past given the present, with a state space defined by the number of individuals in each of the three classes, and changes in the state space characterized by probabilistic transition events. And as future events are independent on past events, the time to the next event follows a negative exponential distribution.

Over the years, a vast number of mathematical modeling approaches has been proposed, tackling the problem from different perspectives. The prevailing taxonomy proposed by Siettos and Russo [58] encompasses three general categories: (1) statistical

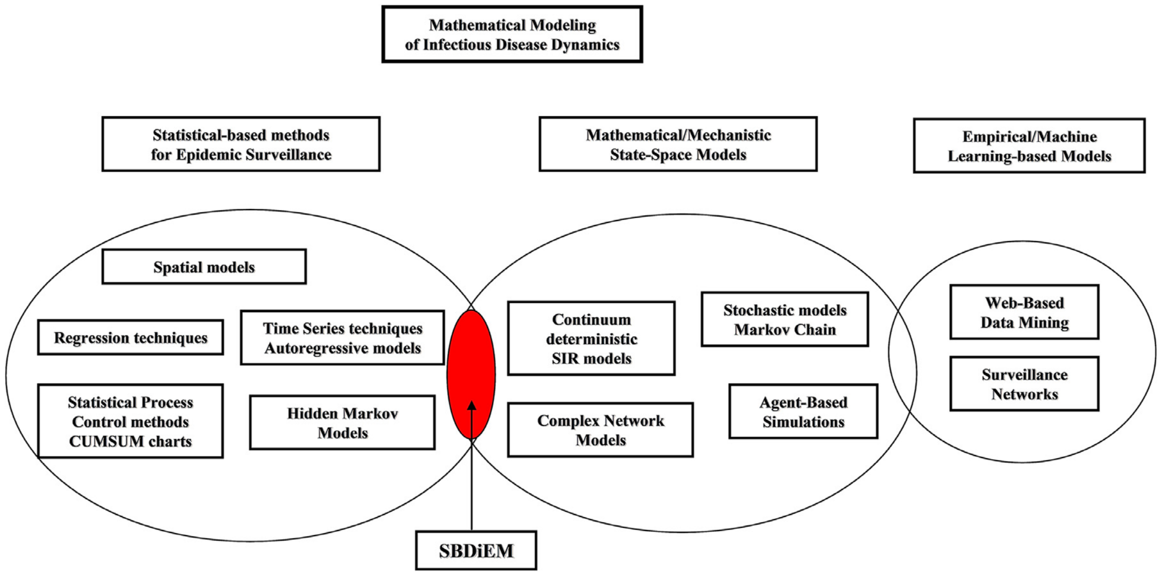

Fig. 1. Updated taxonomy of mathematical models for contagious diseases (source [58]). The new stochastic model lays in the intersection of categories (1) statistical methods and (2) state-space models of epidemic spreads.

methods of outbreaks and their identification of spatial patterns in real epidemics, (2) state-space models of the evolution of a "hypothetical" or on-going epidemic spread, and (3) machine learning methods, all utilized also for predictability purposes vis-à-vis an ongoing epidemic. In particular, the first category includes i) regression methods [59–64], ii) times series analysis, namely ARIMA and seasonal ARIMA approaches [65–68], iii) process control methods including cumulative sum (CUSUM) charts [69–74] and exponentially weighted moving average (EWMA) methods [75,76], as well as iv) Hidden Markov models (HMM) [77,78]. The second category incorporates i) "continuum" models in the form of differential and/or (integro)-partial differential equations [79–82], ii) discrete and continuous-time Markov-chain models [83–85], iii) complex network models which relax the hypotheses of the previous stochastic models that interactions among individuals are instantaneous and homogeneous [86–91], and iv) Agent-based models [92–95]. Lastly, the third category includes well-known machine learning approaches widely used in computer science, such as i) artificial neural networks [96], ii) web-based data mining [97,98] and iii) surveillance networks [99], to name a few.

For the first time in the relevant literature, we introduce a new stochastic model laying in the intersection of categories (1) and (2), called "**Stereographic Brownian Diffusion Epidemiology Model (SBDiEM)**". Fig. 1 presents a graphical overview of the models utilized so far, and the "positioning" of our novel approach for modelling infectious diseases.

4. Mathematical formulation

4.1. Preliminaries

4.1.1. The n-Sphere S^n

Definition 4.1. Let $n \in \mathbb{N}^* = \{1, 2, 3, \ldots\}$. The n-dimensional sphere S^n with center $(c_1, \ldots, c_{n+1})$ and radius $a > 0$ is (defined to be) the set of all points $x = (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}$ satisfying $(x_1 - c_1)^2 + \ldots + (x_{n+1} - c_{n+1})^2 = a^2$. Thus,

$$S^n = \{ (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \,\Big|\, (x_1 - c_1)^2 + \ldots + (x_{n+1} - c_{n+1})^2 = a^2 \}$$

4.1.2. Stereographic projection coordinates

Definition 4.2. We consider $\mathbb{R}^n \subset \mathbb{R}^{n+1}$ to be the hyperplane given by $x^{n+1} = 0$. For convenience, we will let $(x_1, x_2, \ldots, x_n, x_{n+1})$ be coordinates on $\mathbb{R}^{n+1}$ and $(\xi_1, \xi_2, \ldots, \xi_n)$ be coordinates on $\mathbb{R}^n \subset \mathbb{R}^{n+1}$. Let $S^n = \{ (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \,\big|\, x_1^2 + \ldots + x_n^2 + (x_{n+1} - a)^2 = a^2 \}$. The stereographic projection coordinates of S^n is the map $\Phi : S^n - \{0, 0, \ldots, 2a\} \to \mathbb{R}^n$ given by

$$\Phi(x_1, x_2, \ldots, x_n, x_{n+1}) = \left(\frac{2ax_1}{2a - x_{n+1}}, \ldots, \frac{2ax_n}{2a - x_{n+1}} \right).$$

This map defines coordinates $(\xi_1, \xi_2, \ldots, \xi_n)$ on S^n so that the point $(x_1, x_2, \ldots, x_n, x_{n+1})$ of S^n has coordinates $(\xi_1, \xi_2, \ldots, \xi_n)$, where

$$\xi_1 = \frac{2ax_1}{2a - x_{n+1}}, \ldots, \xi_n = \frac{2ax_n}{2a - x_{n+1}}.$$

The inverse map is given by

$$x_1 = \frac{4a^2 \xi_1}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2}, \ldots, x_n = \frac{4a^2 \xi_n}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2},$$
$$x_{n+1} = \frac{2a\left(\xi_1^2 + \ldots + \xi_n^2\right)}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2}.$$

4.1.3. Spherical coordinates

The points of the 2-sphere with center at the origin and radius a may also be described in spherical coordinates in the following way: $x_2 = a \sin\varphi$, where $0 \le \varphi < 2\pi$.

$S^2 = \{x = (a\cos\theta \sin\varphi, a\sin\theta \sin\varphi, a\cos\varphi) \in \mathbb{R}^3 \big| 0 \le \theta < 2\pi,\ 0 \le \varphi \le \pi\}$ i.e.

$$x_1 = a\cos\theta \sin\varphi$$

$$x_2 = a\sin\theta \sin\varphi$$

$$x_3 = a\cos\varphi,$$

where $0 \le \theta < 2\pi$ and $0 \le \varphi \le \pi$.

4.1.4. The Laplace-Beltrami operator

Definition 4.3. A C^∞ differentiable manifold of dimension n is a set M together with a family of one-to-one maps $x_\alpha \colon U_\alpha \to M$ of open sets $U_\alpha \subset \mathbb{R}^n$ into M such that

1. $\bigcup_\alpha x_\alpha(U_\alpha) = M$.
2. For each pair α, β with $x_\alpha(U_\alpha) \bigcap x_\beta(U_\beta) = W \neq \varnothing$, we have that $x_\alpha^{-1}(W), x_\beta^{-1}(W)$ are open sets in $\mathbb{R}^n$ and that $x_\beta^{-1} \circ x_\alpha$, $x_\alpha^{-1} \circ x_\beta$ are C^∞ differentiable maps.
3. The family $\{U_\alpha, x_\alpha\}$ is maximal relative to conditions 1 and 2.

Each pair (x_α, U_α) is called a coordinate chart on M. (For more details see [100])

Definition 4.4. A C^r function $f: M \to \mathbb{R}$, where M is a C^∞ differential manifold is a function f, such that $f \circ x_\alpha : U_\alpha \to \mathbb{R}$ is C^r for every cordinate chart (x_α, U_α) on M.

Let $g = [g_{ij}]$ be the Riemmanian metric tensor on a Riemmanian manifold M. This means that, in any coordinate chart $(x_1, x_2, \ldots, x_n)$ on M, the length element can be computed by

$$ds^2 = \sum_{j=1}^{n} \sum_{i=1}^{n} g_{ij} dx_i dx_j.$$

Given local coordinates $(x_1, \ldots, x_n)$, we can easily compute the matrix $g = [g_{ij}]$ by the inner product

$$g_{ij} = \frac{\partial x_a}{\partial x_i} \cdot \frac{\partial x_a}{\partial x_j}$$

(see [100]). We denote by g^{ij} the elements of the inverse matrix g^{-1}.

Definition 4.5. The Laplace-Beltrami operator Δ_M associated with the metric g is defined by

$$\Delta_M f = \frac{1}{\sqrt{det(g)}} \cdot \sum_i \frac{\partial}{\partial x_i} \left(\sqrt{det(g)} \cdot \sum_j g^{ij} \frac{\partial f}{\partial x_j} \right), \tag{4.1}$$

where f is a C^r function on M.

In this work we are interested in the case where $M = S^2$, i.e., the 2 -dimensional sphere. We will denote the corresponding Laplace-Beltrami operator of S^2 by Δ_2 or just Δ using the spherical coordinates. If $M = S^2$, i.e.

$$M = S^2 = \{x = (a\cos\theta \sin\varphi, a\sin\theta\sin\varphi, a\cos\varphi) \in \mathbb{R}^3 \,|\, 0 \le \theta < 2\pi,\ 0 \le \varphi \le \pi\},$$

we have

$$x_\theta = \frac{\partial x}{\partial \theta} = (-a\sin\theta\sin\varphi, a\cos\theta\sin\varphi, 0)$$

$$x_\varphi = \frac{\partial x}{\partial \varphi} = (a\cos\theta\cos\varphi, a\sin\theta\cos\varphi, -a\sin\varphi)$$

$$g = [g_{ij}] = \begin{pmatrix} x_\theta x_\theta & x_\theta x_\varphi \\ x_\varphi x_\theta & x_\varphi x_\varphi \end{pmatrix},$$

i.e.,

$$g = [g_{ij}] = \begin{pmatrix} a^2 \sin^2\varphi & 0 \\ 0 & a^2 \end{pmatrix}$$

and

$$g^{-1} = [g^{ij}] = \begin{pmatrix} \frac{1}{a^2 \sin^2\varphi} & 0 \\ 0 & \frac{1}{a^2} \end{pmatrix}.$$

Hence the Laplace-Beltrami operator of a smooth function f on S^2 is

$$\Delta_2 f = \frac{1}{a^2 \sin\varphi} \sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(a^2 \sin\varphi \sum_{j=1}^{2} g^{ij} \frac{\partial f}{\partial x_j} \right), \tag{4.2}$$

where $x_1 = \theta$ and $x_2 = \varphi$. Thus

$$\Delta_2 f = \frac{1}{a^2 \sin\varphi} \sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left[a^2 \sin\varphi \left(g^{i1} f_\theta + g^{i2} f_\varphi \right) \right]$$

or

$$\Delta_2 f = \frac{1}{a^2 \sin\varphi} \left(\frac{f_{\theta\theta}}{\sin\varphi} + f_\varphi \cos\varphi + f_{\varphi\varphi} \sin\varphi \right). \tag{4.3}$$

In case where the function f is independent of θ the Laplace-Beltrami operator of f is

$$\Delta_2 f = \frac{1}{a^2 \sin\varphi} \left(f_\varphi \cos\varphi + f_{\varphi\varphi} \sin\varphi \right). \tag{4.4}$$

Generally the Laplace-Beltrami operator of a smooth function f on S^n is

$$\Delta_n f = \frac{1}{\sqrt{\det(g)}} \cdot \sum_{i=1}^{n} \frac{\partial}{\partial \theta_i} \left(\sqrt{\det(g)} \cdot \sum_{j=1}^{n} g^{ij} \frac{\partial f}{\partial \theta_j} \right), \tag{4.5}$$

where

$$\det(g) = a^{2n} \prod_{k=2}^{n} (\sin\theta_k)^{2(k-1)}, \tag{4.6}$$

$$g^{ij} = 0, \text{ if } i \neq j, \quad g^{ii} = \frac{1}{a^2 \sin^2\theta_{i+1} \cdot \ldots \cdot \sin^2\theta_n} \quad \text{and } \theta_n = \varphi.$$

If f is independent of $\theta_1, \theta_2, \ldots, \theta_{n-1}$, the Laplace Beltrami operator of f is

$$\Delta_n f = \frac{1}{a^2} \left((n-1)\cot\varphi \cdot \frac{\partial f}{\partial \varphi} + \frac{\partial^2 f}{\partial \varphi^2} \right). \tag{4.7}$$

Using Stereographic projection coordinates, if $M = S^n$, i.e.

$$M = S^n = \left\{ x = \left(\frac{4a^2 \xi_1}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2}, \ldots, \frac{4a^2 \xi_n}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2}, \frac{2a(\xi_1^2 + \ldots + \xi_n^2)}{\xi_1^2 + \ldots + \xi_n^2 + 4a^2} \right) \in \mathbb{R}^{n+1} \middle|, \xi_1, \ldots, \xi_n \in \mathbb{R} \right\},$$

we have

$$x_{\xi_k} = \frac{\partial x}{\partial \xi_k} = \left(\frac{-8a^2 \xi_1 \xi_k}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2}, \ldots, \frac{-8a^2 \xi_{k-1} \xi_k}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2}, \frac{4a^2 \left(\sum_{i=1}^{n} \xi_i^2 - 2\xi_k^2 + 4a^2 \right)}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2}, \right.$$
$$\left. \frac{-8a^2 \xi_{k+1} \xi_k}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2}, \ldots, \frac{-8a^2 \xi_n \xi_k}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2}, \frac{16a^3 \xi_k}{(\xi_1^2 + \ldots + \xi_n^2 + 4a^2)^2} \right).$$

Hence

$$g_{ii}=\frac{16a^4}{(\xi_1^2+\ldots+\xi_n^2+4a^2)^2}\quad \text{and}\quad g_{ij}=0,\quad \text{if}\quad i\neq j.$$

Thus we have

$$g^{ii}=\frac{(\xi_1^2+\ldots+\xi_n^2+4a^2)^2}{16a^4},\quad g^{ij}=0,\quad \text{if}\quad i\neq j,\quad \text{and}$$

$$\sqrt{det(g)}=\frac{(4a^2)^n}{\left(\xi_1^2+\ldots+\xi_n^2+4a^2\right)^n}.$$

Therefore, the Laplace Beltrami operator of a smooth function f on S^2, using Stereographic projection coordinates is

$$\Delta_2 f=\frac{\left(\xi_1^2+\xi_2^2+4a^2\right)^2}{16a^4}\left(\frac{\partial^2 f}{\partial\xi_1^2}+\frac{\partial^2 f}{\partial\xi_2^2}\right) \tag{4.8}$$

4.1.5. Brownian motion on a riemannian manifold

Definition 4.6. Let M be a Riemannian manifold (see **definition 1.5**) and Δ its corresponding Laplace-Beltrami operator. Any function $P(t, x, y)$ on $(0, \infty)\times M\times M$ satisfying the differential equation

$$\frac{\partial P}{\partial t}-\frac{1}{2}\Delta_x P=0, \tag{4.9}$$

where Δ_x is Δ acting on the x-variables and the initial condition

$$P(t,x,y)\to\delta_x(y)\quad \text{as},\quad t\to 0^+ \tag{4.10}$$

(where $\delta_x(y)$ is the delta mass at $x\in M$) is called a fundamental solution of the heat Eq. (4.9) on M.

The smallest positive fundamental solution of the heat Eqs. (4.9) and (4.10) is the heat kernel on M. It has been proved by J. Dodziak [101], that the heat kernel always exists, and is smooth in (t, x, y). Moreover the heat kernel possesses the following properties.

1. Symmetry in x, y, that is

   $$P(t,x,y)=P(t,y,x)$$

2. The semigroup identity: For any $s\in(0, t)$

   $$P(t,x,y)=\int_M P(s,x,z)P(t-s,z,y)d\mu(z),$$

   where $d\mu$ is the area measure element of M. In polar coordinates $d\mu=\sqrt{|g|}d\theta_1\ldots\theta_n$, where $\theta_n=\varphi$ and $|g|$ is given by (4.6).
3. The total mass inequality, i.e., for all $t>0$ and $x\in M$

   $$\int_M P(t,x,y)d\mu(y)\leq 1. \tag{4.11}$$

In case where M is compact and smooth, there is only one solution of (4.9) and (4.10) which is positive and satisfies

$$\int_M P(t,x,y)d\mu(y)=1 \tag{4.12}$$

Definition 4.7. A process X_t, $t\geq 0$ is a Markov process if for any t, $s\geq 0$, the conditional distribution of X_{t+s}, given the information about the process up to time t, is the same as the conditional distribution of X_{t+s}, given X_t.

Definition 4.8. The Brownian motion X_t, $t\geq 0$, on a Riemannian manifold M is a Markov process with transition density function $P(t, x, y)$ the heat kernel associated with the Laplace-Beltrami operator.

Remark 4.1. In the case where $M=S^n$, $n\geq 2$, the transition density function $P(t, x, y)$ of the Brownian motion X_t depends only on t and $d(x, y)$, the distance between x and y. Thus in spherical coordinates it depends on t and the angle φ between x and y. Hence, the transition density function of the Brownian motion can be written as

$$P(t,x,y)=p(t,\varphi), \tag{4.13}$$

where $p(t, \varphi)$ is the solution of

$$\frac{\partial p}{\partial t}=\frac{1}{2}\Delta_n p=\frac{1}{2a^2}\left((n-1)\cot\varphi\cdot\frac{\partial p}{\partial\varphi}+\frac{\partial^2 p}{\partial\varphi^2}\right) \tag{4.14}$$

and

$$\lim_{t\to 0^+} aA_{n-1}p(t,\varphi)\cdot\sin^{n-1}(\varphi)=\delta(\varphi). \tag{4.15}$$

Here $\delta(\,\cdot\,)$ is the Dirac delta function on $\mathbb{R}$ and A_n denotes the area of the n-dimensional sphere S^n with radius a. It is well known that [102]

$$A_n=\frac{2\pi^{\frac{n+1}{2}}a^n}{\Gamma(\frac{n+1}{2})}, \tag{4.16}$$

where $\Gamma(\,\cdot\,)$ is the Gamma function. More precisely

$$A_n=\frac{2\pi^{\frac{n+1}{2}}a^n}{(\frac{n-1}{2})!}\qquad \text{for n odd} \tag{4.17}$$

$$A_n=\frac{2^n(\frac{n}{2}-1)!\pi^{\frac{n}{2}}a^n}{(n-1)!}\qquad \text{for n even} \tag{4.18}$$

Remark 4.2. The fact that S^n is a compact and smooth manifold implies that (4.14) and (4.15) has a unique positive solution which also satisfies

$$\int_{S^n} P(t,x,y)d\mu(y)=1. \tag{4.19}$$

Furthermore, as $t\to\infty$, $P(t, x, y)$ approaches the uniform density on S^n, i.e. $P(t, x, y)\to c$, where

$$c=\frac{1}{A_n}.$$

In the sequel for typographical convenience we will write X_t instead of $\{X_t\}_t\geq 0$.

4.2. Transition density function $p(t,\varphi)$ of X_t, $t>0$

In this section we shall represent the transition density function $p(t, \varphi)$ of the position $X(t)$ of a biological carrier (infected individual) of virus Y at any given time t. For the next sections we suppose that the infected individual is at position $X(t)$ at any given time t, namely the path defined by its motion is considered infectious. $X_t, t\geq 0$ describes a Brownian motion on a 2-dimensional sphere S^2 of radius a. From the (4.14), (4.15) and (4.17) the transition density function $p(t, \varphi)$ of X_t is the unique solution of

$$\frac{\partial p}{\partial t}=\frac{1}{2a^2\sin\varphi}\left(\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}\sin\varphi+\frac{\partial p}{\partial\varphi}\cos\varphi\right) \tag{4.20}$$

and

$$\lim_{t\to 0^+} 2\pi a^2\sin\varphi\cdot p(t,\varphi)=\delta(\varphi). \tag{4.21}$$

The solution of the diffusion equation

$$\frac{\partial K(t,\varphi)}{\partial t}=\frac{1}{\sin\varphi}\left(\cos\varphi\frac{\partial K(t,\varphi)}{\partial\varphi}+\sin\varphi\frac{\partial^2 K(t,\varphi)}{\partial\varphi^2}\right) \tag{4.22}$$

with initial condition

$$\lim_{t\to 0^+} 2\pi\,\sin(\varphi)K(t,\varphi)=\delta(\varphi) \tag{4.23}$$

is given by the function

$$K(t,\varphi)=\frac{1}{4\pi}\sum_{n\in\mathbb{N}}(2n+1)\exp\left(-n(n+1)\sqrt{2t}\right)P_n^0(\cos\varphi). \quad (4.24)$$

see[103]. Here P_n^0, $n=0,1,2,\dots$, is the associated Legendre polynomials of order zero, i.e.

$$P_n^0(x)=\frac{1}{2^n n!}\cdot\frac{d^n}{dx^n}\left[(x^2-1)^n\right] \quad (4.25)$$

This fact implies the following

Proposition 4.1. *The transition density function of the Brownian motion X_t, $t\geq 0$ on S^2 with radius a it is given by the function*

$$p(t,\varphi)=\frac{1}{4\pi a^2}\sum_{n\in\mathbb{N}}(2n+1)\exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right)P_n^0(\cos\varphi) \quad (4.26)$$

Proof. First we prove that $p(t, \varphi)$ satisfies the differential equation

$$\frac{\partial p}{\partial t}=\frac{1}{2a^2\sin\varphi}\left(\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}\sin\varphi+\frac{\partial p}{\partial\varphi}\cos\varphi\right).$$

We have that

$$p(t,\varphi)=\frac{1}{a^2}K\left(\frac{t}{2a^2},\varphi\right),$$

where $K(t, \varphi)$ is given by the (4.24), therefore

$$\frac{\partial p(t,\varphi)}{\partial t}=\frac{1}{2a^4}\frac{\partial K}{\partial t},$$

$$\frac{\partial p(t,\varphi)}{\partial\varphi}=\frac{1}{a^2}\frac{\partial K}{\partial\varphi} \quad \text{and}$$

$$\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}=\frac{1}{a^2}\frac{\partial^2 K}{\partial\varphi^2}.$$

However from the (4.22)

$$\frac{\partial K}{\partial t}=\frac{1}{\sin\varphi}\left(\cos\varphi\frac{\partial K}{\partial\varphi}+\sin\varphi\frac{\partial^2 K}{\partial\varphi^2}\right), \quad (4.27)$$

hence

$$2a^4\frac{\partial p(t,\varphi)}{\partial t}=\frac{1}{\sin\varphi}\left(a^2\cos\varphi\frac{\partial p(t,\varphi)}{\partial\varphi}+a^2\sin\varphi\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}\right),$$

i.e.

$$\frac{\partial p(t,\varphi)}{\partial t}=\frac{1}{2a^2\sin\varphi}\left(\cos\varphi\frac{\partial p(t,\varphi)}{\partial\varphi}+\sin\varphi\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}\right).$$

Furthermore $p(t, \varphi)$ satisfies the

$$\lim_{t\to 0^+}2\pi\sin(\varphi)p(t,\varphi)=\lim_{t\to 0^+}2\pi a^2\frac{1}{a^2}\sin(\varphi)K\left(\frac{t}{2a^2},\varphi\right)$$

and if we set $u=\frac{t}{2a^2}$ we imply that

$$\lim_{t\to 0^+}2\pi\sin(\varphi)K\left(\frac{t}{2a^2},\varphi\right)=\lim_{u\to 0^+}2\pi\sin(\varphi)K(u,\varphi)=\delta(\varphi).$$

Therefore

$$\lim_{t\to 0^+}2\pi a^2\sin(\varphi)p(t,\varphi)=\delta(\varphi).$$

and this complete the proof. □

4.2.1. Stochastic differential equation of the brownian motion in local coordinates

We recall the following well-known fact

Theorem 4.1. *Let*

$$\sigma(x)=\left[\sigma_{jk}(x)\right], \quad \text{with} \quad 1\leq j\leq n, 1\leq k\leq m,$$

be such that $a(x)=\sigma(x)\cdot\sigma^T(x)$ is positive definite. If X_t is the Ito diffusion process

$$dX_t=b(X_t)dt+\sigma(X_t)dB_t, \quad (4.28)$$

then, its generator A is given by the formula

$$Af(x)=\sum_i b_i(x)\frac{\partial f}{\partial x_i}+\frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(x)\frac{\partial^2 f}{\partial x_i\partial x_j}.$$

Conversely, the operator A given above is the generator of diffusion (4.28). For the proof see [104].

Case of spherical coordinates

The generator of Brownian motion on S^2 in spherical coordinates is

$$Af=\frac{1}{2}\Delta_2 f,$$

i.e.

$$Af=\frac{\cos\varphi}{2a^2\sin\varphi}\frac{\partial f}{\partial\varphi}+\frac{1}{2}\left(\frac{1}{a^2\sin^2\varphi}\frac{\partial^2 f}{\partial\theta^2}+\frac{1}{a^2}\frac{\partial^2 f}{\partial\varphi^2}\right).$$

Therefore, the Brownian motion on S^2 in spherical coordinates is the solution of the stochastic differential equation

$$dX_t=\left(0,\frac{\cos\varphi(t)}{2a^2\sin\varphi(t)}\right)dt+\begin{pmatrix}\frac{1}{a\sin\varphi(t)} & 0\\ 0 & \frac{1}{a}\end{pmatrix}\begin{pmatrix}dB_1(t)\\ dB_2(t)\end{pmatrix},$$

where

$$X_t=(\theta(t),\varphi(t)).$$

Case of sterographic projection coordinates

Expressed in stereographic projection coordinates, the generator of Brownian motion on S^2 is

$$Af=\frac{1}{2}\frac{\left(\xi_1^2+\xi_2^2+4a^2\right)^2}{16a^4}\left(\frac{\partial^2 f}{\partial\xi_1^2}+\frac{\partial^2 f}{\partial\xi_2^2}\right).$$

Hence, the Brownian motion on S^2 in stereographic projection coordinates is the solution of the stochastic differential equation

$$dX_t=\begin{pmatrix}\frac{\left(x_1(t)^2+x_2^2(t)+4a^2\right)}{4a^2} & 0\\ 0 & \frac{\left(x_1^2(t)+x_2^2(t)+4a^2\right)}{4a^2}\end{pmatrix}\begin{pmatrix}dB_1(t)\\ dB_2(t)\end{pmatrix}, \quad (4.29)$$

where

$$X_t=(x_1(t),x_2(t)).$$

4.3. Expectations of exit times of X(t)

We recall some basic definitions.

Definition 4.9. A measurable space $\{\Omega,\mathcal{F}\}$ is said to be equipped with a filtration $\{\mathcal{F}_t\}$, $t\in[0,+\infty)$, if for every $t\geq 0$ $\{\mathcal{F}_t\}$ is a σ-algebra of subsets of Ω such that $\mathcal{F}_t\subset\mathcal{F}$ and for every $t_1,t_2\in[0,+\infty)$ such that $t_1<t_2$, we have that $\mathcal{F}_{t_1}\subset\mathcal{F}_{t_2}$. (i.e. $\{\mathcal{F}_t\}$ is an increasing family of sub σ-algebras of $\mathcal{F}$).

Definition 4.10. Let us consider a measurable space $\{\Omega,\mathcal{F}\}$ equipped with a filtration $\{\mathcal{F}_t\}$. A random variable T is a stopping time with respect to the filtration $\{\mathcal{F}_t\}$, if for every $t\geq 0$

$$\{\omega\in\Omega|T(\omega)\leq t\}\in\mathcal{F}_t.$$

Let X_t be the Brownian motion in S^n and $D \subset S^n$ a domain. Then

$$T = \inf\{t \geq 0 | X_t \notin D\}$$

is a stopping time with respect to $\mathcal{F}_t = \sigma\{X_s | 0 \leq s \leq t\}$, called the exit time on ∂D.

Proposition 4.2. *Let $\varphi_0 \in [0, \pi)$ be fixed. We consider the set D in S^2, such that*

$$D = \{(\theta, \varphi) | \theta \in [0, 2\pi) \quad \text{and} \quad \varphi \in [0, \varphi_0)\}.$$

Of course,

$$\partial D = \{(\theta, \varphi) | \theta \in [0, 2\pi) \quad \text{and} \quad \varphi = \varphi_0\}.$$

If X_t is the position of the biological carrier of the virus Y at a given time t starting at the point

$$A = (\theta, \varphi) \in D,$$

and

$$T = \inf\{t \geq 0 | X_t \notin D\},$$

then the expectation of T is given by

$$E^A[T] = 2a^2 \ln\left(\frac{1 + \cos\varphi}{1 + \cos\varphi_0}\right). \tag{4.30}$$

Proof. Based on [105],

$$u(\theta, \varphi) = E^A[T]$$

we have the unique solution of the differential equation

$$\frac{1}{2}\Delta_2 u = -1, \tag{4.31}$$

with boundary condition as

$$u(\theta, \varphi_0) = 0.$$

Here Δ_2 is the Laplace-Beltrami operator on S^2. By symmetry of D, it follows that the expectation value of T is independent of θ. From (4.4) the differential Eq. (4.31) takes the form

$$\frac{1}{2a^2}\left[\cot(\varphi)\frac{du}{d\varphi} + \frac{d^2 u}{d\varphi^2}\right] = -1, \tag{4.32}$$

with boundary condition

$$u(\varphi_0) = 0. \tag{4.33}$$

Set

$$f(\varphi) = \frac{du}{d\varphi}.$$

hence from (4.32)

$$\frac{1}{2a^2}\left[\cot(\varphi) f(\varphi) + \frac{df(\varphi)}{d\varphi}\right] = -1,$$

or

$$\cos(\varphi) f(\varphi) + \sin(\varphi)\frac{df(\varphi)}{d\varphi} = -2a^2 \sin(\varphi),$$

Thus

$$f(\varphi) = -\frac{2a^2}{\sin\varphi}\int_0^{\varphi} \sin\omega d\omega + \frac{c_1}{\sin\varphi}.$$

Therefore,

$$u(\varphi) = -2a^2 \int_{\varphi_0}^{\varphi} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx + c_1 \int_{\varphi_0}^{\varphi} \frac{1}{\sin x} dx + c_2. \tag{4.34}$$

However (see [104])

$$u(\varphi) = E^A[T] < \infty, \quad \text{for} \quad \varphi \in [0, \varphi_0)$$

hence

$$c_1 = 0.$$

Furthermore, we have

$$u(\varphi_0) = 0, \quad \text{i.e.} \quad c_2 = 0$$

thus,

$$u(\varphi) = 2a^2 \int_{\varphi}^{\varphi_0} \frac{\int_0^x (\sin\omega) d\omega}{(\sin x)} dx.$$

Consequently,

$$E^A[T] = 2a^2 \int_{\varphi}^{\varphi_0} \frac{\int_0^x (\sin\omega) d\omega}{(\sin x)} dx.$$

Thus

$$E^A[T] = 2a^2 \int_{\varphi}^{\varphi_0} \frac{1 - \cos x}{\sin x} dx,$$

or

$$E^A[T] = 2a^2 \left(\int_{\varphi}^{\varphi_0} \frac{1}{\sin x} dx - \int_{\varphi}^{\varphi_0} (\cot x) dx\right),$$

hence

$$E^A[T] = 2a^2 \Big[\ln\left(\tan\left(\frac{\varphi_0}{2}\right)\right) - \ln\left(\tan\left(\frac{\varphi}{2}\right)\right) - \ln(\sin\varphi_0) + \ln(\sin\varphi)\Big].$$

Finally,

$$E^A[T] = 2a^2 \ln\left(\frac{1 + \cos\varphi}{1 + \cos\varphi_0}\right). \tag{4.35}$$

□

Proposition 4.3. *Let $\varphi_1, \varphi_2 \in (0, \pi)$, such that $\varphi_1 < \varphi_2$, are both fixed. We consider the set D in S^2, such that*

$$D = \{(\theta, \varphi) | \theta \in [0, 2\pi) \quad \text{and} \quad \varphi \in (\varphi_1, \varphi_2)\}.$$

We have,

$$\partial D = \{(\theta, \varphi) | \theta \in [0, 2\pi), \quad \text{and} \quad \varphi = \varphi_1 \quad \text{or} \quad \varphi = \varphi_2\}.$$

Let X_t be the position of the infectious individual Y is at a given time t starting at the point

$$A = (\theta, \varphi) \in D,$$

and

$$T = \inf\{t \geq 0 | X_t \notin D\},$$

then the expectation of T is given by

$$E^A[T] = \frac{4a^2}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}\left[\ln\left(\frac{\cos\left(\frac{\varphi_1}{2}\right)}{\cos\left(\frac{\varphi_2}{2}\right)}\right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)}\right) - \ln\left(\frac{\cos\left(\frac{\varphi_1}{2}\right)}{\cos\left(\frac{\varphi}{2}\right)}\right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi_2}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)}\right)\right] \tag{4.36}$$

Proof. According to [105], $E^{\varphi}[t]$ satisfies the Poisson equation on D with Dirichlet boundary data. By uniqueness

$$u(\theta, \varphi) = E^A[T]$$

is the unique solution of the differential Eq. (4.31), i.e.,

$$\frac{1}{2}\Delta_2 u = -1,$$

with boundary condition

$$u(\theta, \varphi_1) = u(\theta, \varphi_2) = 0.$$

Here Δ_2 is the Laplace-Beltrami operator on S^2. By the symmetry of D, it follows that the expectation value of T is independent of θ. From (4.4) the differential Eq. (4.31) takes the form (4.32) with boundary condition

$$u(\theta, \varphi_1) = u(\theta, \varphi_2) = 0. \tag{4.37}$$

Hence from (4.34)

$$u(\varphi) = -2a^2 \int_{\varphi_1}^{\varphi} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx + c_1 \int_{\varphi_1}^{\varphi} \frac{1}{\sin x} dx + c_2.$$

However

$$u(\theta, \varphi_1) = u(\theta, \varphi_2) = 0,$$

i.e.

$$-2a^2 \int_{\varphi_1}^{\varphi_1} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx + c_1 \int_{\varphi_1}^{\varphi_1} \frac{1}{\sin x} dx + c_2 = 0$$

and

$$-2a^2 \int_{\varphi_1}^{\varphi_2} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx + c_1 \int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx + c_2 = 0.$$

Thus

$$c_1 = 2a^2 \frac{\int_{\varphi_1}^{\varphi_2} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx}{\int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx}$$

and

$$c_2 = 0.$$

Consequently,

$$E^A[T] = 2a^2 \left(\int_{\varphi}^{\varphi_1} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx + \frac{\int_{\varphi_1}^{\varphi_2} \frac{\int_0^x \sin\omega d\omega}{\sin x} dx}{\int_{\varphi 1}^{\varphi_2} \frac{1}{\sin x} dx} \cdot \int_{\varphi_1}^{\varphi} \frac{1}{\sin x} dx \right).$$

namely,

$$E^A[T] = 2a^2 \left(\int_{\varphi_1}^{\varphi} (\cot x) dx - \frac{\int_{\varphi 1}^{\varphi_2} (\cot x) dx}{\int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx} \cdot \int_{\varphi_1}^{\varphi} \frac{1}{\sin x} dx \right),$$

hence

$$E^A[T] = \frac{2a^2}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)} \left[\ln\left(\frac{\sin\left(\frac{\varphi}{2}\right)\cos\left(\frac{\varphi}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)\cos\left(\frac{\varphi_1}{2}\right)} \right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi_2}{2}\right)\cos\left(\frac{\varphi_1}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)\cos\left(\frac{\varphi_2}{2}\right)} \right) - \ln\left(\frac{\sin\left(\frac{\varphi_2}{2}\right)\cos\left(\frac{\varphi_2}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)\cos\left(\frac{\varphi_1}{2}\right)} \right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi}{2}\right)\cos\left(\frac{\varphi_1}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)\cos\left(\frac{\varphi}{2}\right)} \right) \right],$$

from which we imply

$$E^A[T] = \frac{4a^2}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)} \left[\ln\left(\frac{\cos\left(\frac{\varphi_1}{2}\right)}{\cos\left(\frac{\varphi_2}{2}\right)} \right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)} \right) - \ln\left(\frac{\cos\left(\frac{\varphi_1}{2}\right)}{\cos\left(\frac{\varphi}{2}\right)} \right) \cdot \ln\left(\frac{\sin\left(\frac{\varphi_2}{2}\right)}{\sin\left(\frac{\varphi_1}{2}\right)} \right) \right] \tag{4.38}$$

□

Proposition 4.4. *We consider the 2-dimensional sphere S^2 of radius a. Let two circles pass through the North pole, such that in stereographic coordinates are represented by the parallel lines $\xi_2 = b$ and $\xi_2 = c$, where $b, c \in \mathbb{R}$, say $b < c$. We consider the set D in S^2, whose stereographic projection is*

$$D = \{(\xi_1, \xi_2) | \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 \in (b, c)\}.$$

Of course

$$\partial D = \{(\xi_1, \xi_2) | \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 = b \quad \text{or} \quad \xi_2 = c\}.$$

If X_t is the position of the carrier of virus Y at a given time t starting at the point A, where the stereogrpaphic projection coordinates of A are

$$(\xi_1, \xi_2) \in D.$$

and

$$T = \inf\{t \geq 0 | X_t \in D\},$$

then,

$$E^A[T] = f(\xi_1, \xi_2) - 2a^2 \ln\left(\xi_1^2 + \xi_2^2 + 4a^2\right), \tag{4.39}$$

where

$$f(\xi_1, \xi_2) = \frac{1}{\pi} \int_0^\infty \frac{g(\xi, c) \exp\left(\frac{\pi\xi_1}{c-b}\right) \sin\left(\frac{\pi(\xi_2 - b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right) \sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right) \cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \eta\right)^2} d\eta$$

$$\frac{1}{\pi} \int_0^\infty \frac{g(\eta, b) \exp\left(\frac{\pi\xi_1}{c-b}\right) \sin\left(\frac{\pi(\xi_2 - b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right) \sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right) \cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) - \eta\right)^2} d\eta \tag{4.40}$$

and

$$g(\xi, t) = 2a^2 \ln\left(\frac{(c-b)^2 \ln^2 |\xi|}{\pi^2} + t^2 + 4a^2 \right) \tag{4.41}$$

Proof. As we have seen the function

$$E^A[T] = U(\xi_1, \xi_2)$$

satisfies the differential equation

$$\frac{1}{2} \Delta_2 U = -1$$

with boundary conditions

$$U(\xi_1, b) = U(\xi_1, c) = 0.$$

Here, Δ_2 is the Laplace-Beltrami operator on S^2 expressed in stereographic projection coordinates. Hence, the differential equation takes the form

$$\frac{1}{2} \frac{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}{16a^4} \cdot \left(\frac{\partial^2 U}{\partial \xi_1^2} + \frac{\partial^2 U}{\partial \xi_2^2} \right) = -1,$$

or

$$\frac{\partial^2 U}{\partial \xi_1^2} + \frac{\partial^2 U}{\partial \xi_2^2} = -\frac{32a^4}{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}. \tag{4.42}$$

However the function

$$U_1(\xi_1, \xi_2) = -2a^2 \ln(\xi_1^2 + \xi_2^2 + 4a^2)$$

satisfies the differential Eq. (4.42). Thus

$$U(\xi_1, \xi_2) = -2a^2 \ln(\xi_1^2 + \xi_2^2 + 4a^2) + f(\xi_1, \xi_2)$$

where $f(\xi_1, \xi_2)$ satisfies

$$\frac{\partial^2 f}{\partial \xi_1^2} + \frac{\partial^2 f}{\partial \xi_2^2} = 0,$$

with boundary conditions

$$f(\xi_1, b) = 2a^2 \ln(\xi_1^2 + b^2 + 4a^2)$$

and

$$f(\xi_1, c) = 2a^2 \ln(\xi_1^2 + c^2 + 4a^2).$$

If we take the transformation of variables $x = \xi_1$ and $y = \xi_2 - b$ and set the function $\phi(x, y) = f(\xi_1, \xi_2)$, then $\phi(x, y)$ we satisfy

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0,$$

with boundary conditions

$$\phi(x,0) = 2a^2 \ln(x^2 + b^2 + 4a^2)$$

and

$$\phi(x,\beta) = 2a^2 \ln(x^2 + c^2 + 4a^2)$$

where $\beta = c - b$. Now let $z = x + yi$ and $w = \exp\left(\frac{\pi z}{\beta}\right)$, i.e. $z = \frac{\beta \ln w}{\pi}$. Thus, if $w = u + vi$, $u, v \in \mathbb{R}$ then

$$u = \exp\left(\frac{\pi x}{\beta}\right)\cos\left(\frac{\pi y}{\beta}\right) \quad \text{and} \quad v = \exp\left(\frac{\pi x}{\beta}\right)\sin\left(\frac{\pi y}{\beta}\right). \tag{4.43}$$

Introducing the function $\psi(u, v) = \phi(x, y)$. It follows that $\psi(u, v)$ satisfies

$$\frac{\partial^2 \psi}{\partial u^2} + \frac{\partial^2 \psi}{\partial v^2} = 0,$$

with boundary conditions

$$\psi(u,0) = 2a^2 \ln\left(\frac{\beta^2 \ln^2 u}{\pi^2} + b^2 + 4a^2\right), \quad \text{for} \quad u > 0$$

and

$$\psi(u,0) = 2a^2 \ln\left(\frac{\beta^2 \ln^2 |u|}{\pi^2} + c^2 + 4a^2\right), \quad \text{for} \quad u < 0.$$

This is the standard Dirichlet boundary value problem for the half line, and it is well known that (see e.g. [106]) its solution is given by the Poisson integral formula for the half-plane:

$$\psi(u,v) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v\psi(\xi,0)}{v^2 + (u-\xi)^2} d\xi,$$

or

$$\psi(u,v) = \frac{1}{\pi}\int_{-\infty}^{0} \frac{v g(\xi,c)}{v^2 + (u-\xi)^2} d\xi + \frac{1}{\pi}\int_{0}^{\infty} \frac{v g(\xi,b)}{v^2 + (u-\xi)^2} d\xi,$$

where

$$g(\xi,t) = 2a^2 \ln\left(\frac{\beta^2 \ln^2 |\xi|}{\pi^2} + t^2 + 4a^2\right).$$

Notice that $g(-\xi,t) = g(\xi,t)$. Hence,

$$\psi(u,v) = \frac{1}{\pi} v \int_{0}^{\infty} \left(\frac{g(\xi,c)}{v^2 + (u+\xi)^2} + \frac{g(\xi,b)}{v^2 + (u-\xi)^2}\right) d\xi,$$

where u, v are given in (4.43). Therefore

$$\phi(x,y) = \frac{1}{\pi}\exp\left(\frac{\pi x}{\beta}\right)\sin\left(\frac{\pi y}{\beta}\right)\int_0^{\infty} \frac{g(\eta,c)}{\exp\left(\frac{2\pi x}{\beta}\right)\sin^2\left(\frac{\pi y}{\beta}\right) + \left(\exp\left(\frac{\pi x}{\beta}\right)\cos\left(\frac{\pi y}{\beta}\right) + \eta\right)^2} d\eta$$

$$+\frac{1}{\pi}\exp\left(\frac{\pi x}{\beta}\right)\sin\left(\frac{\pi y}{\beta}\right)\int_0^{\infty} \frac{g(\eta,b)}{\exp\left(\frac{2\pi x}{\beta}\right)\sin^2\left(\frac{\pi y}{\beta}\right) + \left(\exp\left(\frac{\pi x}{\beta}\right)\cos\left(\frac{\pi y}{\beta}\right) - \eta\right)^2} d\eta,$$

i.e.

$$f(\xi_1,\xi_2) = \frac{1}{\pi}\int_0^{\infty} \frac{g(\eta,c)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \eta\right)^2} d\eta$$

$$+\frac{1}{\pi}\int_0^{\infty} \frac{g(\eta,b)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) - \eta\right)^2} d\eta.$$

Finally

$$E^A[T] = f(\xi_1,\xi_2) - 2a^2 \ln\left(\xi_1^2 + \xi_2^2 + 4a^2\right).$$

□

4.3.1. Hitting probabilities

Proposition 4.5. *Let φ_1, $\varphi_2 \in (0, \pi)$, such that $\varphi_1 < \varphi_2$, are both fixed. We consider the sets D_1, D_2 in S^2, such that*

$$D_1 = \{(\theta,\varphi) \mid \theta \in [0,2\pi) \quad \text{and} \quad \varphi \in (\varphi_1, \pi]\}$$

and

$$D_2 = \{(\theta,\varphi) \mid \theta \in [0,2\pi) \quad \text{and} \quad \varphi \in [0, \varphi_2)\}.$$

We have,

$$\partial D_1 = \{(\theta,\varphi) \mid \theta \in [0,2\pi), \quad \text{and} \quad \varphi = \varphi_1\}$$

and

$$\partial D_2 = \{(\theta,\varphi) \mid \theta_1 \in [0,2\pi), \quad \text{and} \quad \varphi = \varphi_2\}.$$

If X_t is the position of the infected (I) at a given time t starting at the point

$$A = (\theta,\varphi) \in D_1 \cap D_2.$$

and in case

$$T_1 = \inf\{t \geq 0 \mid X_t \notin D_1\},$$

$$T_2 = \inf\{t \geq 0 \mid X_t \notin D_2\}$$

and

$$T = \inf\{t \geq 0 \mid X_t \notin D_1 \cap D_2\},$$

then the probabilities

$$Pr^A\{T = T_1\} \quad \text{and} \quad Pr^A\{T = T_2\}$$

are given by

$$Pr^A\{T = T_1\} = \frac{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi}{2}\right)}\right)}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)} \tag{4.44}$$

and

$$Pr^A\{T = T_2\} = \frac{\ln\left(\frac{\tan\left(\frac{\varphi}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}. \tag{4.45}$$

Proof. It is known that (see [21]),

$$u(\theta,\varphi) = Pr^A\{T = T_1\}$$

is the unique solution of the differential equation

$$\frac{1}{2}\Delta_n u = 0, \tag{4.46}$$

with boundary condition

$$u(\theta, \varphi_1) = 1 \quad \text{and} \quad u(\theta_1, \ldots, \theta_{n-1}, \varphi_2) = 0$$

Here Δ_n is the Laplace-Beltrami operator on S^2. By the symmetry of D, it follows that the probability value of $T = T_1$ is independent of θ. From (4.7) the differential Eq. (4.46) takes the form

$$\frac{1}{2a^2}\left[(n-1)\cot(\varphi)\frac{du}{d\varphi} + \frac{d^2u}{d\varphi^2}\right] = 0, \tag{4.47}$$

with boundary condition

$$u(\varphi_1) = 1 \quad \text{and} \quad u(\varphi_2) = 0. \tag{4.48}$$

In we set

$$f(\varphi) = \frac{du}{d\varphi},$$

hence from (4.47)

$$\frac{1}{2a^2}\left[\cot(\varphi) f(\varphi) + \frac{df(\varphi)}{d\varphi}\right] = 0,$$

or

$$\cos(\varphi) f(\varphi) + \sin(\varphi)\frac{df(\varphi)}{d\varphi} = 0.$$

Thus

$$f(\varphi) = \frac{c_1}{\sin\varphi)},$$

i.e.

$$u(\varphi) = \int_{\varphi_2}^{\varphi} \frac{c_1}{\sin x} dx + c_2. \tag{4.49}$$

However,

$$u(\varphi_1) = 1 \quad \text{and} \quad u(\varphi_2) = 0,$$

hence

$$c_1 = -\frac{1}{\int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx}$$

and

$$c_2 = 0.$$

Thus

$$u(\varphi) = \frac{\int_{\varphi}^{\varphi_2} \frac{1}{\sin x} dx}{\int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx}.$$

Consequently,

$$Pr^A\{T = T_1\} = \frac{\int_{\varphi}^{\varphi_2} \frac{1}{\sin x} dx}{\int_{\varphi_1}^{\varphi_2} \frac{1}{\sin x} dx}.$$

or

$$Pr^A\{T = T_1\} = \frac{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi}{2}\right)}\right)}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)} \tag{4.50}$$

Moreover,

$$Pr^A\{T = T_2\} = 1 - Pr^A\{T = T_1\},$$

hence

$$Pr^A\{T = T_2\} = \frac{\ln\left(\frac{\tan\left(\frac{\varphi}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}. \tag{4.51}$$

□

Proposition 4.6. *We consider the 2-dimensional sphere S^2 of radius a. Let two circles pass through the North Pole, such that in stereographic coordinates are represented by the parallel lines $\xi_2 = b$ and $\xi_2 = c$, where $b, c \in \mathbb{R}$, with $b < c$. Next consider the sets D_1, D_2 in S^2, for which the stereographic projections are*

$$D_1 = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 \in (b, +\infty)\}$$

and

$$D_2 = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 \in (-\infty, c)\}.$$

Of course,

$$\partial D_1 = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 = b\}$$

and

$$\partial D_2 = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \text{and} \quad \xi_2 = c\}.$$

Let X_t be the position of the carrier of virus Y at a given time t starting at the point A, and the stereographic projection coordinates of A are

$$(\xi_1, \xi_2) \in D_1 \cap D_2.$$

If

$$T_1 = \inf\{t \geq 0 | X_t \notin D_1\},$$

$$T_2 = \inf\{t \geq 0 | X_t \notin D_2\}$$

and

$$T = \inf\{t \geq 0 | X_t \notin D_1 \cap D_2\},$$

then

$$Pr^A\{T = T_1\} = \frac{c - \xi_2}{c - b} \quad \text{and} \quad Pr^A\{T = T_2\} = \frac{\xi_2 - b}{c - b}. \tag{4.52}$$

Proof. It is known that (see [21]) the function

$$u(\xi_1, \xi_2) = Pr^A\{T = T_1\}$$

is the unique solution of the differential equation

$$\frac{1}{2}\Delta_2 u = 0 \tag{4.53}$$

with boundary condition

$$u(\xi_1, b) = 1 \quad \text{and} \quad u(\xi_1, c) = 0. \tag{4.54}$$

Here, Δ_2, is the Laplace-Beltrami operator on S^2 expressed in the stereographic projection coordinates. Hence from (4.8) the differential Eq. (4.53) takes the form

$$\frac{1}{2}\frac{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}{16a^4} \cdot \left(\frac{\partial^2 u}{\partial \xi_1^2} + \frac{\partial^2 u}{\partial \xi_2^2}\right) = 0,$$

or

$$\frac{\partial^2 u}{\partial \xi_1^2} + \frac{\partial^2 u}{\partial \xi_2^2} = 0. \tag{4.55}$$

From (4.54) and (4.55) we see easily that

$$u(\xi_1, \xi_2) = \frac{c - \xi_2}{c - b}.$$

Therefore,

$$Pr^A\{T = T_1\} = \frac{c - \xi_2}{c - b} \quad \text{and} \quad Pr^A\{T = T_2\} = \frac{\xi_2 - b}{c - b}.$$

□

4.3.2. Moment generating functions

Proposition 4.7. *Let $\varphi_0 \in [0, \pi)$ be fixed. We consider the set D in S^2, such that*

$$D = \{(\theta, \varphi) | \theta \in [0, 2\pi), \quad \text{and} \quad \varphi \in [0, \varphi_0)\}.$$

Then,

$$\partial D = \{(\theta, \varphi) | \theta \in [0, 2\pi) \quad \text{and} \quad \varphi = \varphi_0\}.$$

If X_t is the infectious position at a given time t starting at the point

$$A = (\theta, \varphi) \in D,$$

and

$$T = \inf\{t \geq 0 | X_t \notin D\},$$

then the expectation of $\exp(-\lambda T)$ is given by

$$E^A[\exp(-\lambda T)] = \frac{P_\nu(\cos\varphi)}{P_\nu(\cos\varphi_0)}, \tag{4.56}$$

where ν is such that $\nu(\nu+1) = -2a^2\lambda$ and $P_\nu(\cdot)$ is the Legendre function

$$P_\nu(z) = P_{-\nu-1}(z) = \frac{1}{\pi}\int_0^\pi (z + \sqrt{z^2-1}\cos\varphi)^\nu d\varphi,$$

where the multiple-valued function $(z + \sqrt{z^2-1}\cos\varphi)^\nu$ is to be determined in such a way that for $\varphi = \frac{\pi}{2}$ it is equal to (the principal value of) z^ν (which is, in particular, real for positive z and real ν).

Proof. If $\lambda > -\frac{\lambda_1}{2}$, where λ_1 is the first Dirichlet eigenvalue of $D \subset S^2$, then

$$E^A[\exp(-\lambda T)]$$

it satisfies the differential equation

$$\frac{1}{2}\Delta_2 u = \lambda u(\varphi) \tag{4.57}$$

with boundary condition

$$u(\varphi_0) = 1. \tag{4.58}$$

Here Δ_2 is the Laplace-Beltrami operator on S^2. By the symmetry of D, it follows that the expectation of $\exp[-\lambda T]$ is independent of θ. Hence u is independent of θ. From (4.4) the differential equation (4.57) takes the form

$$\frac{1}{2a^2\sin\varphi}\left(\frac{du}{d\varphi}\cos\varphi + \frac{d^2u}{d\varphi^2}\sin\varphi\right) = \lambda u(\varphi),$$

i.e.

$$\frac{d}{d\varphi}\left(\frac{du}{d\varphi}\sin\varphi\right) - \left(2\lambda a^2\sin\varphi\right)u(\varphi) = 0. \tag{4.59}$$

If we set

$$z = \cos\varphi,$$

then

$$\frac{du}{d\varphi} = -\sin\varphi\frac{du}{dz}$$

and (4.59) transforms to

$$(1-z^2)\frac{d^2u}{dz^2} - 2z\frac{du}{dz} - 2\lambda a^2 u = 0,$$

or

$$(1-z^2)\frac{d^2u}{dz^2} - 2z\frac{du}{dz} + \nu(\nu+1)u = 0.$$

This is Legendre's differential equation. However, $u(\varphi)$ is bounded for all $\varphi \in [0, \pi]$ and $u(\varphi_0) = 1$. Therefore (see [106]), the solution of (4.59) is

$$u(\varphi) = \frac{P_\nu(\cos\varphi)}{P_\nu(\cos\varphi_0)},$$

i.e.

$$E^A[\exp(-\lambda T) = \frac{P_\nu(\cos\varphi)}{P_\nu(\cos\varphi_0)},$$

where ν is such that $\nu(\nu+1) = -2a^2\lambda$. □

4.4. Reflection principle

Theorem 4.2. *Let X_t be the position of the infectious carrier of virus Y at a given time t starting at the point*

$$A = (\theta, \varphi) \in D,$$

where

$$D = \left\{(\theta, \varphi) \in S^2 \,\middle|\, \theta \in [\right.$$

If

$$T = \inf\{t \geq 0 | X_t \notin D\},$$

then

$$Pr^A\{T < t\} = 2Pr^A\{X_t \notin D\}. \tag{4.60}$$

Proof.

$$Pr^A\{T < t\} = Pr^A\{T < t, X_t \notin D\} + Pr^A\{T < t, X_t \in D\}. \tag{4.61}$$

However, if $X_t \notin D$ then of course $T < t$.

Thus,

$$Pr^A\{T < t, X_t \notin D\} = Pr^A\{X_t \notin D\}. \tag{4.62}$$

On the other hand, if we set

$$\tilde{X}_t = \begin{cases} X_t, & \text{if} \quad t \leq T \\ \hat{X}_t, & \text{if} \quad t > T \end{cases},$$

then by the strong Markov property of X_t

$$Pr^A\{T < t, X_t \in D\} = Pr^A\left\{T < t, \tilde{X}_t \in D\right\},$$

but if $\tilde{X}_t \in D$ then $X_t \notin D$. Hence,

$$Pr^A\{T < t, X_t \in D\} = Pr^A\{T < t, X_t \notin D\},$$

or

$$Pr^A\{T < t, X_t \in D\} = Pr^A\{X_t \notin D\}. \tag{4.63}$$

Therefore from (4.61)–(4.63) we obtain that

$$Pr^A\{T < t\} = 2Pr^A\{X_t \notin D\}.$$

□

4.4.1. Applications of the reflection principle

The reflection principle can help to calculate the distribution functions of certain exit times.

Let X_t be the position of the infected individual at a given time t starting at the point $N(0, 0)$ in spherical coordinates. If

$$D = \left\{(\theta, \varphi) \in S^2 \,\middle|\, \theta \in [0, 2\pi), \varphi \in \left(\frac{\pi}{2}, \pi\right]\right\}$$

then

$$Pr^N\{X_t \notin D\} = \int_0^{\frac{\pi}{2}}\int_0^{2\pi} p(t, \varphi)a^2\sin(\varphi)d\theta d\varphi,$$

i.e.

$$Pr^N\{X_t \notin D\} = 2\pi a^2\int_0^{\frac{\pi}{2}} p(t, \varphi)\sin(\varphi)d\varphi,$$

where $p(t, \varphi)$ is the transition density function of the Brownian motion on S^2 of radius a. Hence from (4.26)

$$Pr^N\{X_t \notin D\} = 2\pi a^2 \int_0^{\frac{\pi}{2}} \frac{1}{4\pi a^2} \sin\varphi \sum_{n\in\mathbb{N}} (2n+1) \exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right) P_n^0(\cos\varphi) d\varphi,$$

or

$$Pr^N\{X_t \notin D\} = \frac{1}{2} + \frac{1}{2}\sum_{n\in\mathbb{N}^*} (2n+1) \exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right) \int_0^{\frac{\pi}{2}} P_n^0(\cos\varphi)\sin(\varphi) d\varphi. \tag{4.64}$$

However for every $n \in \mathbb{N}^*$

$$I_n = \int_0^{\frac{\pi}{2}} P_n^0(\cos\varphi)\sin(\varphi)d\varphi = \int_0^1 P_n^0(x)dx.$$

It is known that (see [106])

$$P_n^0(x) = \frac{1}{2n+1}\frac{d}{dx}\left[P_{n+1}^0(x) - P_{n-1}^0(x)\right].$$

However, $P_n^0(1) = 1$ for every $n \in \mathbb{R}$. Thus

$$I_n = \frac{1}{2n+1}\left(P_{n+1}^0(1) - P_{n-1}^0(1) - P_{n+1}^0(0) + P_{n-1}^0(0)\right),$$

or

$$I_n = \frac{1}{2n+1}\left(P_{n-1}^0(0) - P_{n+1}^0(0)\right).$$

It is also known that for every $n \in \mathbb{N}^*$

$$P_{2n}^0(0) = (-1)^n \frac{(2n)!}{2^{2n}(n!)^2} \quad \text{and} \quad P_{2n+1}^0(0) = 0.$$

Thus, if n is even then $I_n = 0$. If n is odd, i.e. $n = 2k+1$, then

$$I_n = \frac{1}{4k+3}\left(P_{2k}^0(0) - P_{2(k+1)}^0(0)\right),$$

i.e.

$$I_n = \frac{(-1)^n(2k)!}{(k+1)(k!)^2 2^{2k+1}}. \tag{4.65}$$

From (4.64) and (4.65) we get that

$$Pr^N\{X_t \notin D\} = \frac{1}{2} + \frac{1}{2}\sum_{n\in\mathbb{N}} (-1)^n \exp\left(-\frac{(2n+1)(2n+2)\sqrt{t}}{a}\right) \times \frac{(2n)!(4n+3)}{2^{2n+1}(n!)^2(n+1)}. \tag{4.66}$$

Furthermore, if $S(0, \pi)$ namely the South Pole of S^2, then

$Pr^S\{X_t \notin D\} = Pr^N\{\hat{X}_t \notin D\} = Pr^N\{X_t \in D\} = 1 - Pr^N\{X_t \notin D\}$.

Therefore

$$Pr^S\{X_t \notin D\} = \frac{1}{2} - \frac{1}{2}\sum_{n\in\mathbb{N}} (-1)^n \exp\left(-\frac{(2n+1)(2n+2)\sqrt{t}}{a}\right) \times \frac{(2n)!(4n+3)}{2^{2n+1}(n!)^2(n+1)}. \tag{4.67}$$

By using **Theorem 3.2**, if $T = \inf\{t > 0 \,|\, X_t \notin D\}$, then

$$Pr^S\{T < t\} = 1 - \sum_{n\in\mathbb{N}} (-1)^n \exp\left(-\frac{(2n+1)(2n+2)\sqrt{t}}{a}\right) \times \frac{(2n)!(2n+3)}{2^{2n+1}(n!)^2(n+1)}. \tag{4.68}$$

4.5. Local time estimation

Definition 4.11. Let $\varphi_1 \in [0, \pi]$. We set

$$D_1 = \{(\theta_1, \ldots, \theta_{n-1}, \varphi) \,\|\, \theta_1 \in [0, 2\pi), \theta_i \in [0, \pi] \quad \text{for} \quad i = 2, \ldots, n-1 \quad \text{and} \quad \varphi \in (0, \varphi_1]\},$$

is a subset of S^2. The reflected Brownian motion in D_1 is the diffusion Y_t whose generator is Δ_n in D_1 with Neuman boundary condition at ∂D_1.

Roughly speaking Y_t behaves like X_t inside D_1 but when it reaches the boundary, it is reflected back in D_1.

Definition 4.12. Let a fixed open set $D \subset S^n$ with C^3-boundary ∂D. If Y_t is the reflected Brownian motion in D, and D_δ the domain

$$D_\delta = \{x \in D : d(x, \partial D) < \delta\},$$

we define the boundary local time L_t of Y_t, as

$$L_t := \lim_{\delta\to 0^+} \frac{1}{2\delta}\int_0^t 1_{D_\delta}(Y_s)ds.$$

It can be shown that the limit exist in the L_2 sense.

4.5.1. Boundary local time until first hitting

Proposition 4.8. *Let $\varphi_0, \varphi_1 \in (0, \pi)$, such that $\varphi_0 < \varphi_1$, both fixed. We consider the sets D, Γ_0 in S^2, such that*

$$D = \{(\theta, \varphi) | \theta \in [0, 2\pi) \quad \text{and} \quad \varphi \in (\varphi_0, \varphi_1)\}.$$

and

$$\Gamma_0 = \{(\theta, \varphi_0) | \theta \in [0, 2\pi)\}.$$

Let Y_t be the reflected Brownian motion in Γ_0 starting at the point

$$A = (\theta, \varphi) \in D$$

if

$$T = \inf\{t \geq 0 | X_t \in \Gamma_0\}$$

and L_t is the boundary local time of Y_t, then,

$$E^A[\exp(\lambda L_T)] = \frac{\frac{1}{\sin\varphi_1} - \lambda\ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi}{2}\right)}\right)}{\frac{1}{\sin\varphi_1} - \lambda\ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)}, \quad \text{if} \quad \lambda < \frac{1}{\sin(\varphi_1)\ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)} \tag{4.69}$$

and

$$E^A[\exp(\lambda L_T)] = +\infty, \quad \text{if} \quad \lambda \geq \frac{1}{\sin(\varphi_1)\ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)}. \tag{4.70}$$

Proof. It is known that the function

$$z(\theta, \varphi) = E^A[\exp(\lambda L_T)]$$

satisfies the differential equation

$$\Delta_2 z = 0$$

with boundary condition

$$z(\theta, \varphi_0) = 1$$

and

$$-\frac{\partial z}{\partial\varphi}(\theta, \varphi_1) + \lambda z(\theta, \varphi_1) = 0.$$

as long as the function z is positive (see [107]). Here Δ_2 is the Laplace-Beltrami operator on S^2. By the symmetry of D it follows that $E^A[\exp(\lambda L_T)]$ is independent of θ. From (4.2) the differential equation takes the form

$$\cot(\varphi)\frac{dz}{d\varphi} + \frac{d^2z}{d^2\varphi} = 0. \tag{4.71}$$

We have shown that the solution of (4.71) is

$$z(\varphi) = c_1 \int_{\varphi_0}^{\varphi} \frac{1}{\sin x} dx + c_2, \quad c_1, c_2 \in \mathbb{R}.$$

However,

$$z(\theta, \varphi_0) = 1$$

and

$$-\frac{\partial z}{\partial \varphi}(\theta, \varphi_1) + \lambda z(\theta, \varphi_1) = 0.$$

Hence

$$c_1 = \frac{\lambda}{(\sin\varphi_1)^{-1} - \lambda \int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx}$$

and

$$c_2 = 1.$$

Thus

$$z(\varphi) = \frac{(\sin\varphi_1)^{-1} - \lambda \int_{\varphi}^{\varphi_1} (\sin x)^{-1} dx}{(\sin\varphi_1)^{-1} - \lambda \int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx}.$$

However,

$$z(\varphi) > 0 \quad \text{if and only if} \quad \lambda < \frac{(\sin\varphi_1)^{-1}}{\int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx}.$$

Therefore,

$$E^A[\exp(\lambda L_T)] = \frac{(\sin\varphi_1)^{-1} - \lambda \int_{\varphi}^{\varphi_1} (\sin x)^{-1} dx}{(\sin\varphi_1)^{-1} - \lambda \int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx},$$
$$\text{if} \quad \lambda < \frac{(\sin\varphi_1)^{-1}}{\int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx}$$

and

$$E^A[\exp(\lambda L_T)] = +\infty, \quad \text{if} \quad \lambda \geq \frac{(\sin\varphi_1)^{-1}}{\int_{\varphi_0}^{\varphi_1} (\sin x)^{-1} dx}$$

i.e

$$E^A[\exp(\lambda L_T)] = \frac{\frac{1}{\sin\varphi_1} - \lambda \ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi}{2}\right)}\right)}{\frac{1}{\sin\varphi_1} - \lambda \ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)},$$
$$\text{if} \quad \lambda < \frac{1}{\sin(\varphi_1) \ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)}$$

and

$$E^A[\exp(\lambda L_T)] = +\infty, \quad \text{if} \quad \lambda \geq \frac{1}{\sin(\varphi_1) \ln\left(\frac{\tan\left(\frac{\varphi_1}{2}\right)}{\tan\left(\frac{\varphi_0}{2}\right)}\right)}.$$

□

5. Discussion and conclusions

A worldwide multilevel interplay among a plethora of factors ranging from micro-pathogens and individual interactions to macro-scale environmental, socio-economic and demographic conditions, necessitate the development of highly sophisticated mathematical models for robust representation of contagious dynamics of infectious diseases that would lead to the establishment of effective control strategies and prevention policies.

Ethical and practical reasons defer from conducting enormous experiments in public health systems, hence mathematical models appear to be an efficient way to explore contagion dynamics. A key aspect of epidemiological models is their link to real data, which is of particular utility toward the design of vaccination policies. Two major vaccination strategies exist currently, i.e., the mass vaccination, which is most applied, and the recently developed pulse vaccination which is used in an increasing number of countries. However, most vaccination strategies are imperfect in the sense that they decrease the number of cases, without however eradicating the disease.

Public-health organizations in the world use the epidemiological models that fall in the three categories already presented in this work, to evaluate disease outbreak policies for epidemics. As we pointed out, many shortcomings exist for those models. All the models already used in the literature assume that the host population has constant size. However, this excludes diseases in exponentially growing populations as in most developing countries, or disease-induced mortality as childhood diseases in developing countries e.g., malaria. Modeling infectious dynamics in non-stationary host populations requires explicit modeling of the host population as well as of the disease per se. Models sometimes can be highly complicated in order to improve best fit to real data. Nonetheless, very complex models do not always perform optimally in real-world applications or in simulations. Real-world models allow for swift decision making, and suitable quantification of the spatiotemporal dynamics of an outbreak. Multidisciplinary research efforts are speeding up, integrating the advances in epidemiology, molecular biology, computational science and applied mathematics. Mathematical modeling allows better understanding of the transmission process of infectious diseases in space and time, by setting forth rigorously the proper assumptions, the variables, the equations and their parameters.

Due to the complexity of the underlying complex interactions, either deterministic or stochastic epidemiological models are built upon incomplete information about e.g., the basic reproduction number, threshold effects, intensity of spread, precise data of infected versus susceptible individuals, and other inaccuracies regarding the entire infectious network. Simulations or brute-force computational techniques have been implemented in that direction to provide approximate solutions with encouraging results. Nevertheless, some of the underlying generating processes of the outbreaks, such as the virus pathogenicity or variant social network topologies, ethnological characteristics and other quantities, may influence the spread of an outbreak. Simulations often prove to be inefficient for the systematic analysis of an emergent epidemic. New rigorous mathematical modeling methodologies, such as the one presented in this work for the first time, can be used to address inherent incomplete data structure and hidden nonlinear complex dynamics, with an aim to enhance forecastability in combating epidemic outbreaks.

In the present study we introduced a novel approach for surveillance and modeling of infectious disease dynamics, called SBDiEM. We explicitly described the mathematical framework underpinning the implementation and conceptualization of our new-age epidemiological model. Our goal is to contribute to the arsenal of models already developed so far. It can be of particular interest, in light of a recent intensive worldwide effort to speed up the establishment of a global surveillance network for combating pandemics of emergent and re-emergent infectious diseases. Toward this aim, mathematical modeling will play a major role in assessing, controlling and forecasting potential outbreaks. We have to better understand and model the impact of numerous variables on contagious dynamics, ranging from the microscopic host–pathogen level, to individual and population interactions, as well as macroscopic environmental, social, economic and demographic factors all over the world.

As a path for future research, we intend to conduct simulations, and empirical analyses based on real-time spatiotemporal datasets, in case of past outbreaks of infectious diseases as well as for COVID-19. Furthermore, we plan to convey an extensive comparative evaluation investigation of SBDiEM vis-à-vis the three major categories set forth by the taxonomy of Siettos and Russo [58], and more specifically versus (1) statistical methods for epidemic surveillance, (2) state-space models of epidemic spread and (3) machine learning methods. In this way, the forecasting and nowcasting capabilities of the new model will be thoroughly explored. We also intend to investigate embedding the proposed analytical model into integrated artificial intelligence systems in the near future.

Our novel methodology apart from offering a much better understanding of the complex and heterogeneous infectious disease dynamics could enhance predictability of epidemic outbreaks as well as have potentially important implications for national health systems, stakeholders and international policy makers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] OWorld Health Organization. Coronavirus. world health organization, cited january 19. 2020. Available:https://www.who.int/health-topics/coronavirus.
[2] Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020. doi:10.1038/s41586-020-2012-7.
[3] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020. doi:10.1056/NEJMoa2001316.
[4] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. China Lancet 2020. doi:10.1016/S0140-6736(20)30183-5.
[5] Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020. doi:10.1016/S0140-6736(20)30154-9.
[6] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China 2019. N Engl J Med 2020. doi:10.1056/NEJMoa2001017.
[7] Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel. J Travel Med 2020. doi:10.1093/jtm/taaa008.
[8] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020. doi:10.1016/S0140-6736(20)30260-9.
[9] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020. doi:10.1016/j.ijid.2020.01.050.
[10] Zhao S, Musa SS, Lin Q, Ran J, Yang G, Wang W, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of january 2020: a data-driven modelling analysis of the early outbreak. J Clin Med 2020. doi:10.3390/jcm9020388.
[11] Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household transmission of 2009 pandemic influenza a (H1N1) virus in the United States. N Engl J Med 2009;361:2619–27.
[12] Cauchemez S, Fraser C, Van KMD, Donnelly CA, Riley S, Rambaut A, et al. Middle east respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. Lancet Infect Dis 2014;14:50–6.
[13] Cauchemez S, Nouvellet P, Cori A, Jombart T, Garske T, Clapham H, et al. Unraveling the drivers of MERS-CoV transmission. Proc Natl Acad Sci USA 2016;113:9081–908.
[14] Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 2003;300:1961–6.
[15] WHO. Ebola virus disease in West Africa — the first 9 months of the epidemic and forward projections. N Engl J Med 2014;371:1481–95.
[16] Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The brazilian Zika virus strain causes birth defects in experimental models. Nature 2016;534:267–71.
[17] Anderson RM, May RM. Infectious diseases of humans: dynamics and control, 28. Oxford: Oxford University Press; 1992.
[18] Noh JD, Rieger H. Random Walks on Complex Networks. Phys Rev Lett 2004;92:118701.
[19] Draief M, Massouli L. Epidemics and rumours in complex networks. Cambridge University Press; 2010.
[20] Bonaventura M, Nicosia V, Latora V. Characteristic times of biased random walks on complex networks. Phys Rev E 2014;89:012803.
[21] Klebaner FC. Introduction to Stochastic Calculus with Applications. Melbourne: Imperial College Press; 2004. Continuous Trading,' Stoch. Process Appl., 11, 215–260
[22] Bernoulli D. Essai d'une nouvelle analyse de la mortalité caus ée par la petite vérole. Mém Math Phys Acad R Sci Paris 1766;1:1–45.
[23] Dietz K, Heesterbeek JAP. Daniel bernoulli's epidemiological model revisited. Math Biosci 2002;180:1–21. doi:10.1016/S0025-5564(02)00122-0. PMID:12387913
[24] Lambert JH. Die toedlichkeit der kinderblattern. beytrage zum gebrauche der mathematik und deren anwendung. Buchhand Realschule 1772;3:568.
[25] Laplace PS. Theorie analytique des probabilites. Paris: Courcier; 1812.
[26] Ross R. The prevention of malaria. London: John Murray; 1911. p. 651–86.
[27] Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE. Ross-macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. PLoS Pathog 2012;8:e1002588. doi:10.1371/journal.ppat.1002588. PMID:22496640
[28] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond 1927;A115:700–21.
[29] Kermack WO, McKendrick AG. Contribution to the mathematical theory of epidemics. Proc R Soc Lond A Contain Pap Math Phys Character 1927;115:700–21. doi:10.1098/rspa.1927.0118.
[30] Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics, part II, 138; 1932. p. 55–83. doi:10.1098/rspa.1932.0171.
[31] Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics, part III, 141; 1933. p. 94–112. doi:10.1098/rspa.1933.0106.
[32] Guldberg CM, Waage P. Studies concerning affinity. C. M. Forhandlinger: Videnskabs-Selskabet Christiana; 1864. p. 111.
[33] Blower S, Bernoulli D. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it 1766. Rev Med Virol 2004;14:275–88. doi:10.1002/rmv.443. PMID:15334536
[34] Bartlett MS. The critical community size for measles in the United States. J R Stat Soc 1960;123:37–44.
[35] Bailey NJT. The mathematical theory of infectious diseases and its application. Griffin, London; 1957.
[36] Anderson RM, May RM. Infectious diseases of humans. Dynamics and control. Oxford: Oxford University Press; 1991.
[37] Enko PD. On the course of epidemics of some infectious diseases. In: Vrach. St. Petersburg; 1889. p. 1008–10. 1039–1042, 1061–1063
[38] Enko PD. On the course of epidemics of some infectious diseases. Int J Epidemiol 1989;18:749–55. doi:10.1093/ije/18.4.749. PMID:2695472
[39] Dietz K. The first epidemic model: a historical note on P.D. Enko Aust J Stat 1988;30A:56–65. doi:10.1111/j.1467-842X.1988.tb00464.x.
[40] Frost WH. Some conceptions of epidemics in general by wade hampton frost. Am J Epidemiol 1976;103:141–51. PMID:766618
[41] Xing Y, Song L, Sun GQ, Jin Z, Zhang J. Assessing reappearance factors OF H7N9 avian influenza in China. Appl Math Comput 2017;309:192–204. doi:10.1016/j.amc.2017.04.007.
[42] Lee EK, Liu Y, Pietz FH. A compartmental model for zika virus with dynamic human and vector populations. AMIA, 2016; 2016. p. 743–52. PMID: 28269870
[43] Sun GQ, Xie JH, Huang SH, Jin Z, Li MT, Liu L. Transmission dynamics of cholera: mathematical modeling and control strategies. Commun Nonlinear Sci Numer Simul 2017;45:235–44. doi:10.1016/j.cnsns.2016.10.007.
[44] Nishiura H, Kinoshita R, Mizumoto K, Yasuda Y, Nah K. Transmission potential of Zika virus infection in the south pacific. Int J Infect Dis 2016;45:95–7. doi:10.1016/j.ijid.2016.02.017.
[45] Khan MA, Parvez M, Islam S, Khan I, Shafie S, Gul T. Mathematical analysis of typhoid model with saturated incidence rate. Adv Stud Biol 2015;7:65–78. doi:10.12988/asb.2015.41059.
[46] Gui HL, Yong XZ. Dynamic behaviors of a modified SIR model in epidemic diseases using nonlinear incidence and recovery rates. PLoS One 2017;12(4):e017578.
[47] Li MG, Sun GQ, Wu YF, Zhang J, Jin Z. Transmission dynamics of a multi-group brucellosis model with mixed cross infection in public farm. Appl Math Comput 2014;237:582–94. doi:10.1016/j.amc.2014.03.094.
[48] Yua P, Linb W. Complex dynamics in biological systems arising from multiple limit cycle bifurcation. J Biol Dyn 2016;10(1):263–85. doi:10.1080/17513758.2016.1166270.
[49] Shi X, Li G, Zhou X, Song X. Analysis of a differential equation model of HIV infection of CD4 + T-cells with saturated reverse function. Turk J Math 2011;35:649–66.
[50] Bonyah E, Dontwi I, Nyabadza T. A theoretical model for the transmission dynamics of the Buruli ulcer with saturated treatment. Comput Math Meth Med 2014;2014:1–13. doi:10.1155/2014/576039.
[51] Zhang J, Jia J, Song X. Analysis of an SEIR epidemic model with saturated incidence and saturated treatment function. Sci World J 2014;2014:1–11.
[52] Daley DJ, Gani J. Epidemic modeling: an introduction. NY: Cambridge University Press; 2005.
[53] Brauer F, Castillo-Chávez C. Mathematical models in population biology and epidemiology. NY: Springer; 2001.
[54] Keeling MJ, Grenfell BT. Understanding the persistence of measles: reconciling theory, simulation and observation. Proc R Soc Lond 2002;B269:335–43.
[55] Lloyd AL. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. Proc R Soc Lond 2001;B268:985–93.

[56] Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. Public Libr Sci – Med 2005;2(7):e174.
[57] Grossman Z. Oscillatory phenomena in a model of infectious diseases. Theor Popul Biol 1980;18:204–43.
[58] Siettos C, Russo L. Mathematical modelling of infectious disease dynamics. Virulence 2013;4(4):295–306.
[59] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963;78:494–506. doi:10.2307/4591848. PMID:19316455
[60] Stroup DF, Thacker SB, Herndon JL. Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age 1962-1983. Stat Med 1988;7:1045–59. doi:10.1002/sim.4780071006. PMID:3264612
[61] Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ. When is the epidemic warning cut-off point exceeded? Eur J Epidemiol 1994;10:475–6. doi:10.1007/BF01719680. PMID:7843360
[62] Greenland S. Regression methods for epidemiologic analysis. In: Handbook of epidemiology. Part II; 2005. p. 625–91. doi:10.1007/978-3-540-26577-1-17.
[63] Pelat C, Boëlle PY, Cowling BJ, Carrat F, Flahault A, Ansart S, et al. Online detection and quantification of epidemics. BMC Med Inform Decis Mak 2007;7:29. doi:10.1186/1472-6947-7-29. PMID:17937786
[64] Dafni UG, Tsiodras S, Panagiotakos D, Gkolfinopoulou K, Kouvatseas G, Tsourti Z, et al. Algorithm for statistical detection of peaks-syndromic surveillance system for the athens 2004 olympic games. MMWR Morb Mortal Wkly Rep 2004;24(Suppl):86–94.
[65] Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962-1979. i. time series forecasts of expected pneumonia and influenza deaths. Am J Epidemiol 1981;113:215–26. PMID:6258426
[66] Abeku TA, de Vlas SJ, Borsboom G, Teklehaimanot A, Kebede A, Olana D, et al. Forecasting malaria incidence from historical morbidity patterns in epidemicprone areas of ethiopia: a simple seasonal adjustment method performs best. Trop Med Int Health 2002;7:851–7. doi:10.1046/j.1365-3156.2002.00924.x. PMID:12358620
[67] Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. PLoS One 2010;5:e9450. doi:10.1371/journal.pone.0009450. PMID:20209164
[68] Nunes B, Viboud C, Machado A, Ringholz C, Rebelode-Andrade H, Nogueira P, et al. Excess mortality associated with influenza epidemics in portugal, 1980 to 2004. PLoS One 2011;6:e20661. doi:10.1371/journal.pone.0020661. PMID:21713040
[69] Page ES. Continuous inspection scheme. Biometrika 1954;41:100–15.
[70] Raubertas RF. An analysis of disease surveillance data that uses the geographic locations of the reporting units. Stat Med 1989;8:267–71. doi:10.1002/sim.4780080306. Discussion 279-81; PMID:2711060
[71] Cowling BJ, Wong IOL, Ho LM, Riley S, Leung GM. Methods for monitoring influenza surveillance data. Int J Epidemiol 2006;35:1314–21. doi:10.1093/ije/dyl162. PMID:16926216
[72] Höhle M, Paul M. Count data regression charts for the monitoring of surveillance time series. Comput Stat Data Anal 2008;52:4357–68. doi:10.1016/j.csda.2008.02.015.
[73] Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ. Applying cusum-based methods for the detection of outbreaks of ross river virus disease in western australia. BMC Med Inform Decis Mak 2008;8:37. doi:10.1186/1472-6947-8-37. PMID:18700044
[74] Spanos A, Theocharis G, Karageorgopoulos DE, Peppas G, Fouskakis D, Falagas ME. Surveillance of community outbreaks of respiratory tract infections based on house-call visits in the metropolitan area of athens, greece. PLoS One 2012;7:e40310. doi:10.1371/journal.pone.0040310. PMID:22905091
[75] Roberts SW. Control chart tests based on geometric moving averages. Technomearics 1959;1:239–50. doi:10.1080/00401706.1959.10489860.
[76] Elbert Y, Burkom HS. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. Stat Med 2009;28:3226–48. doi:10.1002/sim.3708. PMID:19725023
[77] Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden markov models. Stat Med 1999;18:3463–78. PMID:10611619
[78] Rath TM, Carreras M, Sebastiani P. Automated detection of influenza epidemics with hidden Markov models. Berlin: Springer; 2003.
[79] Anderson RM, May RM. Population biology of infectious diseases: Part i. Nature 1979;280:361–7. doi:10.1038/280361a0. PMID:460412
[80] Brauder F, Castillo-Chavez C. Mathematical models in population biology and epidemiology. New York: Springer-Verlag; 2001.
[81] Murray JD. Mathematical biology II. New York: Springer-Verlag; 2002.
[82] Feng Z., Dieckmann U., Levin S., editors. Disease evolution: models, concepts and data analysis. AMS; 2006.
[83] Severo N. Generalizations of some stochastic epidemic models. Math Biosci 1969;4:395–402. doi:10.1016/0025-5564(69)90019-4.
[84] Prosperi MCF, D'Autilia R, Incardona F, De Luca A, Zazzi M, Ulivi G. Stochastic modelling of genotypic drug-resistance for human immunodeficiency virus towards long-term combination therapy optimization. Bioinformatics 2009;25:1040–7. doi:10.1093/bioinformatics/btn568. PMID:18977781
[85] Vardavas R, Blower S. The emergence of HIV transmitted resistance in botswana: "when will the WHO detection threshold be exceeded?". PLoS One 2007;2:e152. doi:10.1371/journal.pone.0000152. PMID:17225857
[86] Newman MEJ. The structure and function of networks. SIAM Rev 2003;45:167–256. doi:10.1137/S003614450342480.
[87] Barabasi AL. Statistical mechanics of complex networks. Rev Mod Phys 2002;74:47–97. doi:10.1103/RevModPhys.74.47.
[88] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. Phys Rep 2006;424:175–308. doi:10.1016/j.physrep.2005.10.009.
[89] Keeling MJ, Danon L, Vernon MC, House TA. Individual identity and movement networks for disease metapopulations. Proc Natl Acad Sci USA 2010;107:8866–70. doi:10.1073/pnas.1000416107. PMID:20421468
[90] Keeling MJ, Eames KT. Networks and epidemic models. J R Soc Interface 2005;2:295–307. doi:10.1098/rsif.2005.0051. PMID:16849187
[91] Read JM, Eames KTF, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. J R Soc Interface 2008;5:1001–7. doi:10.1098/rsif.2008.0013. PMID:18319209
[92] Eubank S, Guclu H, Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. Nature 2004;429:180–4. doi:10.1038/nature02541. PMID:15141212
[93] Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in southeast asia. Nature 2005;437:209–14. doi:10.1038/nature04017. PMID:16079797
[94] Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, et al. Individual-based computational modeling of smallpox epidemic control strategies. Acad Emerg Med 2006;13:1142–9. doi:10.1111/j.1553-2712.2006.tb01638.x. PMID:17085740
[95] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA 2009;106:21484–9. doi:10.1073/pnas.0906910106. PMID:20018697
[96] Bai Y, Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy. Chaos Solitons Fractals 2005;26:559–69. doi:10.1016/j.chaos.2005.01.064.
[97] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457:1012–14. doi:10.1038/nature07634. PMID:19020500
[98] Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PLoS One 2009;4:e4378. doi:10.1371/journal.pone.0004378. PMID:19197389
[99] Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 2011;5:e1206. doi:10.1371/journal.pntd.0001206. PMID:21647308
[100] Do Carmo MP. Differential geometry of curves and surfaces. Upper Saddle River, New Jersey: Prentice-Hall, Inc.; 1976.
[101] Dodziuk J. Maximum principle for parabolic inequalities and the heat flow on open manifolds. Indiana Oniv Math J 1983;32(5):115–42.
[102] John F. Partial differential equations. USA: Springer; 1982.
[103] Chung MK. Heat kernel smoothing on unit sphere. In: Proceedings of IEEE International symposium on biomedical imaging (ISBI); 2006. p. 992–5.
[104] Oksendal B. Stochastic differential equations. Springer-Verlag; 1995.
[105] Dynkin EB. Markov processes, 2. Berlin: Springer; 1965.
[106] Spiegel MR. Complex variables with an introduction to conformal mapping and its applications. Schaum's Outline Series; 1999.
[107] Papanicolaou VG. The probabilistic solution of the third boundary value problem for second order elliptic equations. Probab Theory Relat Fields 1990;87:27–77.

Functional limit theorems for the Pólya and q-Pólya urns

Dimitris Cheliotis* Dimitra Kouloumpou †

June 3, 2019

Abstract

For the plain Pólya urn with two colors, black and white, we prove a functional central limit theorem for the number of white balls assuming that the initial number of black balls is large. Depending on the initial number of white balls, the limit is either a pure birth process or a diffusion. We also prove analogous results for the q-Pólya urn, which is an urn where, when picking a ball, the balls of one color have priority over those of the other.

1 Introduction and results

1.1 The models

The Pólya urn. This is the model where in an urn that has initially r white and s black balls we draw, successively, uniformly, and at random, a ball from it and then we return the ball back together with k balls of the same color as the one drawn. The number $k \in \mathbb{N}^+$ is fixed. Call A_n and B_n the number of white and black balls respectively after n drawings. The most notable result regarding its asymptotic behavior is that the proportion of white balls in the urn after n drawings, $A_n/(A_n + B_n)$, converges almost surely as $n \to \infty$ to a random variable with distribution $\text{Beta}(r/k, s/k)$. Our aim in this work is to examine whether the entire path $(A_n)_{n\geq 0}$ after appropriate natural transformations converges to a stochastic process.

Standard references for the theory and the applications of Pólya urn and related models are [11] and [14].

The q-Pólya urn. This is a q-analog of the Pólya urn (see [7], [12] for more on q-analogs) introduced in [13] and studied further in [3] (see also [4]). A q-analog of a mathematical object A is another object $A(q)$ so that when $q \to 1$, $A(q)$ "tends" to A. Take $q \in (0, \infty)\backslash\{1\}$. The q-analog of any $x \in \mathbb{C}$ is defined as

$$[x]_q := \frac{q^x - 1}{q - 1}. \tag{1.1}$$

Note that $\lim_{q\to 1}[x]_q = x$. Now consider an urn that has initially r white and s black balls, where $r, s \in \mathbb{N}, r + s > 0$. We perform a sequence of additions of balls in the urn according to the following rule. If at a given time the urn contains w white and b black balls ($w, b \in \mathbb{N}, w + b > 0$), then we add k white balls with probability

$$\mathbf{P}_q(\text{white}) = \frac{[w]_q}{[w + b]_q}. \tag{1.2}$$

*Department of Mathematics, Panepistimiopolis, Athens 15784, Greece. Email: dcheliotis@math.uoa.gr.

†Techological Educational Institute of Pelloponnese, Antikalamos 24100, Kalamata, Greece. Email: dimkouloubou@teipel.gr.

2010 Mathematics Subject Classification: 60F17; 60K99; 60C05.

Keywords: Pólya urn, q-Pólya urn, q-calculus, functional limit theorems.

Otherwise, we add k black balls, and this has probability

$$\mathbf{P}_q(\text{black}) = 1 - \mathbf{P}_q(\text{white}) = q^w \frac{[b]_q}{[w+b]_q}. \tag{1.3}$$

To understand how the q-Pólya urn works, it helps to realize the probabilities $\mathbf{P}_q(\text{white}), \mathbf{P}_q(\text{black})$ through a natural experiment.

If $q \in (0,1)$, then we put the balls in a line with the w white coming first and the b black following. To pick a ball, we go through the line, starting from the beginning and picking each ball with probability $1-q$ independently of what happened with the previous balls. If we finish the line without picking a ball, we start from the beginning. Once we pick a ball, we return it to its position together with k balls of the same color. Given these rules, the probability of picking a white ball is

$$(1-q^w)\sum_{j=0}^{\infty}(q^{w+b})^j = \frac{1-q^w}{1-q^{w+b}} = \frac{[w]_q}{[w+b]_q}, \tag{1.4}$$

which is (1.2), because before picking a white ball, we will go through the entire list a random number of times, say j, without picking any ball and then, going through the white balls, we pick one (probability $1-q^w$).

If $q > 1$, we place in the line first the black balls and we go through the list picking each ball with probability $1-q^{-1}$. According to the above computation, the probability of picking a black ball is

$$\frac{[b]_{q^{-1}}}{[w+b]_{q^{-1}}} = q^w \frac{[b]_q}{[w+b]_q},$$

which is (1.3).

We extend the notion of drawing a ball from a q-Pólya urn to the case where exactly one of w, b is infinity. Then the probability to pick a white (resp. black) ball is determined again by (1.2) (resp. (1.3)), where this is understood as the limit of the right hand side as w or b goes to ∞. For example, assuming that $w = \infty$ and $b \in \mathbb{N}$, we have $\mathbf{P}_q(\text{white}) = 1$ if $q < 1$ and $\mathbf{P}_q(\text{white}) = q^{-b}$ if $q > 1$. Again these probabilities are realized through the experiment described above. Thus, we can run the process even if we start with an infinite number of balls from one color and finite from the other.

1.2 Pólya urn. Scaling limits

For the results of this section, we consider an urn whose initial composition depends on $m \in \mathbb{N}^+$. It is $A_0^{(m)}$ and $B_0^{(m)}$ white and black balls respectively. After n drawings, the composition is $A_n^{(m)}, B_n^{(m)}$.

To see a new process arising out of the path of $(A_n^{(m)})_{n\geq 0}$ we start with an initial number of balls that tends to infinity as $m \to \infty$. We assume then that $B_0^{(m)}$ grows linearly with m. Regarding $A_0^{(m)}$, we study three regimes:

a) $A_0^{(m)}$ stays fixed with m.

b) $A_0^{(m)}$ grows to infinity but sublinearly with m.

c) $A_0^{(m)}$ grows linearly with m.

The regime where $A_0^{(m)}$ grows superlinearly with m follows by regime b) by changing the roles of the two colors.

In the regimes a) and b), the scarcity of white balls has as a result that the time between two consecutive drawings of a white ball is of order $m/A_0^{(m)}$ (the probability of picking a white ball in the first few drawings is approximately $A_0^{(m)}/m$, which is small). We expect then that speeding up time by this factor we will see a birth process. And indeed this is the case as our first two theorems show.

All processes appearing in this work with index set $[0,\infty)$ and values in some Euclidean space $\mathbb{R}^d$ are elements of $D_{\mathbb{R}^d}[0,\infty)$, the space of functions $f:[0,\infty)\to\mathbb{R}^d$ that are right continuous and have limits from the left of each point of $[0,\infty)$. This space is endowed with the Skorokhod topology, and convergence in distribution of processes with values on that space is defined through that topology.

We remind the reader that the negative binomial distribution with parameters $\nu\in(0,\infty)$ and $p\in(0,1)$ is the distribution with support in $\mathbb{N}$ and probability mass function

$$f(x)=\binom{x+\nu-1}{x}p^\nu(1-p)^x \tag{1.5}$$

for all $x\in\mathbb{N}$. When $\nu\in\mathbb{N}^+$, this is the distribution of the number of failures until we see the ν-th success in a sequence of independent trials, each having probability of success p. For a random variable X with this distribution, we write $X\sim NB(\nu,p)$.

Theorem 1.1. *Fix $w_0\in\mathbb{N}^+$ and $b_0\geq 0$. If $A_0^{(m)}=w_0$ and $\lim_{m\to\infty}B_0^{(m)}/m=b_0$, then the process $(k^{-1}\{A_{[mt]}^{(m)}-A_0^{(m)}\})_{t\geq 0}$ converges in distribution, as $m\to\infty$, to an inhomogeneous in time pure birth process $Z=(Z_t)_{t\geq 0}$ such that for all $0\leq t_1<t_2, j\in\mathbb{N}$, the random variable $Z(t_2)-Z(t_1)|Z(t_1)=j$ has distribution $NB\big(\frac{w_0}{k}+j,\frac{t_1+(b_0/k)}{t_2+(b_0/k)}\big)$. Equivalently, Z has rates $\lambda_{t,j}=(kj+w_0)/(kt+b_0)$ for all $(t,j)\in[0,\infty)\times\mathbb{N}$.*

Theorem 1.2. *If $A_0^{(m)}=:g_m$ with $g_m\to\infty, g_m=o(m)$ and $\lim_{m\to\infty}B_0^{(m)}/m=b_0$ with $b_0>0$ constant, then the process $(k^{-1}\{A_{[tm/g_m]}^{(m)}-A_0^{(m)}\})_{t\geq 0}$, as $m\to\infty$, converges in distribution to the Poisson process on $[0,\infty)$ with rate $1/b_0$.*

Next, we look at regime c), i.e., in the case that at time 0 both black and white balls are of order m. In this case, the normalized process of the number of white balls has a non-random limit, which we determine, and then we study the fluctuations of the process around this limit.

Theorem 1.3. *Assume that $A_0^{(m)},B_0^{(m)}$ are such that $\lim_{m\to\infty}\frac{A_0^{(m)}}{m}=a$, $\frac{B_0^{(m)}}{m}=b$ where $a,b\in[0,\infty)$ are not both zero. Then the process $(A_{[mt]}^{(m)}/m)_{t\geq 0}$, as $m\to\infty$, converges in distribution to the deterministic process $X_t=\frac{a}{a+b}(a+b+kt), t\geq 0$.*

The limit X is the same as in an urn in which we add at each step k white or black balls with corresponding probabilities $a/(a+b), b/(a+b)$, that is, irrespective of the composition of the urn at that time.

To determine the fluctuations of the process $(A_{[mt]}^{(m)}/m)_{t\geq 0}$ around its $m\to\infty$ limit, X, we let

$$C_t^{(m)}=\sqrt{m}\bigg(\frac{A_{[mt]}^{(m)}}{m}-X_t\bigg)$$

for all $m\in\mathbb{N}^+$ and $t\geq 0$.

Theorem 1.4. *Let $a,b\in[0,\infty)$, not both zero, $\theta_1,\theta_2\in\mathbb{R}$, and assume that $A_0^{(m)}:=[am+\theta_1\sqrt{m}], B_0^{(m)}=[bm+\theta_2\sqrt{m}]$ for all large $m\in\mathbb{N}$. Then the process $(C_t^{(m)})_{t\geq 0}$ converges in distribution, as $m\to\infty$, to the unique strong solution of the stochastic differential equation*

$$Y_0=\theta_1, \tag{1.6}$$

$$dY_t=\frac{k}{a+b+kt}\bigg\{Y_t-\frac{a}{a+b}(\theta_1+\theta_2)\bigg\}\,dt+k\frac{\sqrt{ab}}{a+b}\,dW_t, \tag{1.7}$$

which is

$$Y_t=\theta_1+\frac{b\theta_1-a\theta_2}{(a+b)^2}kt+k\frac{\sqrt{ab}}{a+b}(a+b+kt)\int_0^t\frac{1}{a+b+ks}\,dW_s. \tag{1.8}$$

W is a standard Brownian motion

Remark. Functional central limit theorems for Pólya type urns have been proven with increasing generality in the works [8], [2], [10]. The major difference with our results is that in theirs, the initial number of balls, $A_0^{(m)}, B_0^{(m)}$, is fixed. More specifically:

1) Gouet ([8]) studies urns with two colors (black and white) in the setting of Bagchi and Pal ([1]). According to that, when a white ball is drawn, we return it in the urn together with a white and b black balls, while if a black ball is drawn, we return it together with c white and d black. The numbers a, b, c, d are fixed integers (possibly negative), the number of balls added to the urn is fixed (that is $a+b=c+d$), and balls are drawn uniformly form the urn. The plain Pólya urn is not studied in that work because, according to the author, it has been studied by Heyde in [9]. However, for the Pólya urn, [9] discusses the central limit theorem and the law of the iterated logarithm. In any case, following the techniques of Heyde and Gouet one can prove the following. Assume for simplicity that $k=1$ and let $L =: \lim_{n\to\infty} \frac{A_n}{n}$. The limit exists with probability one because of the martingale convergence theorem. Then

$$\left\{\sqrt{n}\left(t\frac{A_{n/t}}{n} - L\right)\right\}_{t\geq 0} \xrightarrow{d} \{W_{L'(1-L')t}\}_{t\geq 0}$$

as $n \to \infty$. W is a standard Brownian motion and L' is a random variable independent of W and having the same distribution as L. On the other hand, de-Finetti's theorem gives easily the more or less equivalent statement that, as $n \to \infty$,

$$\left\{\sqrt{n}\left(\frac{A_{nt}}{nt} - L\right)\right\}_{t\geq 0} \xrightarrow{d} \{W_{L'(1-L')/t}\}_{t\geq 0}$$

with W, L' as before.

2) Bai, Hu, and Zhang ([2]) work again in the setting of Bagchi and Pal, but now the numbers a, b, c, d depend on the order of the drawing and are random. The requirement that each time we add the same number of balls is relaxed.

3) Janson ([10]) considers urns with many colors, labeled $1, 2, \ldots, l$, where after each drawing, if we pick a ball of color i, we place in the urn balls of every color according to a random vector $(\xi_{i,1}, \ldots, \xi_{i,l})$ whose distribution depends on i ($\xi_{i,j}$ is the number of balls of color j that we add in the urn). Also, each ball is assigned a certain nonrandom activity that depends only on its color, and then the probability to pick a certain color at a drawing equals the ratio of the total of the activities of all balls of that color to the total of the activities of all balls present in the urn at that time. A restriction in that work is that there is a color i_0 so that starting the urn with just one ball and this ball has this color, there is positive probability to see in the future every other color. This excludes the classical Pólya urn that we study.

1.3 q-Pólya urn. Basic results

We recall some notation from q-calculus (see [4], [12]). For $q \in (0,\infty)\backslash\{1\}, x \in \mathbb{C}, k \in \mathbb{N}^+$, we define

$$[x]_q := \frac{q^x - 1}{q - 1} \qquad \text{the } q\text{-number of } x, \tag{1.9}$$

$$[k]_q! := [k]_q[k-1]_q \cdots [1]_q \qquad \text{the } q\text{-factorial}, \tag{1.10}$$

$$[x]_{k,q} := [x]_q[x-1]_q \cdots [x-k+1]_q \qquad \text{the } q\text{-factorial of order } k, \tag{1.11}$$

$$\begin{bmatrix} x \\ k \end{bmatrix}_q := \frac{[x]_{k,q}}{[k]_q!} \qquad \text{the } q\text{-binomial coefficient} \tag{1.12}$$

$$(x;q)_\infty := \prod_{i=0}^{\infty}(1 - xq^i) \text{ when } q \in [0,1) \qquad \text{the } q\text{-Pochhammer symbol}, \tag{1.13}$$

We extend these definitions in the case $k=0$ by letting $[0]_q! = 1, [x]_{0,q} = 1$.

Now consider a q-Pólya urn that has initially r white and s black balls, where $r \in \mathbb{N} \cup \{\infty\}$ and $s \in \mathbb{N}$. Call X_n the number of drawings that give white ball in the first n drawings. Its distribution is specified by the following.

Fact 1: Let $a := r/k$ and $b := s/k$.
(i) If $r \in \mathbb{N}$, then the probability mass function of X_n is

$$\mathbf{P}(X_n = x) = q^{k(n-x)(a+x)} \frac{\begin{bmatrix}-a\\x\end{bmatrix}_{q^{-k}} \begin{bmatrix}-b\\n-x\end{bmatrix}_{q^{-k}}}{\begin{bmatrix}-a-b\\n\end{bmatrix}_{q^{-k}}} = q^{-sx} \frac{\begin{bmatrix}a+x-1\\x\end{bmatrix}_{q^{-k}} \begin{bmatrix}b+n-x-1\\n-x\end{bmatrix}_{q^{-k}}}{\begin{bmatrix}a+b+n-1\\n\end{bmatrix}_{q^{-k}}} \tag{1.14}$$

$$= q^{-kx(b+n-x)} \frac{\begin{bmatrix}-a\\x\end{bmatrix}_{q^{k}} \begin{bmatrix}-b\\n-x\end{bmatrix}_{q^{k}}}{\begin{bmatrix}-a-b\\n\end{bmatrix}_{q^{k}}} \tag{1.15}$$

for all $x \in \mathbb{N}$.
(ii) If $r = \infty$ and $q > 1$, then the probability mass function of X_n is

$$\mathbf{P}(X_n = x) = q^{-sx}(1-q^{-k})^{n-x} \begin{bmatrix}b+n-x-1\\n-x\end{bmatrix}_{q^{-k}} \frac{[n]_{q^{-k}}!}{[x]_{q^{-k}}!} \tag{1.16}$$

for all $x \in \mathbb{N}$.

Relation (1.14) is (3.1) in [3] where it is proved through recursion. In Section 2 we give an alternative proof.

According to the experiment described in Section 1.1, the balls that are placed first in the line have an advantage to be picked (the white if $q \in (0,1)$, the black if $q > 1$). In fact, this leads to the extinction of drawings from the balls of the other color; there is a point after which the number of balls in the urn of that color stays fixed to a random number. In the next theorem, we identify the distribution of this number. We treat the case $q > 1$.

Theorem 1.5 (Extinction of the second color). *Assume that $q > 1, r \in \mathbb{N} \cup \{\infty\}, s \in \mathbb{N}$. As $n \to \infty$, with probability one, $(X_n)_{n\geq 1}$ converges to a random variable X with values in $\mathbb{N}$ and probability mass function*
(i)

$$f(x) = q^{-sx} \begin{bmatrix}\frac{r}{k}+x-1\\x\end{bmatrix}_{q^{-k}} \frac{(q^{-s};q^{-k})_\infty}{(q^{-r-s};q^{-k})_\infty} \tag{1.17}$$

for all $x \in \mathbb{N}$ in the case $r \in \mathbb{N}$ and
(ii)

$$f(x) = \left(\frac{q^{-s}}{1-q^{-k}}\right)^x \frac{1}{[x]_{q^{-k}}!} (q^{-s};q^{-k})_\infty \tag{1.18}$$

for all $x \in \mathbb{N}$ in the case $r = \infty$.

When $r \in \mathbb{N}$ and $k|r$, X has the negative q-binomial distribution of the second kind with parameters $r/k, q^{-s}, q^{-k}$ (see §3.1 in [4] for its definition). When $r = \infty$, X has the Euler distribution with parameters $q^{-s}/(1-q^{-k}), q^{-k}$ (see §3.3 in [4] again).

1.4 q-Pólya urn. Scaling limits

As in Section 1.2, we consider an urn whose composition after n drawings is $A_n^{(m)}$ white and $B_n^{(m)}$ black balls. $m \in \mathbb{N}^+$ is a parameter. Our objective is to find limits of the entire path of the process $(A_n^{(m)})_{n\in\mathbb{N}}$ analogous to the ones of Section 1.2 for the Pólya urn. Assume that $q > 1$.

If we keep q fixed, nothing new appears because: (a) If $A_0^{(m)}, B_0^{(m)}$ are fixed for all m, then after some point we pick only black balls (Theorem 1.5(i)). (b) If $\lim_{m\to\infty} B_0^{(m)} = \infty$ then the process converges to the one where we pick only black balls. (c) If $B_0^{(m)}$ is fixed for all m and $\lim_{m\to\infty} A_0^{(m)} = \infty$ then the process converges to the one where $r = \infty$ and again, after some point, we pick only black balls (Theorem 1.5(ii)).

Interesting limits appear once we take $q = q_m$ to depend on m and approach 1 as $m \to \infty$. We study two regimes for q_m. In the first, the distance of q_m from 1 is $\Theta(1/m)$ while in the second, the distance is $o(1/m)$.

1.4.1 The regime $q = 1 + \Theta(m^{-1})$

Assume that $q_m = c^{1/m}$ with $c > 1$.

Theorem 1.6. *Fix $w_0 \in \mathbb{N}^+$ and $b_0 \geq 0$. If $A_0^{(m)} = w_0$ and $\lim_{m\to\infty} B_0^{(m)}/m = b_0$, then the process $(k^{-1}(A_{[mt]}^{(m)} - A_0^{(m)}))_{t\geq 0}$ converges in distribution as $m \to \infty$ to an inhomogeneous in time pure birth process Z with starting value 0 and such that for all $0 \leq t_1 < t_2, j \in \mathbb{N}$, the random variable $Z(t_2) - Z(t_1)|Z(t_1) = j$ has distribution $NB\big(\frac{w_0}{k} + j, \frac{1-c^{-b_0-kt_1}}{1-c^{-b_0-kt_1}}\big)$. Equivalently, Z has rates*

$$\lambda_{t,j} = \frac{w_0 + jk}{c^{b_0+kt} - 1} \log c \tag{1.19}$$

for all $(t,j) \in [0,\infty) \times \mathbb{N}$.

Theorem 1.7. *Assume that $A_0^{(m)} = g_m$ and $\lim_{m\to\infty} B_0^{(m)}/m = b_0$, where $b_0 \in (0,\infty)$ and $g_m \in \mathbb{N}^+, g_m \to \infty, g_m = o(m)$ as $m \to \infty$. Then the process $(k^{-1}(A_{[tm/g_m]}^{(m)} - A_0^{(m)}))_{t\geq 0}$ converges in distribution, as $m \to \infty$, to the Poisson process on $[0,\infty)$ with rate*

$$\frac{\log c}{c^{b_0} - 1}. \tag{1.20}$$

Theorem 1.8. *Assume that $A_0^{(m)}, B_0^{(m)}$ are such that $\lim_{m\to\infty} A_0^{(m)}/m = a, \lim_{m\to\infty} B_0^{(m)}/m = b$, where $a, b \in [0,\infty)$ are not both zero. Then the process $\big(A_{[mt]}/m\big)_t \geq 0$ converges in distribution, as $m \to +\infty$, to the unique solution of the differential equation*

$$\hat{X}_0 = a, \tag{1.21}$$

$$d\hat{X}_t = k\frac{1 - c^{\hat{X}_t}}{1 - c^{a+b+kt}} dt, \tag{1.22}$$

which is

$$\hat{X}_t := a - \frac{1}{\log c} \log\left(\frac{c^b - 1 + c^{-kt}(1 - c^{-a})}{c^b - c^{-a}}\right). \tag{1.23}$$

As for the Pólya urn, we determine the fluctuations of the process $(A_{[mt]}^{(m)}/m)_{t\geq 0}$ around its $m \to \infty$ limit, $\hat{X}$. Let

$$\hat{C}_t^{(m)} = \sqrt{m}\bigg(\frac{A_{[mt]}^{(m)}}{m} - \hat{X}_t\bigg)$$

for all $m \in \mathbb{N}^+$ and $t \geq 0$.

Theorem 1.9. *Let $a, b \in [0,\infty)$, not both zero, $\theta_1, \theta_2 \in \mathbb{R}$, and assume that $A_0^{(m)} := [am + \theta_1\sqrt{m}], B_0^{(m)} = [bm + \theta_2\sqrt{m}]$ for all large $m \in \mathbb{N}$. Then the process $(\hat{C}_t^{(m)})_{t\geq 0}$ converges in distribution, as $m \to \infty$, to the unique solution of the stochastic differential equation*

$$\begin{aligned}
\hat{Y}_0 &= \theta_1, \\
d\hat{Y}_t &= \frac{k \log c}{c^{a+b+kt} - 1}\left\{\frac{(c^{a+b} - 1)\hat{Y}_t - c^b(c^a - 1)(\theta_1 + \theta_2)}{c^b - 1 + c^{-kt}(1 - c^{-a})}\right\} dt \\
&\quad + k\sqrt{(c^a - 1)(c^b - 1)}\frac{c^{(a+kt)/2}}{c^{a+b+kt} - c^{a+kt} + c^a - 1} dW_t,
\end{aligned} \tag{1.24}$$

which is

$$\begin{aligned}
\hat{Y}_t = \frac{c^{a+b+kt} - 1}{c^{a+b+kt} - c^{a+kt} + c^a - 1}\bigg(&\theta_1 - (\theta_1 + \theta_2)\frac{c^{a+b}(c^a - 1)}{c^{a+b} - 1}\frac{c^{kt} - 1}{c^{a+b+kt} - 1} \\
&+ k\sqrt{(c^a - 1)(c^b - 1)} \int_0^t \frac{c^{(a+kt)/2}}{c^{a+b+kt} - 1} dW_s\bigg).
\end{aligned} \tag{1.25}$$

W is a standard Brownian motion

1.4.2 The regime $q = 1 + o(m^{-1})$

In this regime, we let $q = q(m) := c^{\varepsilon_m/m}$ where $c > 1$ and $\varepsilon_m \to 0^+$ as $m \to \infty$. With computations analogous to those of the results of the previous subsection, it is easy to see that Theorems 1.1, 1.2 , 1.3, 1.4 hold exactly the same for the q-Pólya urn in this regime.

1.5 q-Pólya urn with many colors

In this paragraph, we give a q-analog for the Pólya urn with more than two colors. The way to do the generalization is inspired by the experiment we used in order to explain relation (1.2).

Let $l \in \mathbb{N}, l \geq 2$, and $q \in (0,1)$. Assume that we have an urn containing w_i balls of color i for each $i \in \{1, 2, \dots, l\}$. To draw a ball from the urn, we do the following. We order the balls in a line, first those of color 1, then those of color 2, and so on. Then we visit the balls, one after the other, in the order that they have been placed, and we select each with probability $1 - q$ independently of what happened with the previous balls. If we go through all balls without picking any, we repeat the same procedure starting from the beginning of the line. Once a ball is selected, the drawing is completed. We return the ball to its position together with another k of the same color. For each $i = 0, 1, \dots, l$, let $s_i = \sum_{1\leq j\leq i} w_j$. Notice that s_l is the total number of balls in the urn. Then, working as for (1.4), we see that

$$\mathbf{P}(\text{color } i \text{ is drawn}) = q^{s_{i-1}} \frac{1 - q^{w_i}}{1 - q^{s_l}} = \frac{q^{s_{i-1}} - q^{s_i}}{1 - q^{s_l}} = q^{s_{i-1}} \frac{[w_i]_q}{[s_l]_q}. \tag{1.26}$$

Call p_i the number in the last display for all $i = 1, 2, \dots, l$. Note that when $q \to 1$, p_i converges to w_i/s_l, which is the probability for the usual Pólya urn with l colors. It is clear that for any given $q \in (0, \infty)\backslash\{1\}$, the numbers $p_1, p_2, \dots, p_l$ are non-negative and add to 1 (the second fraction in (1.26) shows this). We define then for this q the q-Pólya urn with colors $1, 2, \dots, l$ to be the sequential procedure in which, at each step, we add k balls of a color picked randomly among $\{1, 2, \dots, l\}$ so that the probability that this color is i is p_i .

When $q > 1$, these probabilities come out of the experiment described above but in which we place the balls in reverse order (that is, first those of color l, then those of color $l - 1$, and so on) and we go through the list selecting each ball with probability $1 - q^{-1}$. It is then easy to see that the probability to pick a ball of color i is p_i.

Theorem 1.10. *Assume that $q \in (0,1)$ and that we start with $a_1, a_2, \dots, a_l$ balls from colors 1, 2, …, l respectively, where $a_1, a_2, \dots, a_l \in \mathbb{N}$ are not all zero. Call $X_{n,i}$ the number of times in the first n drawings that we picked color i. The probability mass function for the vector $(X_{n,2}, X_{n,3}, \dots, X_{n,l})$ is*

$$\mathbf{P}\left(X_{n,2} = x_2, \dots, X_{n,l} = x_l\right) = q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1}(a_j + k x_j)} \frac{\prod_{i=1}^{l} \left[\begin{smallmatrix} -\frac{a_i}{k} \\ x_i \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} -\frac{a_1 + a_2 \dots + a_l}{k} \\ n \end{smallmatrix}\right]_{q^{-k}}} \tag{1.27}$$

$$= \begin{bmatrix} n \\ x_1, x_2, \dots, x_l \end{bmatrix}_{q^{-k}} \frac{q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1}(a_j + k x_j)} \prod_{i=1}^{l} \left[-\frac{a_i}{k}\right]_{x_i, q^{-k}}}{\left[-\frac{a_1 + a_2 + \dots + a_l}{k}\right]_{n, q^{-k}}} \tag{1.28}$$

for all $x_2, \dots, x_l \in \{0, 1, 2, \dots, n\}$ with $x_2 + \cdot + x_l \leq n$, where $x_1 := n - \sum_{i=2}^{l} x_2$ and $\left[\begin{smallmatrix} n \\ x_1, x_2, \dots, x_l \end{smallmatrix}\right]_{q^{-k}} := \frac{[n]_{q^{-k}}!}{[x_1]_{q^{-k}}! \cdot \dots \cdot [x_l]_{q^{-k}}!}$ is the q-multinomial coefficient.

It follows from Theorem 1.5 that when $q \in (0, 1)$, after some random time, we will be picking only balls of color 1. So that the number of times that we pick each of the other colors $2, 3, \dots, l$, say $X_2, X_3, \dots, X_n$ are finite. We determine the joint distribution of these numbers.

Theorem 1.11. *Under the assumptions of Theorem 1.10, as $n \to +\infty$, with probability one, the vector $(X_{n,2}, X_{n,3}, \dots, X_{n,l})$ converges to a random vector $(X_2, X_3, \dots, X_l)$ with values in $\mathbb{N}^{l-1}$ and probability* *mass function*

$$f(x_2, x_3, \ldots, x_l) = q^{\sum_{i=2}^{l} x_i \sum_{j=1}^{i-1} a_j} \prod_{i=2}^{l} \begin{bmatrix} x_i + \frac{a_i}{k} - 1 \\ x_i \end{bmatrix}_{q^k} \frac{(q^{a_1}; q^k)_\infty}{(q^{a_1+\cdots+a_l}; q^k)_\infty} \tag{1.29}$$

for all $x_2, \ldots, x_l \in \mathbb{N}$.

Note that the random variables $X_2, \ldots, X_l$ are independent although $(X_{n,2}, X_{n,3}, \ldots, X_{n,l})$ are dependent.

Next, we look for a scaling limit for the path of the process. Assume that $c \in (0,1)$ and $q_m = c^{1/m}$. Let $A_{j,i}^{(m)}$ be the number of balls of color i in this urn after j drawings.

Theorem 1.12. *Let m be a positive integer and assume that in the q-Pólya urn with l different colors of balls it holds $\frac{1}{m}\left(A_{0,1}^{(m)}, A_{0,2}^{(m)}, \ldots, A_{0,l}^{(m)}\right) \overset{m\to\infty}{\to} (a_1, a_2, \ldots, a_l)$, where $a_1, \ldots, a_l \in [0, \infty)$ are not all zero. Set $\sigma_0 = 0$ and $\sigma_i := \sum_{j \le i} a_j$ for all $i = 1, 2, \ldots, l$. Then the process $\left(\frac{1}{m}A_{[mt],1}^{(m)}, \frac{1}{m}A_{[mt],2}^{(m)}, \ldots, \frac{1}{m}A_{[mt],l}^{(m)}\right)_{t\ge 0}$ converges in distribution, as $m \to +\infty$, to $(X_{t,1}, X_{t,2}, \ldots, X_{t,l})_{t\ge0}$ with*

$$X_{t,i} = a_i + \frac{1}{\log c} \log \frac{(1 - c^{\sigma_l + kt}) - c^{\sigma_{i-1}}(1 - c^{kt})}{(1 - c^{\sigma_l + kt}) - c^{\sigma_i}(1 - c^{kt})} \tag{1.30}$$

for all $i = 1, 2, \ldots, l$.

As in the case of two colors, we study the regime where $q_m = c^{\epsilon_m/m}$, with $c \in (0,1)$ and $\epsilon_m \to 0^+$.

Theorem 1.13. *Let m be a positive integer and assume that in the q-Pólya urn with l different colors of balls that $\frac{1}{m}\left(A_{0,1}^{(m)}, A_{0,2}^{(m)}, \ldots, A_{0,l}^{(m)}\right) \overset{m\to\infty}{\to} (a_1, a_2, \ldots, a_l)$, where $a_1, \ldots, a_l \in [0, \infty)$ are not all zero. Then the process $\left(\frac{1}{m}A_{[mt],1}^{(m)}, \frac{1}{m}A_{[mt],2}^{(m)}, \ldots, \frac{1}{m}A_{[mt],l}^{(m)}\right)_{t\ge0}$ converges in distribution, as $m \to +\infty$, to $(X_t)_{t\ge0}$ with*

$$X_t = \left(1 + \frac{kt}{a_1 + \cdots + a_l}\right)(a_1, a_2, \ldots, a_l) \tag{1.31}$$

for all $t \ge 0$.

Orientation. In Section 2, we prove Fact 1 and Theorem 1.5, which are basic results for the q-Pólya urn. Section 3 (Section 4) contains the proofs of the theorems for the Pólya and q-Pólya urns that give convergence to a jump process (to a continuous process). Finally, Section 5 contains the proofs for the results that refer to the q-Pólya urn with arbitrary, finite number of colors.

2 q-Pólya urn. Prevalence of a single color

In this section, we prove the claims of Section 1.3. Before doing so, we mention three properties of the q-binomial coefficient. For all $q \in (0,\infty)\backslash\{1\}, x \in \mathbb{C}, n, k \in \mathbb{N}$ with $k \le n$ it holds

$$[-x]_q = -q^{-x}[x]_q, \tag{2.1}$$

$$\begin{bmatrix} -x \\ k \end{bmatrix}_q = (-1)^k q^{-k(k+2x-1)/2} \begin{bmatrix} x+k-1 \\ k \end{bmatrix}_q, \tag{2.2}$$

$$\begin{bmatrix} x \\ k \end{bmatrix}_{q^{-1}} = q^{-k(x-k)} \begin{bmatrix} x \\ k \end{bmatrix}_q, \tag{2.3}$$

$$\sum_{1\le i_1<i_2<\cdots<i_k\le n} q^{i_1+i_2+\cdots+i_k} = q^{\binom{k+1}{2}} \begin{bmatrix} n \\ k \end{bmatrix}_q. \tag{2.4}$$

The first is trivial, the second follows from the first, the third is easily shown, while the last is Theorem 6.1 in [12].

Proof of Fact 1. (i) The probability to get black balls exactly at the drawings $i_1 < i_2 < \cdots < i_{n-x}$ is

$$g(i_1, i_2, \ldots, i_{n-x}) = \frac{\prod_{j=0}^{x-1}[r+jk]_q \prod_{j=0}^{n-x-1}[s+jk]_q}{\prod_{j=0}^{n-1}[r+s+jk]_q} q^{\sum_{\nu=1}^{n-x} r+(i_\nu-\nu)k}. \tag{2.5}$$

To see this, note that, due to (1.2) and (1.3), the required probability would be equal to the above fraction if in (1.3) the term q^w were absent. This term appears whenever we draw a black ball. Now, when we draw the ν-th black ball, there are $r+(i_\nu-\nu)k$ white balls in the urn, and this explains the exponent of q in (2.5).

Since $[x+jk]_q = \frac{1-q^{x+jk}}{1-q} = [-\frac{x}{k}-j]_{q^{-k}}[-k]_q$ for all $x, j \in \mathbb{R}$, the fraction in (2.5) equals

$$\frac{[-a]_{x,q^{-k}}[-b]_{n-x,q^{-k}}}{[-a-b]_{n,q^{-k}}}. \tag{2.6}$$

Then

$$\sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} q^{\sum_{\nu=1}^{n-x} r+(i_\nu-\nu)k} = q^{(n-x)r-k(n-x)(n-x+1)/2} \sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} (q^k)^{i_1+i_2+\cdots+i_{n-x}} \tag{2.7}$$

$$= q^{(n-x)r-k(n-x)(n-x+1)/2} q^{k\binom{n-x+1}{2}} \begin{bmatrix} n \\ x \end{bmatrix}_{q^k} \tag{2.8}$$

$$= q^{(n-x)r} q^{kx(n-x)} \begin{bmatrix} n \\ x \end{bmatrix}_{q^{-k}} = q^{k(n-x)(a+x)} \begin{bmatrix} n \\ x \end{bmatrix}_{q^{-k}}. \tag{2.9}$$

The second equality follows from (2.4) and the equality $\left[\begin{smallmatrix} n \\ x \end{smallmatrix}\right]_{q^k} = \left[\begin{smallmatrix} n \\ n-x \end{smallmatrix}\right]_{q^k}$. The third, from (2.3). Thus, the sum $\sum_{1\le i_1<i_2<\cdots<i_{n-x}\le n} g(i_1, i_2, \ldots, i_{n-x})$ equals the first expression in (1.14). The second expression in (1.14) and (1.15) follow by using (2.2) and (2.3) respectively.

(ii) In this scenario, we take $r \to \infty$ in the last expression in (1.14). We will explain shortly why this gives the probability we want. Since $q^{-k} \in (0,1)$, we have $\lim_{t\to\infty}[t]_{q^{-k}} = (1-q^{-k})^{-1}$ and thus, for each $\nu \in \mathbb{N}$, it holds

$$\lim_{t\to\infty} \begin{bmatrix} t+\nu-1 \\ \nu \end{bmatrix}_{q^{-k}} = \frac{1}{[\nu]_{q^{-k}}!} \frac{1}{(1-q^{-k})^\nu}. \tag{2.10}$$

Applying this twice in the last expression in (1.14) (there $a = r/k \to \infty$), we get as limit the right hand side of (1.16).

Now, to justify that passage to the limit $r \to \infty$ in (1.14) gives the required result, we argue as follows. For clarity, denote the probability $\mathbf{P}_q(\text{white})$ when there are w white and b black balls in the urn by $\mathbf{P}_q^{w,b}(\text{white})$. And when there are r white and s black balls in the urn in the beginning of the procedure, denote the probability of the event $X_n = x$ by $\mathbf{P}^{r,s}(X_n = x)$. It is clear that the probability $\mathbf{P}^{r,s}(X_n = x)$ is a continuous function (in fact, a polynomial) of the quantities

$$\mathbf{P}_q^{r+ki,s+kj}(\text{white}) : i = 0, 1, \ldots, x-1, j = 0, 1, \ldots, n-x-1,$$

for all values of $r \in \mathbb{N} \cup \{\infty\}, s \in \mathbb{N}$. In $\mathbf{P}^{\infty,s}(X_n = x)$, each such quantity, $\mathbf{P}_q^{\infty,m}(\text{white})$, equals $\lim_{r\to\infty} \mathbf{P}^{r,m}(\text{white})$. Thus, $\mathbf{P}^{\infty,s}(X_n = x) = \lim_{r\to\infty} \mathbf{P}^{r,s}(X_n = x)$. ∎

Before proving Theorem 1.5, we give a simple argument that shows that eventually we will be picking only black balls. That is, the number $X := \lim_{n\to\infty} X_n$ of white balls drawn in an infinite sequence of drawings is finite. It is enough to show it in the case that $r = \infty$ and $s = 1$ since, by the experiment that realizes the q-Pólya urn, we have (using the notation from the proof of Fact 1 (ii))

$$\mathbf{P}^{r,s}(X = \infty) \le \mathbf{P}^{\infty,1}(X = \infty).$$

For each $n \in \mathbb{N}^+$, call E_n the event that at the n-th drawing we pick a white ball, B_n the number of black balls present in the urn after that drawing (also, $B_0 := 1$), and write $\hat{q} := 1/q$. Then $\mathbf{P}(E_n) =$

$\mathbf{E}(\mathbf{P}(E_n|B_{n-1})) = \mathbf{E}(\hat{q}^{B_{n-1}})$. We will show that this decays exponentially with n. Indeed, since at every drawing there is probability at least $1-\hat{q}$ to pick a black ball, we can construct in the same probability space the random variables $(B_n)_{n\geq 1}$ and $(Y_i)_{i\geq 1}$ so that the Y_i are i.i.d. with $Y_1 \sim$ Bernoulli$(1-\hat{q})$ and $B_n \geq 1 + k(Y_1 + \cdots + Y_n)$ for all $n \in \mathbb{N}^+$. Consequently,

$$\mathbf{P}(E_n) \leq \mathbf{E}(\hat{q}^{1+k(Y_1+\cdots+X_{n-1})}) = \hat{q}\{\mathbf{E}(\hat{q}^{kY_1})\}^{n-1}.$$

This implies that $\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty$, and the first Borel-Cantelli lemma gives that $\mathbf{P}^{\infty,1}(X_\infty = \infty) = 0$.

Proof of Theorem 1.5. Since $(X_n)_{n\geq 1}$ is increasing, it converges to a random variable X with values in $\mathbb{N} \cup \{\infty\}$. In particular, it converges to this variable in distribution. Our aim is to take the limit as $n \to \infty$ in the last expression in (1.14) and in (1.16) in order to determine the distribution of X. Note that for $a \in \mathbb{R}$ and $\theta \in [0,1)$ it is immediate that (recall (1.13) for the notation)

$$\lim_{n\to\infty} \begin{bmatrix} a+n \\ n \end{bmatrix}_\theta = \frac{(\theta^{a+1};\theta)_\infty}{(\theta;\theta)_\infty}. \tag{2.11}$$

(i) Taking $n \to \infty$ in the last expression in (1.14) and using (2.11), we get the required expression, (1.17), for f. Then relation (2.2) in [3] (or (8.1) in [12]) shows that $\sum_{x\in\mathbb{N}} f(x) = 1$, so that it is a probability mass function of a random variable X with values in $\mathbb{N}$.
(ii) This follows after taking limit in (1.16) and using (2.11) and $\lim_{n\to\infty}(1-q^{-k})^n [n]_{q^{-k}}! = (q^{-k};q^{-k})_\infty$. ∎

3 Jump process limits. Proof of Theorems 1.1, 1.2, 1.6, 1.7

In the case of Theorems 1.1, 1.6, we let $g_m := 1$ for all $m \in \mathbb{N}^+$, and in all four theorems we let $v := v_m := m/g_m$. Our interest is in the sequence of the processes $(Z^{(m)})_{m\geq 1}$ with

$$Z^{(m)}(t) = \frac{1}{k}(A^{(m)}_{[vt]} - A^{(m)}_0) \tag{3.1}$$

for all $t \geq 0$.

We apply Theorem 7.8 in [6], that is, we show that the sequence $(Z^{(m)})_{m\geq 1}$ is tight and its finite dimensional distributions converge. Tightness gives that there is a subsequence of this sequence that converges in distribution to a process $Z = (Z_t)_{t\geq 0}$ with paths in the space $D_{\mathbb{R}}[0,\infty)$ of real valued functions on $[0,\infty)$ right continuous with left limits. Then tightness together with convergence of finite dimensional distributions shows that the whole sequence $(Z^{(m)})_{m\geq 1}$ converges in distribution to Z.

Notation: (i) For sequences $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ with values in $\mathbb{R}$, we will say that they are asymptotically equivalent, and will write $a_n \sim b_n$ as $n \to \infty$, if $\lim_{n\to\infty} a_n/b_n = 1$. We use the same expressions for functions f, g defined in a neighborhood of ∞ and satisfy $\lim_{x\to\infty} f(x)/g(x) = 1$.
(ii) For $a \in \mathbb{C}$ and $k \in \mathbb{N}^+$, let

$$(a)_k := a(a-1)\cdots(a-k+1), \tag{3.2}$$
$$a^{(k)} := a(a+1)\cdots(a+k-1), \tag{3.3}$$

the falling and rising factorial respectively. Also let $(a)_0 := a^{(0)} := 1$.

3.1 Convergence of finite dimensional distributions

Since for each $m \geq 1$ the process $Z^{(m)}$ is Markov taking values in $\mathbb{N}$ and increasing in time, it is enough to show that the conditional probability

$$\mathbf{P}(Z^{(m)}(t_2) = k_2 | Z^{(m)}(t_1) = k_1) \tag{3.4}$$

converges as $m \to \infty$ for each $0 \le t_1 < t_2$ and nonnegative integers $k_1 \le k_2$.
Consider first the case of Pólya urn and define

$$n := [vt_2] - [vt_1], \tag{3.5}$$

$$x := k_2 - k_1, \tag{3.6}$$

$$\sigma := \frac{A_0^{(m)} + kk_1}{k}, \tag{3.7}$$

$$\tau := \frac{k[vt_1] - kk_1 + B_0^{(m)}}{k}. \tag{3.8}$$

Then, the above probability equals

$$\begin{aligned}
&\mathbf{P}(A_{[vt_2]}^{(m)} = kk_2 + w_0 | A_{[vt_1]}^{(m)} = kk_1 + w_0) \\
&= \binom{n}{x} \frac{k\sigma(k\sigma+k)(k\sigma+2k)\cdots(k\sigma+(x-1)k)k\tau(k\tau+k)(k\tau+2k)\cdots(k\tau+(n-x-1)k)}{(k\sigma+k\tau)(k\sigma+k\tau+k)(k\sigma+k\tau+2k)\cdots(k\sigma+k\tau+(n-1)k)} \qquad (3.9)\\
&= \frac{(n)_x}{x!}\frac{\sigma^{(x)}\tau^{(n-x)}}{(\sigma+\tau)^{(n)}} = \frac{(n)_x}{x!}\sigma^{(x)}\frac{\Gamma(\tau+n-x)}{\Gamma(\tau)}\frac{\Gamma(\sigma+\tau)}{\Gamma(\sigma+\tau+n)}. \qquad (3.10)
\end{aligned}$$

To compute the limit as $m \to \infty$ of (3.10), we will use Stirling's approximation for the Gamma function,

$$\Gamma(y) \sim \left(\frac{y}{e}\right)^y \sqrt{\frac{2\pi}{y}} \tag{3.11}$$

as $y \to \infty$, and its consequence

$$\Gamma(y+a) \sim \Gamma(y) y^a \tag{3.12}$$

as $y \to \infty$ for all $a \in \mathbb{R}$.

Theorem 1.1. Recall that $v = m$ in this theorem. Using (3.12) twice, with the role of a played by $-x$ and σ, we see that the last quantity in (3.10), for $m \to \infty$, is asymptotically equivalent to

$$\begin{aligned}
&\frac{(m(t_2-t_1))^x}{x!}\sigma^{(x)}\tau^{\sigma}\frac{(\tau+n)^{-x}}{(\tau+n)^{\sigma}} \sim \frac{(m(t_2-t_1))^x}{x!}\sigma^{(x)}\frac{\{m(t_1+(b_0/k))\}^{\sigma}}{\{m(t_2+(b_0/k))\}^{\sigma+x}} \\
&= \frac{(t_2-t_1)^x}{x!}\sigma^{(x)}\frac{\{t_1+(b_0/k)\}^{\sigma}}{\{t_2+(b_0/k)\}^{\sigma+x}} = \binom{\sigma+x-1}{x}\left(\frac{t_2-t_1}{t_2+(b_0/k)}\right)^x\left(1-\frac{t_2-t_1}{t_2+(b_0/k)}\right)^{\sigma}.
\end{aligned}$$

Thus, as $m \to \infty$, the distribution of $\{Z^{(m)}(t_2) - Z^{(m)}(t_1)\}|Z^{(m)}(t_1) = k_1$ converges to the negative binomial distribution with parameters σ, $\frac{t_1+(b_0/k)}{t_2+(b_0/k)}$ [recall (1.5)]. ■

Theorem 1.2. Using (3.11), we see that the last quantity in (3.10), for $m \to \infty$, is asymptotically equivalent to

$$\begin{aligned}
&\frac{(m(t_2-t_1))^x}{x! g_m^x}\frac{g_m^x}{k^x}e^x\frac{(\tau+n-x)^{\tau+n-x}}{\tau^{\tau}}\frac{(\sigma+\tau)^{\sigma+\tau}}{(\sigma+\tau+n)^{\sigma+\tau+n}} \\
&\sim \frac{m^x(t_2-t_1)^x}{x!k^x}e^x(\tau+n-x)^{-x}\left(\frac{\tau+n-x}{\sigma+\tau+n}\right)^n\left(\frac{\sigma+\tau}{\sigma+\tau+n}\right)^{\sigma}\left(\frac{(\tau+n-x)(\sigma+\tau)}{\tau(\sigma+\tau+n)}\right)^{\tau} \\
&\sim \frac{m^x(t_2-t_1)^x}{x!k^x}e^x\tau^{-x}e^{-(t_2-t_1)/b_0}e^{-(t_2-t_1)/b_0}e^{-x+(t_2-t_1)/b_0} \sim \frac{1}{x!}\left(\frac{t_2-t_1}{b_0}\right)^x e^{-(t_2-t_1)/b_0}.
\end{aligned}$$

Here it was crucial that $b_0 > 0$. Thus, as $m \to \infty$, the distribution of $\{Z^{(m)}(t_2) - Z^{(m)}(t_1)\}|Z^{(m)}(t_1) = k_1$ converges to the Poisson distribution with parameter $(t_2 - t_1)/b_0$. ■

Now we treat the cases of Theorems 1.6, 1.7, which concern the q-Pólya urn. Define again n, x, σ, τ as in (3.5)-(3.8), and $r := q_m^{-k} = c^{-k/m}$. Then, the probability in (3.4), with the help of the last expression in (1.14), is computed as

$$r^{\tau x}\begin{bmatrix}\sigma+x-1 \\ x\end{bmatrix}_r \frac{\begin{bmatrix}\tau+n-x-1 \\ n-x\end{bmatrix}_r}{\begin{bmatrix}\sigma+\tau+n-1 \\ n\end{bmatrix}_r} = r^{\tau x}\begin{bmatrix}\sigma+x-1 \\ x\end{bmatrix}_r\Big(\prod_{i=n-x+1}^{n}(1-r^i)\Big)\frac{1}{\prod_{i=n-x}^{n-1}(1-r^{\tau+i})}\frac{[\tau+n-1]_{n,r}}{[\sigma+\tau+n-1]_{n,r}}. \tag{3.13}$$

The last ratio is

$$\prod_{i=0}^{n-1}\frac{1-r^{\tau+i}}{1-r^{\sigma+\tau+i}}=\prod_{i=0}^{n-1}\left(1-(1-r^{\sigma})r^{\tau}\frac{r^{i}}{1-r^{\sigma+\tau+i}}\right). \tag{3.14}$$

Denote by $1-a_{m,i}$ the i-th term of the product. The logarithm of the product equals

$$-(1-r^{\sigma})r^{\tau}\sum_{i=0}^{n-1}\frac{r^{i}}{1-r^{\sigma+\tau+i}}+o(1) \tag{3.15}$$

as $m\to\infty$. To justify this, note that $1-r^{\sigma}\sim\frac{1}{m}(A_0^{(m)}+kk_1)\log c$ and $r^{\tau+i}/(1-r^{\sigma+\tau+i})\le 1/(1-c^{-b_0})$ for all $i\in\mathbb{N}$. Thus, for all large m, $|a_{m,i}|<1/2$ for all $i=0,1,\dots,n-1$, and the error in approximating the logarithm of $1-a_{m,i}$ by $-a_{m,i}$ is at most $|a_{m,i}|^2$ (by Taylor's expansion, we have $|\log(1-y)+y|\le|y|^2$ for all y with $|y|\le 1/2$). The sum of all errors is at most $n\max_{0\le i<n}|a_{m,i}|^2$, which goes to zero as $m\to\infty$ because $1-r^{\sigma}\sim C/n$ for some appropriate constant $C>0$.

We will compute the limit of the right hand side of (3.13) as $m\to\infty$ under the assumptions of Theorems 1.6, 1.7.

Theorem 1.6. As $m\to\infty$, the first term of the product in (3.13) converges to $c^{-x(b_0+kt_1)}$. The q-binomial coefficient converges to $\binom{k^{-1}w_0+k_2-1}{k_2-k_1}$. The third term converges to $(1-c^{-k(t_2-t_1)})^x$, while the denominator of the fourth term converges to $(1-\rho_2)^x$, where we set $\rho_i:=c^{-b_0-kt_i}$ for $i=1,2$. The expression preceding $o(1)$ in (3.15) is asymptotically equivalent to

$$-\frac{k}{m}\sigma(\log c)\rho_1\sum_{i=0}^{n-1}\frac{c^{-ki/m}}{1-r^{\sigma+\tau}c^{-ki/m}}=-\rho_1k\sigma(\log c)\frac{1}{m}\sum_{i=0}^{n-1}\frac{c^{-ki/m}}{1-\rho_1c^{-ki/m}}+o(1) \tag{3.16}$$

$$=-\rho_1k\sigma\log c\int_0^{t_2-t_1}\frac{1}{c^{ky}-\rho_1}\,dy+o(1)=\sigma\log\frac{1-\rho_1}{1-\rho_2}+o(1). \tag{3.17}$$

The equality in the first line is true because $\lim_{m\to\infty}r^{\sigma+\tau}=\rho_1$ and the function $x\mapsto c^{-ki/m}/(1-xc^{-ki/m})$ has derivative bounded uniformly in i,m when x is confined to a compact subset of $[0,1)$. Thus, the limit of (3.13), as $m\to\infty$, is

$$\binom{\sigma+x-1}{x}\left(\frac{\rho_1-\rho_2}{1-\rho_2}\right)^{x}\left(\frac{1-\rho_1}{1-\rho_2}\right)^{\sigma}, \tag{3.18}$$

which means that, as $m\to\infty$, the distribution of $\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1)=k_1$ converges to the negative binomial distribution with parameters $\sigma,(1-\rho_1)/(1-\rho_2)$. ■

Theorem 1.7. Now the term $r^{\tau x}$ converges to c^{-xb_0}, while

$$\begin{bmatrix}\sigma+x-1\\ x\end{bmatrix}_r\Big(\prod_{i=n-x+1}^{n}(1-r^{i})\Big)=\frac{\prod_{i=0}^{x-1}(1-r^{\sigma+i})}{\prod_{i=1}^{x}(1-r^{i})}\Big(\prod_{i=n-x+1}^{n}(1-r^{i})\Big) \tag{3.19}$$

$$\sim\frac{\prod_{i=0}^{x-1}(\sigma+i)}{\prod_{i=1}^{x}i}\frac{((t_2-t_1)k\log c)^x}{g_m^x}\sim\frac{1}{x!}((t_2-t_1)\log c)^x. \tag{3.20}$$

The denominator of the fourth term in (3.13) converges to $(1-c^{-b_0})^x$. The expression in (3.15) is asymptotically equivalent to

$$-r^{\tau}(1-r^{\sigma})\sum_{i=0}^{n-1}\frac{r^{i}}{1-r^{\sigma+\tau+i}}\sim-c^{-b_0}\frac{g_m}{m}\log c\frac{n}{1-c^{-b_0}}\sim-\frac{\log c}{c^{b_0}-1}(t_2-t_1). \tag{3.21}$$

In the first $\sim$, we used the fact that the terms of the sum, as $m\to\infty$, converge uniformly in i to $(1-c^{-b_0})^{-1}$. Thus, the limit of (3.13), as $m\to\infty$, is

$$\frac{1}{x!}\left(\frac{\log c}{c^{b_0}-1}(t_2-t_1)\right)^{x}e^{-\frac{\log c}{c^{b_0}-1}(t_2-t_1)}, \tag{3.22}$$

which means that, as $m\to\infty$, the distribution of $\{Z^{(m)}(t_2)-Z^{(m)}(t_1)\}|Z^{(m)}(t_1)=k_1$ converges to the Poisson distribution with parameter $\frac{t_2-t_1}{c^{b_0}-1}\log c$. ■

For use in the following section, we define

$$U(t_1, t_2, k_1, x) := \lim_{m\to\infty} \mathbf{P}(Z^{(m)}(t_2) = k_1 + x | Z^{(m)}(t_1) = k_1) \tag{3.23}$$

for all $0 \le t_1 \le t_2, k_1 \in \mathbb{N}, x \in \mathbb{N}$. The results of this section show that U as a function of $x \in \mathbb{N}$ is a probability mass function of an appropriate random variable with values in $\mathbb{N}$.

3.2 Tightness

We apply Corollary 7.4 of Chapter 3 in [6]. According to it, it is enough to show that

(i) For each $t \ge 0$, it holds $\lim_{R\to\infty} \overline{\lim}_{m\to\infty} \mathbf{P}(|Z^{(m)}(t)| \ge R) = 0$.

(ii) For each $T, \varepsilon > 0$, it holds $\lim_{\delta\to 0} \overline{\lim}_{m\to\infty} \mathbf{P}(w'(Z^{(m)}, \delta, T) \ge \varepsilon) = 0$.

Here, for any function $f : [0, \infty) \to \mathbb{R}$, we define

$$w'(f, \delta, T) := \inf_{\{t_i\}} \max_i \sup_{s,t\in[t_{i-1},t_i)} |f(s) - f(t)|,$$

where the infimum is over all partitions of the form $0 = t_0 < t_1 < \cdots t_{n-1} < T \le t_n$ with $t_i - t_{i-1} > \delta$ for all $i = 1, 2, \dots, n$.

The first requirement holds because $Z^{(m)}(t)$ converges in distribution as we showed in the previous subsection. The second requirement, since $Z^{(m)}$ is a jump process with jump sizes only 1, is equivalent to

$$\lim_{\delta\to 0^+} \overline{\lim_{m\to\infty}} \mathbf{P}(\text{There are at least two jump times of } Z^{(m)} \text{ in } [0, T] \text{ with distance} \le \delta) = 0. \tag{3.24}$$

Call $A_{m,\delta}$ the event inside the probability and for $j = 1, 2, \dots, [T/\delta]$ define $I_j := ((j-1)\delta, (j+1)\delta]$. Then, for each $\ell \in \mathbb{N}$, the probability $\mathbf{P}(A_{m,\delta} \cap \{Z^{(m)}(T) \le \ell\})$ is bounded above by

$$\sum_{j=1}^{[T/\delta]} \mathbf{P}(\{Z^{(m)}(T) \le \ell\} \cap \{\text{There are at least two jump times of } Z^{(m)} \text{ in } I_j\}) \tag{3.25}$$

$$\le \sum_{j=1}^{[T/\delta]} \mathbf{P}(\{Z^{(m)}(T) \le \ell\} \cap \{Z^{(m)}((j+1)\delta) - Z^{(m)}((j-1)\delta) \ge 2\}) \tag{3.26}$$

$$\le \sum_{j=1}^{[T/\delta]} \max_{0\le\mu\le\ell} \mathbf{P}(Z^{(m)}((j+1)\delta) - Z^{(m)}((j-1)\delta) \ge 2 | Z^{(m)}((j-1)\delta) = \mu). \tag{3.27}$$

The limit of the last quantity as $m \to \infty$, with the use of the function U of (3.23), is written as

$$\sum_{j=1}^{[T/\delta]} \max_{0\le\mu\le\ell} \sum_{x=2}^{\infty} U((j-1)\delta, (j+1)\delta, \mu, x) \le \frac{T}{\delta} \max_{\substack{0\le\mu\le\ell \\ 1\le j\le[T/\delta]}} \sum_{x=2}^{\infty} U((j-1)\delta, (j+1)\delta, \mu, x). \tag{3.28}$$

Claim: The max in (3.28) is bounded above by $\delta^2 C(\ell, T)$ for an appropriate constant $C(\ell, T) \in (0, \infty)$ that does not depend on m or δ.

Assuming the claim and taking $m \to \infty$ in $\mathbf{P}(A_{m,\delta}) = \mathbf{P}(A_{m,\delta} \cap \{Z^{(m)}(T) \le \ell\}) + \mathbf{P}(A_{m,\delta} \cap \{Z^{(m)}(T) > \ell\})$, we get

$$\overline{\lim_{m\to\infty}} \mathbf{P}(A_{m,\delta}) \le \delta C(\ell, T) + \overline{\lim_{m\to\infty}} \mathbf{P}(\{Z^{(m)}(T) > \ell\}).$$

Now let $\varepsilon > 0$. Because of the validity of (i) in the tightness requirements, there is ℓ large enough so that the second term is $< \varepsilon$. Fixing this ℓ and taking $\delta \to 0$ in the inequality, we get (3.24).

Proof of the claim: We establish the above claim for each of the Theorems 1.1, 1.2, 1.6, 1.7. We use the following bounds. If X, Y are random variables with $X \sim \text{Poisson}(\lambda)$ and $Y \sim NB(\nu, p)$ then

$$\mathbf{P}(X \ge 2) \le \lambda^2, \tag{3.29}$$

$$\mathbf{P}(Y \ge 2) \le \frac{\nu(\nu+1)}{2}(1-p)^2. \tag{3.30}$$

The first inequality is elementary, while the second is true because the difference of the two sides

$$\mathbf{P}(Y \geq 2) - \frac{\nu(\nu+1)}{2}(1-p)^2 = 1 - p^\nu - rp^\nu(1-p) - \frac{\nu(\nu+1)}{2}(1-p)^2$$

is an increasing function of p in $[0,1]$ with value 0 at $p = 1$.

According to the results of Section 3.1, the sum after the max in (3.28) equals $\mathbf{P}(X \geq 2)$ where

$$X \sim \begin{cases} NB\big(\frac{w_0}{k} + \mu, \frac{t_1 + (b_0/k)}{t_2 + (b_0/k)}\big) & \text{for Theorem 1.1,} \\ \text{Poisson}\big(\frac{2\delta}{b_0}\big) & \text{for Theorem 1.2,} \\ NB\big(\frac{w_0}{k} + \mu, \frac{1-c^{-b_0-kt_1}}{1-c^{-b_0-kt_2}}\big) & \text{for Theorem 1.6,} \\ \text{Poisson}\big(2\delta \frac{\log c}{c^{b_0}-1}\big) & \text{for Theorem 1.7,} \end{cases} \tag{3.31}$$

and $t_1 := (j-1)\delta, t_2 := (j+1)\delta$. The claim then follows from (3.29) and (3.30).

3.3 Conclusion

It is clear from the form of the finite dimensional distributions that in all Theorems 1.1, 1.2, 1.6, 1.7 the limiting process Z is a pure birth process that does not explode in finite time. Its rate at the point $(t,j) \in [0,\infty) \times \mathbb{N}$ is

$$\lambda_{t,j} = \lim_{h \to 0^+} \frac{1}{h} \mathbf{P}(Z(t+h) = j+1 | Z(t) = j)$$

and is found as stated in the statement of each theorem.

4 Deterministic and diffusion limits. Proof of Theorems 1.3, 1.4, 1.8, 1.9

These theorems are proved with the use of Theorem 7.1 in Chapter 8 of [5], which is concerned with convergence of time-homogeneous Markov processes to diffusions. Since our basic Markov chain, $(A_n^{(m)})_{n \in \mathbb{N}}$, is not time-homogeneous, we do the standard trick of considering the chain $\{(A_n^{(m)}, n)\}_{n \in \mathbb{N}}$ which is time-homogeneous.

4.1 Proof of Theorems 1.3, 1.8

For each $m \in \mathbb{N}^+$, consider the discrete time-homogeneous Markov chain

$$Z_n^{(m)} = \Big(\frac{A_n^{(m)}}{m}, \frac{n}{m}\Big).$$

From any given state (x_1, x_2) of $Z_n^{(m)}$, the chain moves to either of $(x_1 + k/m, x_2 + m^{-1}), (x_1, x_2 + m^{-1})$ with corresponding probabilities $p(x_1, x_2, m), 1 - p(x_1, x_2, m)$, where

$$p(x_1, x_2, m) := \begin{cases} \frac{mx_1}{A_0^{(m)} + B_0^{(m)} + kmx_2} & \text{in the case of Theorem 1.3,} \\ \frac{1 - q_m^{mx_1}}{1 - q_m^{A_0^{(m)} + B_0^{(m)} + kmx_2}} & \text{in the case of Theorem 1.8.} \end{cases}$$

This is true because when the chain is at the point (x_1, x_2), then the time n is $n = mx_2$ and $A_n^{(m)} + B_n^{(m)} = A_0^{(m)} + B_0^{(m)} + kn$. Define also

$$p(x_1, x_2) := \lim_{m \to \infty} p(x_1, x_2, m) = \begin{cases} \frac{x_1}{a+b+kx_2} & \text{in the case of Theorem 1.3,} \\ \frac{1-c^{x_1}}{1-c^{a+b+kx_2}} & \text{in the case of Theorem 1.8.} \end{cases} \tag{4.1}$$

We compute the mean and the covariance matrix for the one step change of $Z^{(m)} = (Z^{(m),1}, Z^{(m),2})$ conditioned on its current position.

$$\mathbf{E}\left[Z^{(m),1}_{n+1} - Z^{(m),1}_{n} | Z^{(m)}_n = (x_1, x_2)\right] = \frac{k}{m} p(x_1, x_2, m), \tag{4.2}$$

$$\mathbf{E}\left[Z^{(m),2}_{n+1} - Z^{(m),2}_{n} | Z^{(m)}_n = (x_1, x_2)\right] = \frac{1}{m}, \tag{4.3}$$

$$\mathbf{E}\left[(Z^{(m),1}_{n+1} - Z^{(m),1}_{n})^2 | Z^{(m)}_n = (x_1, x_2)\right] = \frac{k^2}{m^2} p(x_1, x_2, m), \tag{4.4}$$

$$\mathbf{E}\left[(Z^{(m),1}_{n+1} - Z^{(m),1}_{n})(Z^{(m),2}_{n+1} - Z^{(m),2}_{n}) | Z^{(m)}_n = (x_1, x_2)\right] = \frac{k}{m^2} p(x_1, x_2, m), \tag{4.5}$$

$$\mathbf{E}\left[(Z^{(m),2}_{n+1} - Z^{(m),2}_{n})^2 | Z^{(m)}_n = (x_1, x_2)\right] = \frac{1}{m^2}. \tag{4.6}$$

For each $m \in \mathbb{N}^+$, we consider the process $\Lambda^{(m)}_t := Z^{(m)}_{[mt]}, t \geq 0$. According to Theorem 7.1 in Chapter 8 of [5], the sequence $(\Lambda^{(m)})_{m\geq 1}$ converges weakly to the solution, $(S_t)_{t\geq 0}$, of the differential equation

$$\begin{aligned} dS_t &= b(S_t)dt, \\ S_0 &= \begin{pmatrix} a \\ 0 \end{pmatrix}, \end{aligned} \tag{4.7}$$

where

$$S_t = \begin{pmatrix} S^{(1)}_t \\ S^{(2)}_t \end{pmatrix}, \; b\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kp(x, y) \\ 1 \end{pmatrix}. \tag{4.8}$$

To apply the theorem, we need to check that the martingale problem MP$(b, \mathbb{O})$ has a unique solution. Here, $\mathbb{O}$ is the 2×2 zero matrix. See [5], §5.4, for details on the martingale problem. The problem indeed has unique solution because the differential equation (4.7) has a unique solution, and by well known results, this implies the claim for the martingale problem.

It follows that the process $(A^{(m)}_{[mt]})_{t\geq 0}$ converges, as $m \to \infty$, to the solution of the differential equation

$$X_0 = a, \tag{4.9}$$

$$dX_t = kp(X_t, t)dt. \tag{4.10}$$

For both theorems, 1.3 and 1.8, this ordinary differential equation is separable and its unique solution is the one stated.

4.2 Proof of Theorems 1.4, 1.9

Proof of Theorem 1.4. Call $\lambda := a/(a+b)$. For each $m \in \mathbb{N}^+$, consider the discrete time-homogeneous Markov chain

$$Z^{(m)}_n = \left(\sqrt{m}\Big(\frac{A^{(m)}_n}{m} - a - \lambda k \frac{n}{m}\Big), \frac{n}{m}\right), n \in \mathbb{N}.$$

From any given state (x_1, x_2) of $Z^{(m)}_n$, the chain moves to either of $(x_1 - km^{-1/2}\lambda, x_2 + m^{-1}), (x_1 + km^{-1/2}(1 - \lambda), x_2 + m^{-1})$ with corresponding probabilities

$$\Pi^{(m)}\left[(x_1, x_2), \left(x_1 - \frac{k}{\sqrt{m}}\lambda, x_2 + \frac{1}{m}\right)\right] = \frac{B^{(m)}_n}{A^{(m)}_n + B^{(m)}_n}, \tag{4.11}$$

$$\Pi^{(m)}\left[(x_1, x_2), \left(x_1 + \frac{k}{\sqrt{m}}(1 - \lambda), x_2 + \frac{1}{m}\right)\right] = \frac{A^{(m)}_n}{A^{(m)}_n + B^{(m)}_n}, \tag{4.12}$$

with

$$A^{(m)}_n = ma + \lambda k m x_2 + x_1\sqrt{m}, \tag{4.13}$$

$$B^{(m)}_n = A^{(m)}_0 + B^{(m)}_0 + kmx_2 - A^{(m)}_n. \tag{4.14}$$

We used the fact that when the chain is at the point (x_1, x_2), then the time n is $n = mx_2$.

We compute the mean and the covariance matrix for the one step change of $Z^{(m)} = (Z^{(m),1}, Z^{(m),2})$ conditioned on its current position.

$$\mathbf{E}\left[Z_{n+1}^{(m),1} - Z_n^{(m),1} | Z_n^{(m)} = (x_1, x_2)\right] = \frac{k}{\sqrt{m}} \frac{(1-\lambda)A_n^{(m)} - \lambda B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}}$$

$$\sim \frac{1}{m} \frac{k\{x_1 - (\theta_1 + \theta_2)\lambda\}}{a + b + kx_2}, \tag{4.15}$$

$$\mathbf{E}\left[Z_{n+1}^{(m),2} - Z_n^{(m),2} | Z_n^{(m)} = (x_1, x_2)\right] = \frac{1}{m}, \tag{4.16}$$

$$\mathbf{E}\left[(Z_{n+1}^{(m),1} - Z_n^{(m),1})^2 | Z_n^{(m)} = (x_1, x_2)\right] = \frac{k^2}{m} \left(\lambda^2 \frac{B_n^{(m)}}{A_n^{(m)} + B_n^{(m)}} + (1-\lambda)^2 \frac{A_n^{(m)}}{A_n^{(m)} + B_n^{(m)}}\right)$$

$$\sim \frac{1}{m} \frac{k^2 ab}{(a+b)^2}, \tag{4.17}$$

$$\mathbf{E}\left[(Z_{n+1}^{(m),1} - Z_n^{(m),1})(Z_{n+1}^{(m),2} - Z_n^{(m),2}) | Z_n^{(m)} = (x_1, x_2)\right] \sim \frac{1}{m^2} \frac{kx_1}{a + b + kx_2}, \tag{4.18}$$

$$\mathbf{E}\left[(Z_{n+1}^{(m),2} - Z_n^{(m),2})^2 | Z_n^{(m)} = (x_1, x_2)\right] = \frac{1}{m^2}. \tag{4.19}$$

Then, for each $m \in \mathbb{N}^+$, we consider the process $\Lambda_t^{(m)} := Z_{[mt]}^{(m)}, t \geq 0$. According to Theorem 7.1 in Chapter 8 of [5], the sequence $(\Lambda^{(m)})_{m \geq 1}$ converges in distribution to the solution, $(S_t)_{t \geq 0}$, of the stochastic differential equation

$$dS_t = b(S_t)dt + \sigma(S_t)dB_t, \tag{4.20}$$

$$S_0 = \begin{pmatrix} \theta_1 \\ 0 \end{pmatrix}, \tag{4.21}$$

where

$$S_t = \begin{pmatrix} S_t^{(1)} \\ S_t^{(2)} \end{pmatrix}, \qquad B_t = \begin{pmatrix} B_t^{(1)} \\ B_t^{(2)} \end{pmatrix},$$

$$b\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{k\{x - (\theta_1 + \theta_2)\lambda\}}{a+b+ky} \\ 1 \end{pmatrix}, \quad \sigma\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k\frac{\sqrt{ab}}{a+b} & 0 \\ 0 & 0 \end{pmatrix}.$$

B is a two dimensional standard Brownian motion. Again, to apply that theorem, we need to check that the martingale problem MP(b, σ) has a unique solution. This follows from the existence and uniqueness of strong solution for the above stochastic differential equation as the coefficients b, σ are Lipschitz and grow at most linearly at infinity.

Thus, the process $(Z_{[mt]}^{(m),1})_{t \geq 0}$ converges in distribution, as $m \to \infty$, to the solution of

$$Y_0 = \theta_1, \tag{4.22}$$

$$dY_t = \frac{k\{Y_t - (\theta_1 + \theta_2)\lambda\}}{a + b + kt} dt + k\frac{\sqrt{ab}}{a+b} dB_t^{(1)}. \tag{4.23}$$

The same is true for $(C_t^{(m)})_{t \geq 0}$ because $\sup_{t \geq 0} |C_t^{(m)} - Z_{[mt]}^{(m),1}| \leq k/\sqrt{m}$. To solve the last SDE, we set $U_t := \{Y_t - (\theta_1 + \theta_2)\lambda\}/(a + b + kt)$. Ito's lemma gives that

$$dU_t = k\frac{\sqrt{ab}}{(a+b)} \frac{1}{a + b + kt} dB_t^{(1)},$$

and since $U_0 = (b\theta_1 - a\theta_2)/(a+b)^2$, we get

$$U_t = \frac{b\theta_1 - a\theta_2}{(a+b)^2} + k\frac{\sqrt{ab}}{a+b} \int_0^t \frac{1}{a + b + ks} dB_s^{(1)}.$$

This gives (1.8). ■

Proof of Theorem 1.9. The proof is analogous to that of Theorem 1.4, only the algebra is a little more involved. For each $m \in \mathbb{N}^+$, consider the discrete time-homogeneous Markov chain

$$Z_n^{(m)} = \Big(\sqrt{m}\Big(\frac{A_n^{(m)}}{m} - X_{n/m}\Big), \frac{n}{m}\Big), n \in \mathbb{N}.$$

From any given state (x_1, x_2) of $Z_n^{(m)}$, the chain moves to either of

$$(x_1, x_2) + (km^{-1/2} + \sqrt{m}(X_{x_2} - X_{x_2+m^{-1}}), m^{-1}), \tag{4.24}$$
$$(x_1, x_2) + (\sqrt{m}(X_{x_2} - X_{x_2+m^{-1}}), m^{-1}) \tag{4.25}$$

with corresponding probabilities $p(x_1, x_2, m), 1 - p(x_1, x_2, m)$, where

$$p(x_1, x_2, m) = \frac{[A_n^{(m)}]_{q_m}}{[A_0^{(m)} + B_0^{(m)} + kmx_2]_{q_m}} \tag{4.26}$$

and

$$A_n^{(m)} = mX_{x_2} + x_1\sqrt{m}, \tag{4.27}$$
$$B_n^{(m)} = A_0^{(m)} + B_0^{(m)} + kmx_2 - A_n^{(m)}. \tag{4.28}$$

We used the fact that when the chain is at the point (x_1, x_2), then the time n is $n = mx_2$. For convenience, let $\Delta X_{x_2} = X_{x_2+m^{-1}} - X_{x_2}$.

We compute the mean and the covariance matrix for the one step change of $Z^{(m)} = (Z^{(m),1}, Z^{(m),2})$ conditioned on its current position. Of the following relations, the first four are immediate, the fifth follows from part (a) of the claim that follows and the fact that $Z_{n+1}^{(m),2} - Z_n^{(m),2} = 1/m$.

$$\mathbf{E}\left[Z_{n+1}^{(m),1} - Z_n^{(m),1} | Z_n^{(m)} = (x_1, x_2)\right] = \frac{k}{\sqrt{m}} p(x_1, x_2, m) - \sqrt{m}\Delta X_{x_2} \tag{4.29}$$
$$\mathbf{E}\left[(Z_{n+1}^{(m),1} - Z_n^{(m),1})^2 | Z_n^{(m)} = (x_1, x_2)\right] = \left(\frac{k^2}{m} - 2k\Delta X_{x_2}\right) p(x_1, x_2, m) + m(\Delta X_{x_2})^2 \tag{4.30}$$
$$\mathbf{E}\left[Z_{n+1}^{(m),2} - Z_n^{(m),2} | Z_n^{(m)} = (x_1, x_2)\right] = \frac{1}{m}, \tag{4.31}$$
$$\mathbf{E}\left[(Z_{n+1}^{(m),2} - Z_n^{(m),2})^2 | Z_n^{(m)} = (x_1, x_2)\right] = \frac{1}{m^2}, \tag{4.32}$$
$$\mathbf{E}\left[(Z_{n+1}^{(m),1} - Z_n^{(m),1})(Z_{n+1}^{(m),2} - Z_n^{(m),2}) | Z_n^{(m)} = (x_1, x_2)\right] = O(m^{-2}) \tag{4.33}$$

We examine now the asymptotics of the first two expectations.

Claim:

(a)
$$\mathbf{E}\left[Z_{n+1}^{(m),1} - Z_n^{(m),1} | Z_n^{(m)} = (x_1, x_2)\right] \sim \frac{1}{m}\frac{k\log c}{c^{a+b+kx_2} - 1}\left(c^{X_{x_2}}x_1 - \frac{(c^{X_{x_2}} - 1)c^{a+b+kx_2}}{c^{a+b+kx_2} - 1}(\theta_1 + \theta_2)\right) + O(\frac{1}{m^{3/2}})$$

(b)
$$\mathbf{E}\left[\{Z_{n+1}^{(m),1} - Z_n^{(m),1}\}^2 | Z_n^{(m)} = (x_1, x_2)\right] \sim \frac{1}{m}k^2 g(x_2)\{1 - g(x_2)\} + O(\frac{1}{m^{3/2}})$$

where $g(x_2) := \lim_{m\to\infty} p(x_1, x_2, m) = \frac{c^{X_{x_2}} - 1}{c^{a+b+kx_2} - 1}$.

Proof of the claim. We examine the asymptotics of $p(x_1, x_2, m)$ and ΔX_{x_2}. We have

$$p(x_1, x_2, m) = \frac{c^{X_{x_2} + \frac{1}{\sqrt{m}x_1}} - 1}{c^{\frac{A_0^{(m)} + B_0^{(m)}}{m} + kx_2} - 1} = \frac{c^{X_{x_2} + \frac{1}{\sqrt{m}x_1}} - 1}{c^{a+b+kx_2 + \frac{\theta_1 + \theta_2}{\sqrt{m}} + O(\frac{1}{m})} - 1} \tag{4.34}$$
$$= g(x_2) + \frac{\log c}{c^{a+b+kx_2} - 1}\left(c^{X_{x_2}}x_1 - \frac{(c^{X_{x_2}} - 1)c^{a+b+kx_2}}{c^{a+b+kx_2} - 1}(\theta_1 + \theta_2)\right)\frac{1}{\sqrt{m}} + O(\frac{1}{m}). \tag{4.35}$$

The second equality follows from a Taylor development. Also

$$\Delta X_{x_2} = X'_{x_2}\frac{1}{m} + O(m^{-2}) = kg(x_2)\frac{1}{m} + O(m^{-2}). \tag{4.36}$$

For X' we used the differential equation, (1.22), that X satisfies instead of the explicit expression for it. Substituting these expressions in (4.29), (4.30), we get the claim.

Relation (1.23) implies that $c^{X_{x_2}} = (c^{a+b}-1)/\{c^b - 1 + c^{-kx_2}(1-c^{-a})\}$, and this gives that the parenthesis following $\frac{1}{m}$ in equation (a) of the claim above equals

$$\frac{(c^{a+b}-1)x_1 - c^b(c^a-1)(\theta_1+\theta_2)}{c^b - 1 + c^{-kx_2}(1-c^{-a})} \tag{4.37}$$

and also that

$$g(x_2)\{1-g(x_2)\} = \frac{(c^a-1)(c^b-1)c^{a+kx_2}}{(c^{a+b+kx_2} - c^{a+kx_2} + c^a - 1)^2}. \tag{4.38}$$

It follows as before that the process $(Z^{(m)}_{[mt]})_{t\geq 0}$ converges, as $m\to\infty$, to the solution of the stochastic differential equation

$$dS_t = b(S_t)dt + \sigma(S_t)dB_t, \tag{4.39}$$

$$S_0 = \begin{pmatrix} \theta_1 \\ 0 \end{pmatrix}, \tag{4.40}$$

where

$$S_t = \begin{pmatrix} S_t^{(1)} \\ S_t^{(2)} \end{pmatrix},\ B_t = \begin{pmatrix} B_t^{(1)} \\ B_t^{(2)} \end{pmatrix},$$

$$b\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{k\log c}{c^{a+b+ky}-1}\left\{\frac{(c^{a+b}-1)x - c^b(c^a-1)(\theta_1+\theta_2)}{c^b-1+c^{-ky}(1-c^{-a})}\right\} \\ 1 \end{pmatrix},$$

$$\sigma\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k\sqrt{(c^a-1)(c^b-1)}\frac{c^{(a+ky)/2}}{c^{a+b+ky}-c^{a+ky}+c^a-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

B is a two dimensional standard Brownian motion. Again, the martingale problem $\mathrm{MP}(b,\sigma)$ has a unique solution due to the form of the functions b,σ. And with analogous arguments as in Theorem 1.4, we get that the process $(\hat{C}_t^{(m)})_{t\geq 0}$ converges to the unique solution of the stochastic differential equation (1.24). To solve that, we remark that a solution of a stochastic differential equation of the form

$$dY_t = (\alpha(t)Y_t + \beta(t))dt + \gamma(t)dW_t \tag{4.41}$$

with $\alpha,\beta,\gamma : [0,\infty)\to\mathbb{R}$ continuous functions is given by

$$Y_t = e^{\int_0^t \alpha(s)\,ds}\left(Y_0 + \int_0^t \beta(s)e^{-\int_0^s \alpha(r)\,dr}\,ds + \int_0^t \gamma(s)e^{-\int_0^s \alpha(r)\,dr}\,dW_s\right). \tag{4.42}$$

[To discover the formula, we apply Itó's rule to $Y_t\exp\{-\int_0^t \alpha(s)\,ds\}$ and use (4.41).] Applying this formula for the values of α,β,γ dictated by (1.24) we arrive at (1.25). ■

5 Proofs for the q-Pólya urn with many colors

Proof of Theorem 1.10. First, the equality of the expressions in (1.27), (1.28) follows from the definition of the q-multinomial coefficient.

We will prove (1.27) by induction on l. When $l=2$, (1.27) holds because of (1.14). In that relation, we have $x_1 = x, x_2 = n-x$. Assuming that (1.27) holds for $l\geq 2$ we will prove the case $l+1$. The probability $\mathbf{P}(X_{n,2}=x_2,\dots,X_{n,l+1}=x_{l+1})$ equals

$$\mathbf{P}(X_{n,3}=x_3,\dots,X_{n,l+1}=x_{l+1})\,\mathbf{P}(X_{n,2}=x_2 \mid X_{n,3}=x_3,\dots,X_{n,l+1}=x_{l+1}) \tag{5.1}$$

$$= q^{\sum_{i=3}^{l+1} x_i \sum_{j=1}^{i-1}(a_j+kx_j)} \frac{\left[\begin{smallmatrix} -\frac{a_1+a_2}{k} \\ x_1+x_2 \end{smallmatrix}\right]_{q^{-k}} \prod_{i=3}^{l+1} \left[\begin{smallmatrix} -\frac{a_i}{k} \\ x_i \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} -\frac{a_1+\dots a_{l+1}}{k} \\ n \end{smallmatrix}\right]_{q^{-k}}} \cdot q^{x_2(a_1+kx_1)} \frac{\left[\begin{smallmatrix} -\frac{a_1}{k} \\ x_1 \end{smallmatrix}\right]_{q^{-k}} \left[\begin{smallmatrix} -\frac{a_2}{k} \\ x_2 \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} -\frac{a_1+a_2}{k} \\ x_1+x_2 \end{smallmatrix}\right]_{q^{-k}}} \tag{5.2}$$

$$= q^{\sum_{i=2}^{l+1} x_i \sum_{j=1}^{i-1}(a_j+kx_j)} \frac{\prod_{i=1}^{l+1} \left[\begin{smallmatrix} -\frac{a_i}{k} \\ x_i \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} -\frac{a_1+\dots a_{l+1}}{k} \\ n \end{smallmatrix}\right]_{q^{-k}}}. \tag{5.3}$$

This finishes the induction provided that we can justify these two equalities. The second is obvious, so we turn to the first. The first probability in (5.1) is specified by the inductive hypothesis. That is, given the description of the experiment, in computing this probability it is as if we merge colors 1 and 2 into one color which is placed in the line before the remaining $l-1$ colors. This color has initially a_1+a_2 balls and we require that in the first n drawings we choose it x_1+x_2 times. The second probability in (5.1) is specified by the $l=2$ case of (1.27), which we know. More specifically, since the number of drawings from colors $3,4,\dots,l+1$ is given, it is as if we have an urn with just two colors $1,2$ that have initially a_1 and a_2 balls respectively. We do x_1+x_2 drawings with the usual rules for a q-Pólya urn, placing in a line all balls of color 1 before all balls of color 2, and we want to pick x_1 times color 1 and x_2 times color 2. ■

Proof of Theorem 1.11. The components of $(X_{n,2},X_{n,3},\dots,X_{n,l})$ are increasing in n, and from Theorem 1.5 we have that each of them has finite limit (we treat all colors $2,\dots,l$ as one color). Thus the convergence of the vector with probability one to a random vector with values is $\mathbb{N}^{l-1}$ follows. In particular, we also have convergence in distribution, and it remains to compute the distribution of the limit. Let $x_1 := n-(x_2+\dots+x_l)$. Then the probability in (1.27) equals

$$\mathbf{P}(X_{n,2}=x_2,\dots,X_{n,l}=x_l) = q^{-\sum_{1\le i<j\le l} a_j x_i} \frac{\prod_{i=1}^{l} \left[\begin{smallmatrix} \frac{a_i}{k}+x_i-1 \\ x_i \end{smallmatrix}\right]_{q^{-k}}}{\left[\begin{smallmatrix} \frac{\sum_{i=1}^{l} a_i}{k}+n-1 \\ n \end{smallmatrix}\right]_{q^{-k}}} \tag{5.4}$$

$$= q^{\sum_{1\le j<i\le l} x_i a_j} \frac{\prod_{i=1}^{l} \left[\begin{smallmatrix} \frac{a_i}{k}+x_i-1 \\ x_i \end{smallmatrix}\right]_{q^{k}}}{\left[\begin{smallmatrix} n+\frac{\sum_{i=1}^{l} a_i}{k}-1 \\ n \end{smallmatrix}\right]_{q^{k}}} \tag{5.5}$$

$$= q^{\sum_{i=2}^{l}\left(x_i \sum_{j=1}^{i-1} a_j\right)} \left\{ \prod_{i=2}^{l} \begin{bmatrix} \frac{a_i}{k}+x_i-1 \\ x_i \end{bmatrix}_{q^k} \right\} \frac{\left[\begin{smallmatrix} x_1+\frac{a_1}{k}-1 \\ x_1 \end{smallmatrix}\right]_{q^{k}}}{\left[\begin{smallmatrix} n+\frac{\sum_{i=1}^{l} a_i}{k}-1 \\ n \end{smallmatrix}\right]_{q^{k}}}. \tag{5.6}$$

In the first equality, we used (2.2) while in the second we used (2.3). When we take $n\to\infty$ in (5.6), the only terms involving n are those of the last fraction, and (2.11) determines their limit. Thus, the limit of (5.6) is found to be the function $f(x_2,\dots,x_l)$ in the statement of the theorem. ■

Proof of Theorem 1.12. For each $m\in\mathbb{N}^+$, we consider the discrete time-homogeneous Markov chain

$$Z_n^{(m)} := \left(\frac{n}{m}, \frac{A_{n,2}^{(m)}}{m}, \frac{A_{n,3}^{(m)}}{m}, \dots, \frac{A_{n,l}^{(m)}}{m}\right), n\in\mathbb{N}.$$

From any given state $(t,x) := (t,x_2,x_3,\dots,x_l)$ that $Z^{(m)}$ finds itself it moves to one of

$$\left(t+\frac{1}{m}, x_2,\dots,x_i+\frac{1}{m},\dots,x_l\right), i=2,\dots,l,$$
$$\left(t+\frac{1}{m}, x_2,\dots,x_i,\dots,x_l\right)$$

with corresponding probabilities

$$p_i(x_2,\dots,x_l,t,m) = q^{ms_{i-1}(t)}\frac{[mx_i]_q}{[ms_l(t)]_q}, i=2,\dots,l, \tag{5.7}$$
$$p_1(x_2,\dots,x_l,t,m) = \frac{[mx_1(t)]_q}{[ms_l(t)]_q}, \tag{5.8}$$

where $s_i(t) = x_1(t)+\sum_{1<j\le i} x_j$ for $i\in\{1,2,\dots,l\}$ and $x_1(t) := m^{-1}\sum_{j=1}^{l} A_{0,j}^{(m)} + kt - \sum_{2\le j\le l} x_i$. These follow from (1.26) once we count the number of balls of each color present at the state (t,x). To do this, we note that $Z_n^{(m)} = (t,x)$ implies that $n=mt$ drawings have taken place so far, the total number of balls is $A_{0,1}^{(m)}+\cdots+A_{0,l}^{(m)}+kmt$, and the number of balls of color i, for $2\le i\le l$, is mx_i. Thus, the number of balls of color 1 is $A_{0,1}^{(m)}+\cdots+A_{0,l}^{(m)}+kmt-m\sum_{2\le j\le l} x_i = mx_1(t)$. The required relations follow.

Let $x_1 := \lim_{m\to\infty} x_1(t) = \sigma_l + kt - \sum_{2\le j\le l} x_i$ and $s_i := \lim_{m\to\infty} s_i(t) = \sum_{1\le j\le i} x_i$ for all $i\in\{1,2,\dots,l\}$. Then, since $q=c^{1/m}$, for fixed $(t,x_2,\dots,x_l)\in[0,\infty)^l$ with $(x_2,\dots,x_l)\neq 0$, we have

$$\lim_{m\to\infty} p_i(x_2,\dots,x_l,t,m) = c^{s_{i-1}}\frac{[x_i]_c}{[s_l]_c} \tag{5.9}$$

for all $i=2,\dots,l$. We also note the following.

$$Z_{n+1,1}^{(m)} - Z_{n,1}^{(m)} = \frac{1}{m}, \tag{5.10}$$
$$\mathbf{E}\left[Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)} | Z_n^{(m)} = (t,x_2,\dots,x_l)\right] = \frac{k}{m} p_i(x_2,\dots,x_l,t,m), \tag{5.11}$$
$$\mathbf{E}\left[(Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)})^2 | Z_n^{(m)} = (t,x_2,\dots,x_l)\right] = \frac{k^2}{m^2} p_i(x_2,\dots,x_l,t,m), \tag{5.12}$$
$$\mathbf{E}\left[(Z_{n+1,i}^{(m)} - Z_{n,i}^{(m)})(Z_{n+1,j}^{(m)} - Z_{n,j}^{(m)}) | Z_n^{(m)} = (t,x_2,\dots,x_l)\right] = 0 \tag{5.13}$$

for $i,j=2,3,\dots,l$ with $i\neq j$.

Therefore, with similar arguments as in the proof of Theorem 1.3, as $m\to+\infty$, $(Z_{[mt]}^{(m)})_{t\ge 0}$ converges in distribution to Y, the solution of the ordinary differential equation

$$\begin{aligned} dY_t &= b(Y_t)dt, \\ Y_0 &= (0,a_2,\dots,a_l), \end{aligned} \tag{5.14}$$

where $b(t,x_2,\dots,x_l) = \big(1, b^{(2)}(t,x), b^{(3)}(t,x),\dots,b^{(l)}(t,x)\big)$ with

$$b^{(i)}(t,x) = kc^{s_{i-1}}\frac{[x_i]_c}{[s_l]_c}$$

for $i=2,3,\dots,l$. Note that $s_l = \sigma_l + kt$ does not depend on x.

Since $A_{[mt],1}^{(m)} + A_{[mt],2}^{(m)} + \cdots + A_{[mt],l}^{(m)} = kmt + A_{0,1}^{(m)} + A_{0,2}^{(m)} + \cdots + A_{0,l}^{(m)}$, we get that the process $(A_{[mt],1}^{(m)}/m, A_{[mt],2}^{(m)}/m + \cdots + A_{[mt],l}^{(m)}/m)_{t\ge 0}$ converges in distribution to a process $(X_{t,1},X_{t,2},\dots,X_{t,l})_{t\ge 0}$ so that $X_{t,1}+\cdots+X_{t,l} = a_1+a_2+\cdots+a_l+kt$, while the $X_{t,i}, i=2,\dots,l$, satisfy the system

$$X'_{t,i} = kc^{\sigma_l+kt-\sum_{j=i}^{l} X_{t,i}}\frac{1-c^{X_{t,i}}}{1-c^{\sigma_l+kt}} \qquad \text{for all } t>0, \tag{5.15}$$
$$X_{0,i} = a_i, \tag{5.16}$$

with $i=2,3,\dots,l$. Letting $Z_{r,i} = c^{X_{\frac{1}{k\log c}\log r, i}}$ for all $r\in(0,1]$ and $i\in\{1,2,\dots,l\}$, we have for the $Z_{r,i}, i\in\{2,3,\dots,l\}$ the system

$$\frac{Z'_{r,i}}{1-Z_{r,i}} = \frac{\sigma_l}{1-\sigma_l r}\frac{1}{\prod_{i<j\le l} Z_{r,j}}, \tag{5.17}$$
$$Z_{1,i} = c^{a_i}. \tag{5.18}$$

In the case $i = l$, the empty product equals 1. It is now easy to prove by induction (starting from $i = l$ and going down to $i = 2$) that

$$Z_{r,i} = \frac{c^{\sigma_l - \sigma_{i-1}}(1 - c^{\sigma_l} r) - c^{\sigma_l}(1 - r)}{c^{\sigma_l - \sigma_i}(1 - c^{\sigma_l} r) - c^{\sigma_l}(1 - r)} \tag{5.19}$$

for all $r \in (0, 1]$. Since $Z_{r,1} Z_{r,2} \cdots Z_{r,l} = c^{\sigma_l} r$, we can check that (5.19) holds for $i = 1$ too. The fraction in (5.19) equals

$$c^{a_i} \frac{(1 - c^{\sigma_l} r) - c^{\sigma_{i-1}}(1 - r)}{(1 - c^{\sigma_l} r) - c^{\sigma_i}(1 - r)}. \tag{5.20}$$

Recalling that $X_{t,i} = (\log c)^{-1} \log Z_{c^{kt}}$, we get (1.30) for all $i \in \{1, 2, \ldots, l\}$. ■

Proof of Theorem 1.13. This is proved in the same way as Theorem 1.12. We keep the same notation as there. The only difference now is that $\lim_{m \to \infty} p_i(t, x_2, \ldots, x_l, m) = x_i / s_l$. As a consequence, the system of ordinary differential equations for the limit process $Y_t := (t, X_{t,2}, \ldots, X_{t,l})$ is (5.14) but with

$$b^{(i)}(t, x) = \frac{k x_i}{s_l}.$$

Recall that $s_l = \sigma_l + kt$. Thus, for $i = 2, 3, \ldots, l$, the process $X_{t,i}$ satisfies $X'_{t,i} = k X_{t,i} / (\sigma_l + kt)$, $X_{0,i} = a_i$, which give immediately the last $l - 1$ coordinates of (1.31). The formula for the first coordinate follows from $X_{t,1} + X_{t,2} + \cdots + X_{t,l} = kt + \sigma_l$. ■

References

[1] Bagchi, Arunabha, and Asim K. Pal. "Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures." SIAM Journal on Algebraic Discrete Methods 6, no. 3 (1985): 394-405.

[2] Bai, Zhi-Dong, Feifang Hu, and Li-Xin Zhang. "Gaussian approximation theorems for urn models and their applications." The Annals of Applied Probability 12, no. 4 (2002): 1149-1173.

[3] Charalambides, Ch A. "A q-Pólya urn model and the q-Pólya and inverse q-Pólya distributions." Journal of Statistical Planning and Inference 142, no. 1 (2012): 276-288.

[4] Charalambides, Charalambos A. Discrete q-distributions. John Wiley & Sons, 2016.

[5] Durrett, Richard. Stochastic calculus: a practical introduction. CRC press, 1996.

[6] Ethier, Stewart N., and Thomas G. Kurtz. Markov processes: characterization and convergence. Vol. 282. John Wiley & Sons, 2009.

[7] G. Gasper and M. Rahman. Basic Hypergeometric Series. Cambridge University Press, Cambridge, 1990

[8] Gouet, Raul. "Martingale functional central limit theorems for a generalized Pólya urn." The Annals of Probability (1993): 1624-1639.

[9] Heyde, C. C. "On central limit and iterated logarithm supplements to the martingale convergence theorem." Journal of Applied Probability 14, no. 4 (1977): 758-775.

[10] Janson, Svante. "Functional limit theorems for multitype branching processes and generalized Pólya urns." Stochastic Processes and their Applications 110, no. 2 (2004): 177-245.

[11] Johnson, Norman Lloyd, and Samuel Kotz. "Urn models and their application; an approach to modern discrete probability theory." (1977).

[12] Kac, Victor, and Pokman Cheung. Quantum calculus. Springer Science & Business Media, 2001.

[13] Kupershmidt, Boris A. "q-probability: I. Basic discrete distributions." Journal of Nonlinear Mathematical Physics 7, no. 1 (2000): 73-93.

[14] Mahmoud, Hosam. Pólya urn models. Chapman and Hall/CRC, 2008.

Chaos, Solitons and Fractals 122 (2019) 153–162

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

On the pricing of exotic options: A new closed-form valuation approach

Stelios Bekiros[a,b,c,*], Dimitra Kouloumpou[b]

[a] European University Institute, Department of Economics, Via delle Fontanelle, 18, I-50014, Florence, Italy
[b] Athens University of Economics and Business, School of Business Administration, Troias, 2, GR10434, Athens, Greece
[c] Rimini Centre for Economic Analysis (RCEA), LH3079, Wilfrid Laurier University, 75 University Ave W., ON, N2L3C5, Waterloo, Canada

A R T I C L E I N F O

Article history:
Received 28 February 2019
Accepted 13 March 2019

JEL classification:
G10
G12
G13
C63

MSC:
60-08
60J75
47A68
42A85
91B28
91G20
91B25
60J28
47H25
46F12

Keywords:
Brownian motion
Option pricing
Crossing probabilities
Diffusion approximations

A B S T R A C T

We provide a novel method to estimate in a closed-form solution the option prices of various exotic options, using techniques based on Laplace–Beltrami operator for estimating diffusion boundary times. We estimate exit times and their expectations, the hitting probabilities, boundary local times until the first hitting and other probabilistic quantities and moment generating functions related to local hitting times. Our findings maybe of paramount importance for traders, investors, speculators and more broadly speaking for financial institutions.

1. Introduction

The payoff of a simple European or American-style call or put option while it depends heavily upon the value of the underlying asset, yet not particularly so on the path taken. In general, options derivatives products are determined by many features and primarily upon the underlying assets, as commonly reported in the relevant literature. However, a plethora of exotic options including Binary options, Cash-or-Nothing, Asset-or-Nothing, Barrier, Double Barrier options amongst other, all depend strongly on the path of the asset as well and on whether price barriers are hit or not. These barriers eventually control the option valuation. Once either of these barriers is breached, the status of the option is immediately determined, namely either the option comes into existence in case the barrier - as called - is in- or knock-in barrier, or ceases to exist if the barrier is out- or knock-out barrier. Other double barrier options of many types also exist (see [1]). In this work, we utilize double barrier options as proper proxies for many categories of exotic derivatives. To the best of our knowledge, we present for the first time new ways of estimating the expectation of the time when various exotic options seize to exist, their hitting probabilities, the exit times and their expectations, boundary local times until the first hitting and other probabilistic quantities related

* Corresponding author.
E-mail addresses: stelios.bekiros@eui.eu (S. Bekiros), dkouloumpou@aueb.gr (D. Kouloumpou).

https://doi.org/10.1016/j.chaos.2019.03.012

to the boundary local times. Importantly, we deliver closed-form solutions.

The diversity of exotic barrier options is indicative of their applicability in modern derivatives markets. Many different combinations of barrier options can be implemented in derivatives markets for equities, FX, commodities and bonds. We present thereafter the most important types: a double knock-out (DKO) or one touch knock-out Barrier option has both lower and upper knock-out barriers. Initially the holder of the option owns a call or a put. If at any time, either barrier is breached, the option seizes to exist (knocked-out). In some cases at knock-out, the holder may receive a rebate. For a double knock-in (DKI), or one touch knock-in barrier, if either barrier is breached, the holder of the barrier option is knocked-in, hence now owns the call or put. In cases where the option is never knocked-in, the holder may receive a rebate. An upper barrier knock-out (UKO) double barrier involves the upper barrier which when breached prior to the lower barrier, the option holder is knocked out, whilst in case the lower is breached prior to the upper or neither barrier is breached, the holder owns the option. Furthermore, in case of an upper barrier knock-out double option (UKO2) if the upper barrier is breached prior to the lower barrier, the option holder gets nothing and instead in case the lower barrier is breached prior to the upper, the holder receives an option. All in all, if neither barrier is breached, the holder gets nothing. For a lower barrier knock-out (LKO) if the lower barrier is hit prior to the upper barrier, the holder is knocked out and when the upper barrier is hit prior to the lower or neither barrier is breached, the holder owns an option. Additionally, for double lower barrier knock-outs (LKO2) if the lower barrier is breached prior to the upper one, the option holder gets nothing while when the upper barrier is breached prior to the lower, the holder receives the option. If neither barrier is breached, the holder is not compensated. An upper barrier knock-in double barrier (UKI) relates to the case whereby, if the upper barrier is reached prior to the lower barrier, the option holder receives a call or put, whilst in case neither barrier is hit, the holder owns an option. A lower barrier knock-in (LKI) performs similarly to the UKI, yet with a switch on the lower and upper barrier. A double touch knock-out option (DTKO) exists when the holder initially holds a call or put option. However, if both the upper and lower barriers are breached during the life of the option, the holder is knocked out. For a double touch knock-in option (DTKI), if both the upper and lower barriers are breached during the life of the option, the holder is knocked-in to a call or put option. With all of the aforementioned barrier variations, the specification of rebate is possible. These rebates (cash or asset amounts) can be specified if one or the other barrier is hit or if neither barrier is reached. Using these rebate features is a way of including digital / binary payoffs that depend on barrier levels. Finally, the type of monitoring conducted upon the barriers is a very important feature as well. Several possibilities exist, namely each barrier is continuously monitored for the life of the option or the barrier is partially monitored for specific windows during the option life. During these windows, the barriers are monitored continuously. Alternatively, each barrier is partially monitored for specific windows during the life of the option and the barriers are monitored at discrete dates or in another case the barrier is discretely monitored at specific dates.

Merton [2] was the first to derive a closed form solution for a down-and-out European call option. Other closed-form pricing formulae of exotic derivatives i.e., particularly for single-barrier options were published by Rubinstein and Reiner [3]. Rich [4] provided a mathematical framework for pricing the single-barrier options. A valuation method for double-barrier options based upon the probabilistic approach was discussed by Kunitomo and Ikeda [5]. The values of the double barrier options can also be obtained by solving the Black–Scholes partial differential equation with the corresponding boundary conditions using the method of separation of variables. Analytical solutions of one-touch double-barrier binary options, in which a fixed payoff is determined by whether it is touching the barrier, are derived by Hui [6]. Hui [7] extends regular single and double barrier options to time-dependent barrier options in which the barrier period covers a segment of time either at the beginning (front end) or the end (rear end) of the option life. This feature makes the time-dependent barrier options more flexible than the regular barrier options for an investor, having a particular view on an underlying asset in a certain period of time. The one-time barrier discontinuity in the time-dependent barrier options makes their pricing formulae different from the regular barrier option-pricing formulae. Roberts and Shortland [8] consider the problem of pricing derivative securities which involve a barrier clause. They present general techniques to calculate, or estimate accurately barrier option prices, using methods for estimating diffusion process with hitting times. Mario Dell'Era [9] discusses the efficiency of the spectral method for computing the value of double barrier options. Using this method, one may write the option price as a Fourier series with suitable coefficients. However, all of the aforementioned methods cannot be generalized vis-a-vis the valuation of many other exotic derivatives, and more importantly they do not tackle with all the analytical closed-form specificities occuring and remain unsolvable.

We contribute to the literature in significant ways. Specifically, to our knowledge this is the first study to estimate a closed-form solution for all the inherent features of a variety of barrier options and other proxies. Specifically, i) the expectation of the time that the option dies out is calculated, ii) the hitting probabilities i.e., the probabilities that the option hits first the upper or the lower barrier are accurately estimated, iii) certain probabilistic quantities related to the boundary local time until first hitting are introduced, iv) the exit times and their expectations are estimated as well as v) the boundary local times until the first hitting, which are of immense importance to the investors, alongside with the moment generating functions for all of the above. In this work, we consider as our proxy the double knock-out barrier with two barriers related to the strike price: an upper and a lower one. The upper barrier defines the level where the trigger price is above the strike price, while the lower barrier establishes a point at which the trigger price is below the strike. If the underlying does not break out of either barrier at any time during the option life, the option acts like a plain vanilla option and the holder would receive a specified payout. However, if one of the barriers has been broken through, the option dies out (gets knocked-out). Our novel concept is that we assume that the underlying asset follows a geometric Brownian motion on a 1−dimensional sphere $S^1 = \{x = (a\cos\varphi, a\sin\varphi) \in \mathbb{R}^2 |\ 0 \le \varphi < 2\pi\}$ (i.e. circle) with center at the origin and radius $a > \frac{H_2}{2\pi}$. In this case, the transformation introduces $\varphi_0, \varphi_1 \in [0, 2\pi)$ such that $H_1 = a\varphi_0$ and $H_2 = a\varphi_1$ i.e., two points on the circle, that denote the upper and lower barriers. The underlying asset then starts from a point $\phi \in D = (H_1, H_2)$. Via this method, we estimate the closed-form solution of the price of every barrier option or any exotic one thereby. Moreover, we calculate the expectation of the time the option dies out. The probabilities that the option hits first the upper or the lower barrier are calculated and we evaluate certain probabilistic quantities related to the boundary local time of the domain D until first hitting. We deliver valuable closed-form mathematical solutions of paramount importance for traders, investors, speculators and more broadly speaking for financial institutions. The paper is organized as follows: Section 2 presents preliminary definitions, propositions and proofs, whilst Section 3 describes the valuation method and a theorem related to that. Section 4 recalls some definitions and proofs, and presents new results on exit times, expectations, hitting probabilities and moments generating fnctions. Section 5 exposes the proofs for estimating boundary local times. Finally, Section 6 concludes with very interesting remarks regarding the extension of the results on spheres of higher dimensions and future applications of the presented methodology to the valuation of other exotic derivatives as well as to other mathematical problems in many topical fields.

2. Preliminaries

2.1. The $n-$sphere S^n

Definition 2.1. Let $n \in \mathbb{N}^* = \{1, 2, 3, \ldots\}$. The $n-$dimensional sphere S^n with center $(c_1, c_{2,\ldots,} c_{n+1})$ and radius $a>0$ is (defined to be) the set of all points $x = (x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1}$ satisfying $(x_1 - c_1)^2 + (x_2 - c_2)^2 + \ldots + (x_{n+1} - c_{n+1})^2 = a^2$. Thus,

$$S^n = \{(x_1, x_2, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} | (x_1 - c_1)^2 + (x_2 - c_2)^2 + \ldots + (x_{n+1} - c_{n+1})^2 = a^2\} \tag{2.1}$$

The points of the $n-$sphere with center at the origin and radius a for $n = 1$ may also be discribed in spherical coordinates in the following way

$$S^1 = \{x = (a\cos\varphi, a\sin\varphi) \in \mathbb{R}^2 |\ 0 < \varphi \leq 2\pi\} \tag{2.2}$$

The Laplace-Beltrami operator of a smooth function f on S^1 is

$$\Delta_1 f = \frac{1}{a^2}\frac{\partial^2 f}{\partial\varphi^2} \tag{2.3}$$

(see [10]).

2.2. Brownian motion on S^n

Definition 2.2. The Brownian motion on S^n is a diffusion (Markov) process $X_t, t \geq 0$, on S^n whose transition density is a function $P(t, x, y)$ on $(0, +\infty) \times S^n \times S^n$ satisfying

$$\frac{\partial P}{\partial t} = \frac{1}{2}\Delta_n P \tag{2.4}$$

$$P(t, x, y) \to \delta_x(y) \text{ as } t \to 0^+ \tag{2.5}$$

where Δ_n is the Laplace–Beltrami operator on S^n acting on the x-variables and $\delta_x(y)$ is the delta mass at x, i.e. $P(t, x, y)$ is the heat kernel of S^n. The heat kernel exists, it is unique, positive and smooth in (t, x, y).

2.2.1. Further properties of the heat kernel P(t, x, y)

Moreover the heat kernel possesses the following properties:

1. Symmetry in x, y, that is $P(t, x, y) = P(t, y, x)$.
2. The semigroup identity: For any $s \in (0, t)$
$$P(t, x, y) = \int_{S^n} P(s, x, z)P(t - s, z, y)d\mu z \tag{2.6}$$
where $d\mu$ is the area measure element of S^n.
3. As $t \to \infty$, $P(t, x, y)$ approaches the uniform density on S^n, i.e. $\lim\limits_{t\to\infty} P(t, x, y) = \frac{1}{A_n}$ where A_n is the area of the S^n with radius a. It is well known that
$A_n = \frac{2\pi^{\frac{n+1}{2}} a^n}{(\frac{n-1}{2})!}$ for n odd and $A_n = \frac{2^n(\frac{n}{2}-1)!\pi^{\frac{n}{2}} a^n}{(n-1)!}$ for n even
4. Finally, the symmetry of S^n implies that $P(t, x, y)$ depends only on t and $d(x, y)$, the distance between x and y. In spherical coordinates it depends on t and the angle φ between x and y. Hence $P(t, x, y) = P(t, \varphi)$ where $P(t, \varphi)$ satisfies
$$\frac{\partial P}{\partial t} = \frac{1}{2}\Delta_n P = \frac{1}{2a^2}\left[(n-1)\cot\varphi\frac{\partial P}{\partial\varphi} + \frac{\partial^2 P}{\partial\varphi^2}\right] \tag{2.7}$$
and
$$\lim_{t\to 0^+} aA_{n-1}P(t, \varphi)\sin^{n-1}\varphi = \delta(\varphi) \tag{2.8}$$

The symbol $\delta(\cdot)$ denotes the standard Dirac delta function on $\mathbb{R}$.

2.2.2. Explicit form of the heat kernel of S^1

Reminder (Poisson Summation Formula)**.** Let $f(x)$ be a function in the Schwartz space $\mathcal{S}(\mathbb{R})$, where $\mathcal{S}(\mathbb{R})$ consists of the set of all infinitely differentiable functions f on $\mathbb{R}$ so that f and all its derivatives $f^{(l)}$ are rapidly decreasing, in the sense that

$$\sup_{x\in\mathbb{R}} |x|^k \left|f^{(l)}(x)\right| < \infty \quad \text{for every} \quad k, l \geq 0.$$

Then

$$\sum_{n\in\mathbb{Z}} f(x + 2\pi n) = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} F(n)\exp(inx),$$

where $F(\xi)$ is the Fourier transform of $f(x)$, i.e.

$$F(\xi) = \int_{-\infty}^{+\infty} f(x)\exp(-i\xi x)dx, \quad \xi \in \mathbb{R}.$$

For example, if

$$f(x) = \exp(-Ax^2 + Bx), \quad A > 0, \quad B \in \mathbb{C},$$

then

$$F(\xi) = \sqrt{\frac{\pi}{A}}\exp\left(\frac{(i\xi - B)^2}{4A}\right)$$

2.3. The case of S^1

Proposition 2.1. *The transition density function of the Brownian motion X_t, $t \geq 0$ on S^1 with radius a is the function*

$$p(t, \varphi) = \frac{1}{2\pi a}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{n^2 t}{2a^2} + in\varphi\right). \tag{2.9}$$

Equivalently

$$p(t, \varphi) = \frac{1}{\pi a}\sum_{n\in\mathbb{N}}\left[\exp\left(-\frac{n^2 t}{2a^2}\right)\cos(n\varphi)\right] - \frac{1}{2\pi a}. \tag{2.10}$$

and

$$p(t, \varphi) = \frac{1}{\sqrt{2\pi t}}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{a^2}{2t}(\varphi - 2\pi n)^2\right). \tag{2.11}$$

Proof. If

$$p(t, \varphi) = \frac{1}{\pi a}\sum_{n\in\mathbb{N}}\left[\exp\left(-\frac{n^2 t}{2a^2}\right)\cos(n\varphi)\right] - \frac{1}{2\pi a},$$

then

$$\frac{\partial p(t, \varphi)}{\partial t} = -\frac{1}{2\pi a^3}\sum_{n\in\mathbb{N}} n^2\cos(n\varphi)\exp\left(-\frac{n^2 t}{2a^2}\right) \tag{2.12}$$

and

$$\frac{\partial^2 p(t, \varphi)}{\partial\varphi^2} = -\frac{1}{\pi a}\sum_{n\in\mathbb{N}} n^2\cos(n\varphi)\exp\left(-\frac{n^2 t}{2a^2}\right). \tag{2.13}$$

Therefore

$$\frac{\partial p(t, \varphi)}{\partial t} = \frac{1}{2a^2}\frac{\partial^2 p(t, \varphi)}{\partial\varphi^2}.$$

We will now show that

$$\lim_{t\to 0^+} ap(t, \varphi) = \delta(\varphi).$$

If $\varphi \in (0, 2\pi)$, then

$$\lim_{t\to 0^+} ap(t, \varphi) = 0. \tag{2.14}$$

Next we observe that

$$\int_0^{2\pi} ap(t,\varphi)d\varphi = \frac{1}{\pi}\int_0^{2\pi}\sum_{n\in\mathbb{N}}\left[\exp\left(-\frac{n^2t}{2a^2}\right)\cos(n\varphi)\right]d\varphi - 1. \tag{2.15}$$

For $t>0$ let us consider the functions

$f_n : [0, 2\pi] \to \mathbb{R}, \quad n\in\mathbb{N},$

with

$$f_n(\varphi) = \cos(n\varphi)\exp\left(-\frac{n^2t}{2a^2}\right).$$

Notice that $f_n(\varphi)$ are integrable functions on $[0, 2\pi]$. Furthermore

$$\sum_{n=1}^{+\infty} f_n(\varphi)$$

converges uniformly on $[0, 2\pi]$ because

$$|f_n(\varphi)| \le \exp\left(-\frac{n^2t}{2a^2}\right)$$

and the series

$$\sum_{n=1}^{\infty}\exp\left(-\frac{n^2t}{2a^2}\right)$$

converges. Therefore (2.15) gives

$$\int_0^{2\pi} ap(t,\varphi)d\varphi = -1 + \frac{1}{\pi}\sum_{n\in\mathbb{N}}\exp\left(-\frac{n^2t}{2a^2}\right)\int_0^{2\pi}\cos(n\varphi)d\varphi,$$

thus

$$\int_0^{2\pi} ap(t,\varphi)d\varphi = 1, \text{ for every } t>0. \tag{2.16}$$

Therefore from (2.15) and (2.16)

$$\lim_{t\to 0^+} ap(t,\varphi) = \delta(\varphi)$$

and this complete the proof. □

2.4. Geometric Brownian motion on a 1-dimentional sphere S^1

Definition 2.3. Let X_t, $t\ge 0$ be the Brownian motion on S^1 of radius a. The geometric Brownian motion on S^1 of radius a with drift is

$$Z_t = Z_0\exp\left[\left(r-\frac{1}{2}\sigma^2\right)t + \sigma aX_t\right] \tag{2.17}$$

i.e. Z_t have stochastic differential $dZ_t = rZ_tdt + \sigma aZ_tdX_t$.

We have already shown that, the Brownian motion on S^1 of radius a, in spherical coordinates is the solution of the stochastic differntial equation $dX_t = \frac{1}{a}dB_t$. Hence $dZ_t = rZ_tdt + \sigma Z_tdB_t$. Therefore, the generator L of Z_t is given by

$$Lf(\varphi) = r\varphi\frac{\partial f}{\partial\varphi} + \frac{1}{2}\sigma^2\varphi^2\frac{\partial^2 f}{\partial\varphi^2} \tag{2.18}$$

2.5. Transition density function of the geometric Brownian motion on S^1

Let $X_t, t\ge 0$ be the Brownian motion on S^1 of radius a. The transition density function of the Brownian motion $X_t, t\ge 0$ on S^1 of radius a is the function (2.9) i.e.

$$p(t,\varphi) = \frac{1}{2\pi a}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{n^2t}{2a^2} + in\varphi\right)$$

This means that $F_{X_t}(\varphi) = P[X_t \le \varphi] = \int_0^{\varphi} p(t,\varphi)d\varphi$. The geometric Brownian motion $Z_t, t\ge 0$ on S^1 of radius a is $Z_t = Z_0\exp[(r-\frac{1}{2}\sigma^2)t + \sigma aX_t]$. Hence,

$F_{Z_t}(\varphi) = P[Z_t \le \varphi] = P[Z_0\exp[(r-\frac{1}{2}\sigma^2)t + \sigma aX_t] \le \varphi] = P[X_t \le \frac{1}{\sigma a}\ln(\frac{\varphi}{Z_0}) + (\frac{\sigma}{2a} - \frac{r}{\sigma a})t] = \int_{-\infty}^{\frac{1}{\sigma a}\ln(\frac{\varphi}{Z_0})+(\frac{\sigma}{2a}-\frac{r}{\sigma a})t} p(t,y)dy.$

Now differentiating with respect to φ, we obtain that the transition density function of the geometric Brownian motion, is the function

$$p_{Z_t}(t,\varphi) = \frac{1}{\sigma a\varphi}p\left(t, \frac{1}{\sigma a}\ln\left(\frac{\varphi}{Z_0}\right) + \left(\frac{\sigma}{2a} - \frac{r}{\sigma a}\right)t\right)$$

i.e.

$$p_{Z_t}(t,\varphi) = \frac{1}{2\pi a^2\sigma\varphi}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{n^2t}{2a^2} + in\left(\frac{1}{\sigma a}\ln\left(\frac{\varphi}{z_0}\right) + \frac{\sigma^2-2r}{2\sigma a}t\right)\right) \tag{2.19}$$

3. Value of the derivative security

We limit ourselves to assume that the underlying asset follows a Geometric Brownian motion with drift, i.e. $dY_t = rY_tdt + \sigma Y_tdB_t$, where Y_t is the asset price and B_t, $t\ge 0$ is the Brownian motion.

Define $\Psi(Y_T) = (Y_T - k)^+ = \max\{(Y_T-k), 0\}$ be the payoff of the derivative security at time T if the underlying security is at Y_T (k is the strike price of the option). Assume that there is a double knock-out Barrier at levels $H_1, H_2 \in \mathbb{R}$ such that $H_1 < H_2$. i.e., if one of the barrier is reached in a double knock-out option, the option is killed. The idea is to consider the geometric Brownian motion on a 1−dimensional sphere $S^1 = \left\{x = (a\cos\varphi, a\sin\varphi) \in \mathbb{R}^2 | \ 0 \le \varphi < 2\pi\right\}$ (circle) with center at the origin and radius $a > \frac{H_2}{2\pi}$.

In this case there exist φ_0, $\varphi_1 \in [0, 2\pi)$ such that $H_1 = a\varphi_0$ and $H_2 = a\varphi_1$ and $k = a\varphi_k$. For arbitrary process S and $H_1, H_2 \in \mathbb{R}$ such that $H_1 < H_2$, we use the following notation according to [8]:

$\tau^Y_{H_1} = \inf\{t: \ Y(t) \le H_1\}$ if $Y(0) = H_1$ *and*

$\tau^Y_{H_2} = \inf\{t: \ Y(t) \ge H_2\}$ if $Y(0)H_2$

where H_{1} and H_2 are the Barriers.

Let $Y(t)$ be the value of the stock at time $t\in[0, T]$, where $T = \min\{\tau^y_{H_1}, \ \tau^y_{H_2}\}$. From the theory of arbitrage-free pricing in a complete market (see [11]), the value of the derivative security can then be expressed as follows

$$V(t,x) = V(T, H_1, H_2, Y(t), t) = E[\Psi(S_T)I(\tau^y_{H_1} < T)I(\tau^y_{H_2} < T) \tag{3.1}$$

where $\Psi(Z_T) = (Y_T - k)^+ = \max\{(Y_T-k), 0\}$ is the payoff of the derivative security at time T if the underlying security is at Y_T. The boundary problem for $V(t, x)$ can be tackled with fast and accurate pricing of Barrier options under Levy processes to solve it

$$\left\{\begin{array}{c} \frac{\partial V}{\partial t} + LV - rV = 0 \\ V(0,x) = (x_0 - k)^+ \\ V(t,H_1) = V(t,H_2) = 0 \\ V(t,x) = 0 \text{ for every } x\in(-\infty, H_1]\cup[H_2, +\infty) \end{array}\right\} \tag{3.2}$$

where L is the generator of Y_t.

In case Z_t is the Geometric Brownian motion without drift on a 1-dimensional sphere S^1 of radius $a > \frac{H_2}{2\pi}$, i.e. $dZ_t = \sigma aZ_tdX_t$ and we let ϕ_1, $\varphi_2 \in [0, 2\pi)$, such that $\varphi_1 < \varphi_2$ with $a\varphi_1 = H_1$ and $\varphi_2 = H_2$, then the problem (3.2) is equivelant to

$$\left\{\begin{array}{c} \frac{\partial u}{\partial t} + r\varphi\frac{\partial u}{\partial \varphi} + \frac{1}{2}\sigma^2\varphi^2\frac{\partial^2 u}{\partial \varphi^2} - ru = 0 \\ u(0,\varphi) = a(\varphi_0 - \varphi_k)^+ \\ u(t,\varphi_1) = u(t,\varphi_2) = 0 \\ u(t,\varphi) = 0 \text{ for every } \varphi \in D^c, \text{ where } D = (\varphi_1,\varphi_2) \subset S^1 \end{array}\right\} \tag{3.3}$$

where $a\varphi_0 = x_0$ and $a\varphi_k = k$

Theorem 3.1. *Under Black–Scholes framework the arbitrage-price of a knock-out call double barrier option is given by relation*

$$V(t,x) = \int_0^{\ln\left(\frac{H_2}{H_1}\right)} \frac{2\exp[-r(T-t)]}{\ln\left(\frac{H_2}{H_1}\right)}\left(e^{\xi}H_1 - k\right)^+ I_{[A(t)<\xi<B(t):\ t\in[0,T]]} \cdot \sum_{n\in\mathbb{Z}}\left[\exp\left[\frac{(n\pi\sigma)^2(T-t)}{2\ln^2\left(\frac{H_2}{H_1}\right)}\right]\sin\left(\frac{n\pi\xi}{\ln\left(\frac{H_2}{H_1}\right)}\right)\sin\left(\frac{n\pi\ln\left(\frac{x}{H_1}\right)}{\ln\left(\frac{H_2}{H_1}\right)}\right)\right]d\xi \tag{3.4}$$

where $A(t) = \ln H_1 + (T-t)(r - \frac{\sigma^2}{2})$ and $B(t) = \ln H_2 + (T-t)(r - \frac{\sigma^2}{2})$ (see [8])

4. Exit times

We recall some basic definitions:

Definition 4.1. A measurable space $\{\Omega, F\}$ *is said to be equipped with a filtration* $\{F_t\}$, $t \in [0,+\infty)$, if for every $t \geq 0$ $\{F_t\}$ is a σ-algebra of subsets of Ω such that $F_t \subset F$ and for every $t_1, t_2 \in [0,+\infty)$ such that $t_1 < t_2$, we have that $F_{t_1} \subset F_{t_2}$. (i.e. $\{F_t\}$ is an increasing family of sub σ-algebras of F).

Definition 4.2. Let us consider a measurable space $\{\Omega, F\}$ equipped with a filtration $\{F_t\}$. A random variable T is a stopping time with respect to the filtration $\{F_t\}$, if for every $t \geq 0$ $\{\omega \in \Omega \mid T(\omega) \leq t\} \in F_t$.

Let Z_t be the Geometric Brownian motion on S^1 and $D \subset S^1$ a domain. Then $T = \inf\{t \geq 0 \mid Z_t \notin D\}$ is a stopping time with respect to $F_t = \sigma\{Z_s \mid 0 \leq s \leq t\}$, called the exit time on ∂D.(For more details see [12]).

4.1. Expectations of exit times on S^1

Proposition 4.1. *Let $\varphi_1, \varphi_2 \in (0, 2\pi]$, such that $\varphi_1 < \varphi_2$, both fixed. We consider the set D in S^1, such that $D = (\varphi_1, \varphi_2)$. If Z_t is the Geometric Brownian motion with drift on S^1 of radius a starting at the point $\varphi \in D$, then the expectation of T is given by*

$$E^{\varphi}[T] = \frac{2}{\sigma^2 - 2r} \cdot \frac{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi - \left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_1 - \left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_2}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{4.1}$$

Proof. Reminder. If $u(x) = E^x[T]$, then $u(x)$ satisfies

$Lu(\varphi) = -1$

$u|_{\partial D} = 0$

(see [13])

Hence from (2.18) the differential equation takes the form

$$r\varphi\frac{\partial u}{\partial \varphi} + \frac{\sigma^2\varphi^2}{2}\frac{\partial^2 u}{\partial \varphi^2} = -1 \tag{4.2}$$

with boundary condition

$$u(\varphi_1) = u(\varphi_2) = 0 \tag{4.3}$$

From (4.2) and (4.3) we imply that

$$u(\varphi) = \frac{2}{\sigma^2 - 2r} \cdot \frac{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi - \left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_1 - \left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_2}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

i.e.

$$E^{\varphi}[T] = \frac{2}{\sigma^2 - 2r} \cdot \frac{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi - \left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_1 - \left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln\varphi_2}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \qquad \square$$

Based on the proof above, if Y_t is the asset price and we have a double knock-out barrier at level $H_1 = a\varphi_1$ and $H_2 = a\varphi_2$ then if its price starts at the point $x \in (H_1, H_2)$ the expectation of $T = \inf\{t \geq 0 \mid \text{the option is killed}\}$ is

$$E^x[T] = \frac{2}{\sigma^2 - 2r} \cdot \frac{\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln x - \left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - x^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln H_1 - \left(x^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)\ln H_2}{\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{4.4}$$

4.2. Expectation of f(Z$_t$)

Proposition 4.2. *Let $\varphi_1, \varphi_2 \in (0, 2\pi]$, such that $\varphi_1 < \varphi_2$, both fixed. We consider the set D in S^1, such that $D = (\varphi_1, \varphi_2)$. If Z_t is the Geometric Brownian motion with drift on S^1 of radius a starting at the point $\varphi \in D$, and f be a function on ∂D , then the expectation of $f(Z_t)$ is given by*

$$E^{\varphi}[f(Z_t)] = \frac{f(\varphi_2)\left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right) + f(\varphi_1)\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{4.5}$$

Proof. It is known that the function $u(\varphi) = E^{\varphi}[Z_t]$ satisfies the differential equation

$$Lu(\varphi) = 0$$

with boundary condition

$$u|_{\partial D} = f$$

(see [12])

Hence from (2.18) the differential equation takes the form

$$r\varphi \frac{\partial u}{\partial \varphi} + \frac{\sigma^2\varphi^2}{2}\frac{\partial^2 u}{\partial \varphi^2} = 0 \tag{4.6}$$

with boundary condition

$$u(\varphi_1) = f(\varphi_1) \text{ and } u(\varphi_2) = f(\varphi_2) \tag{4.7}$$

From (4.6) and (4.7) we imply that

$$u(\varphi) = \frac{f(\varphi_2)\left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right) + f(\varphi_1)\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

$$E^{\varphi}[f(Z_t)] = \frac{f(\varphi_2)\left(\varphi^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right) + f(\varphi_1)\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

□

4.3. Hitting probabilities

Proposition 4.3. *Let $\varphi_1, \varphi_2 \in (0, 2\pi]$, such that $\varphi_1 < \varphi_2$, both fixed. We consider the sets D_1, D_2 in S^1, such that $D_1 = (\varphi_1, 2\pi]$ and $D_2 = (0, \varphi_2)$. Let Z_t is the Geometric Brownian motion with drift on S^1 of radius a starting at the point $\varphi \in D_1 \cap D_2$. If*

$$T_1 = \inf\{t \geq 0 \mid Z_t \notin D_1\} \tag{4.8}$$

$$T_2 = \inf\{t \geq 0 \mid Z_t \notin D_2\} \tag{4.9}$$

and

$$T = \inf\{t \geq 0 \mid Z_t \notin D_1 \cap D_2\} \tag{4.10}$$

then the probabilities $\Pr^{\varphi}\{T = T_1\}$ and $\Pr^{\varphi}\{T = T_2\}$ are given by

$$\Pr^{\varphi}\{T = T_1\} = \frac{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}}{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}} \tag{4.11}$$

and

$$\overset{\varphi}{\Pr}\{T=T_2\}=\frac{\varphi^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}}{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}} \tag{4.12}$$

Proof. From (4.5) for $f(x)=x$ we have $E^{\varphi}[f(Z_t)]=\frac{\varphi_2(\varphi^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}})+\varphi_1(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi^{\frac{\sigma^2-2r}{\sigma^2}})}{(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}})}$.

However,

$$E^{\varphi}[f(Z_t)]=\varphi_1\overset{\varphi}{\Pr}\{T=T_1\}+\varphi_2\overset{\varphi}{\Pr}\{T=T_2\}$$

and

$$\overset{\varphi}{\Pr}\{T=T_1\}+\overset{\varphi}{\Pr}\{T=T_2\}=1$$

Therefore,

$$\overset{\varphi}{\Pr}\{T=T_1\}=\frac{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi^{\frac{\sigma^2-2r}{\sigma^2}}}{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}}$$

and

$$\overset{\varphi}{\Pr}\{T=T_2\}=\frac{\varphi^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}}{\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}}-\varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}}$$

□

Based on the proof above, in case Y_t is the asset price and we have a double knock-out Barrier at levels $H_1=a\varphi_1$ and $H_2=a\varphi_2$ then if its price starts at the point $x\in(H_1,H_2)$ the probability the option is killed because it reaches the barrier level H_1 is given as

$$\overset{x}{\Pr}\{T=T_1\}=\frac{H_2^{\frac{\sigma^2-2r}{\sigma^2}}-x^{\frac{\sigma^2-2r}{\sigma^2}}}{H_2^{\frac{\sigma^2-2r}{\sigma^2}}-H_1^{\frac{\sigma^2-2r}{\sigma^2}}} \tag{4.13}$$

and the proability the option is killed because it reaches the barrier level H_2 is

$$\overset{x}{\Pr}\{T=T_2\}=\frac{x^{\frac{\sigma^2-2r}{\sigma^2}}-H_1^{\frac{\sigma^2-2r}{\sigma^2}}}{H_2^{\frac{\sigma^2-2r}{\sigma^2}}-H_1^{\frac{\sigma^2-2r}{\sigma^2}}} \tag{4.14}$$

4.4. Moment generating functions

Proposition 4.4. *Let $\varphi_0,\varphi_1\in(0,2\pi]$, such that $\varphi_0<\varphi_1$ both fixed. We consider the set D on S^1such that $D=(\varphi_0,\varphi_1)$. If Z_t is the geometric Brownian motion on S^1 of radius a starting at the point $\varphi\in D$, then the expectation of* $\exp(-\lambda T)$ *is given by*

$$E^{\varphi}[\exp(-\lambda T)]$$
$$=\frac{\left(\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}-\varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)\varphi^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}-\varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}$$
$$+\frac{\left(\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}-\varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)\varphi^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}-\varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}$$

if $\lambda>-\dfrac{(2r-\sigma^2)^2}{8\sigma^2}$

$$E^{\varphi}[\exp(-\lambda T)]=\left(\frac{\varphi}{\varphi_1\varphi_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}}\frac{(\ln\varphi_2-\ln\varphi)\varphi_2^{\frac{\sigma^2-2r}{2\sigma^2}}-(\ln\varphi-\ln\varphi_1)\varphi_1^{\frac{\sigma^2-2r}{2\sigma^2}}}{\ln\varphi_2-\ln\varphi_1},\ \text{if } \lambda=\frac{\sigma^2-2r}{2\sigma^2} \tag{4.15}$$

$$E^{\varphi}[\exp(-\lambda T)] = \left(\frac{\varphi}{\varphi_1\varphi_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}} \frac{\varphi_2^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi_1}\right)\right]} + \frac{\varphi_1^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi}{\varphi_1}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi_1}\right)\right]} \tag{4.17}$$

$$\text{if } \lambda < -\frac{\left(2r-\sigma^2\right)^2}{8\sigma^2} \tag{4.18}$$

Proof. Reminder: Assume that $\lambda > -\frac{\lambda_1}{2}$, where λ_1 is the first Dirichlet eigenvalue of $D\subset S^1$. If $u(x) = E[\exp(-\lambda T)]$ then $u(x)$ satisfies $Lu(\varphi) = \lambda u(\varphi)$ with boundary conditions $u|_{\partial D} = 1$, (see [12])

The first dirichlet eigenvalue of $D\subset S^1$ is $\lambda = \frac{\pi^2}{a^2(\varphi_2-\varphi_1)^2}$. Hence if $\lambda > -\frac{\pi^2}{a^2(\varphi_2-\varphi_1)^2}$, then $E^{\varphi}[\exp(-\lambda T)]$ satisfies the differential equation

$$r\varphi\frac{\partial u}{\partial\varphi} + \frac{\sigma^2\varphi^2}{2}\frac{\partial^2 u}{\partial\varphi^2} = \lambda u(\varphi) \tag{4.19}$$

with boundary condition

$$u(\varphi_1) = u(\varphi_2) = 1 \tag{4.20}$$

This is a Cauchy–Euler equation. The solution is

$$E^{\varphi}[\exp(-\lambda T)] = \frac{\left(\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - \varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)\varphi^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - \varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}$$
$$+ \frac{\left(\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - \varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)\varphi^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{\varphi_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - \varphi_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\varphi_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}},$$
$$\text{if } \lambda > -\frac{\left(2r-\sigma^2\right)^2}{8\sigma^2}$$

$$E^{\varphi}[\exp(-\lambda T)] = \left(\frac{\varphi}{\varphi_1\varphi_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}} \frac{(\ln\varphi_2 - \ln\varphi)\varphi_2^{\frac{\sigma^2-2r}{2\sigma^2}} - (\ln\varphi - \ln\varphi_1)\varphi_1^{\frac{\sigma^2-2r}{2\sigma^2}}}{\ln\varphi_2 - \ln\varphi_1},\ \text{if } \lambda = \frac{\sigma^2-2r}{2\sigma^2} \tag{4.21}$$

$$E^{\varphi}[\exp(-\lambda T)] = \left(\frac{\varphi}{\varphi_1\varphi_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}} \frac{\varphi_2^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi_1}\right)\right]} + \frac{\varphi_1^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi}{\varphi_1}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2-8\lambda\sigma^2}\ln\left(\frac{\varphi_2}{\varphi_1}\right)\right]} \tag{4.22}$$

$$\text{if } \lambda < -\frac{\left(2r-\sigma^2\right)^2}{8\sigma^2} \tag{4.23}$$

i.e. if Y_t is the asset price and we have a double knock-out Barrier at levels $H_1 = a\varphi_1$ and $H_2 = a\varphi_2$, then if its price starts at the point $x\in(H_1, H_2)$

$$E^{x}[\exp(-\lambda T)] = \frac{\left(H_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - H_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)x^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{H_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}H_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - H_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}H_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}$$
$$+ \frac{\left(H_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - H_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}\right)H^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}}{H_1^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}H_2^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}} - H_1^{-\frac{(2r-\sigma^2)+\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}H_2^{-\frac{(2r-\sigma^2)-\sqrt{(2r-\sigma^2)^2+8\lambda\sigma^2}}{2\sigma^2}}},$$

$$\text{if } \lambda > -\frac{\left(2r-\sigma^2\right)^2}{8\sigma^2} \tag{4.24}$$

$$E^x[\exp(-\lambda T)] = \left(\frac{x}{H_1H_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}} \frac{(\ln H_2 - \ln x)H_2^{\frac{\sigma^2-2r}{2\sigma^2}} - (\ln x - \ln H_1)H_1^{\frac{\sigma^2-2r}{2\sigma^2}}}{\ln H_2 - \ln H_1}, \text{ if } \lambda = \frac{\sigma^2-2r}{2\sigma^2} \tag{4.25}$$

$$E^x[\exp(-\lambda T)] = \left(\frac{x}{H_1H_2}\right)^{\frac{\sigma^2-2r}{2\sigma^2}} \frac{H_2^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2 - 8\lambda\sigma^2} \ln\left(\frac{H_2}{x}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2 - 8\lambda\sigma^2} \ln\left(\frac{H_2}{H_1}\right)\right]} + \frac{H_1^{\frac{\sigma^2-2r}{2\sigma^2}} \sin\left[\sqrt{-\left(2r-\sigma^2\right)^2 - 8\lambda\sigma^2} \ln\left(\frac{x}{H_1}\right)\right]}{\sin\left[\sqrt{-\left(2r-\sigma^2\right)^2 - 8\lambda\sigma^2} \ln\left(\frac{H_2}{H_1}\right)\right]} \tag{4.26}$$

$$\text{if } \lambda < -\frac{\left(2r-\sigma^2\right)^2}{8\sigma^2} \tag{4.27}$$

□

5. Local time estimation

Definition 5.1. Let φ_0, $\varphi_1 \in (0, 2\pi]$, such that $\varphi_0 < \varphi_1$ both fixed. We consider the set D in S^1such that $D = (\varphi_0, \varphi_1)$. The Reflected Geometric Brownian motion in D is the diffusion W_t whose generator is L in D with Neuman boundary condition at ∂D.

Roughly speaking W_t behaves like Z_t inside D but when it reaches the boundary it is reflected back in D.

Definition 5.2. Let a fixed open set $D \subset S^1$ with C^3-boundary ∂D. If Y_t is the Reflected Geometric Brownian motion in$D$, and D_δ the domain

$D_\delta = \{x \in D \mid d(x, \partial D) < \delta\}$

we define the boundary local time $L_t$of W_t as

$$L_t = \lim_{\delta\to 0} \frac{1}{2\delta} \int_0^t 1_{D_\delta}(W_s) ds$$

It can be shown that the limit exists in L_2 sense.

5.1. Boundary local time until first hitting

Proposition 5.1. *Let φ_1, $\varphi_2 \in (0, 2\pi]$, such that $\varphi_1 < \varphi_2$ both fixed. We consider the set D in S^1such that $D = (\varphi_1, \varphi_2)$. Let W_t be the The Reflected Geometric Brownian motion on D starting at the point $\varphi \in D$. If*

$T = \inf\{t \geq 0 \mid Z_t = \varphi_1\}$

and L_t is the boundary local time of W_t, then

$$E^\varphi[\exp(\lambda L_t)] = \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}, \quad \text{if}$$

$$\lambda < \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{5.1}$$

and

$$E^\varphi[\exp(\lambda L_t)] = +\infty, \quad \text{if} \quad \lambda \geq \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{5.2}$$

Proof. It is known that the function $Z(\varphi) = E^\varphi[\exp(\lambda L_t)]$ satisfies the differential equation

$$Lz = 0 \tag{5.3}$$

with boundary conditions

$$z(\varphi_1) = 1 \tag{5.4}$$

and

$$-\frac{dz}{d\varphi}(\varphi_2) + \lambda z(\varphi_2) = 0 \tag{5.5}$$

as long as the function z is positive (see [14])

Thus

$$z(\varphi) = \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)} \tag{5.6}$$

However $z(\varphi) > 0$ for every $\varphi \in (\varphi_1, \varphi_2)$ if and only if $\lambda < \frac{(\sigma^2-2r)\varphi_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$. Therefore

$$E^{\varphi}[\exp(\lambda L_t)] = \frac{(\sigma^2 - 2r)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{(\sigma^2 - 2r)\varphi_2^{-\frac{2r}{\sigma^2}} - \lambda\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}, \quad \text{if}$$

$$\lambda < \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

and

$$E^{\varphi}[\exp(\lambda L_t)] = +\infty, \quad \text{if} \quad \lambda \geq \frac{\left(\sigma^2 - 2r\right)\varphi_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(\varphi_2^{\frac{\sigma^2-2r}{\sigma^2}} - \varphi_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

□

As a consequence of the proof, if we set $\tilde{\lambda} = a\lambda$, Y_t the asset price and we consider the double knock-out barrier at levels $H_1 = a\varphi_1$ and $H_2 = a\varphi_2$ then if its price starts at the point $x \in (H_1, H_2)$, then the expectation is given as

$$E^{x}\left[\exp\left(\tilde{\lambda}L_t\right)\right] = \frac{\left(\sigma^2 - 2r\right)H_2^{-\frac{2r}{\sigma^2}} - \tilde{\lambda}\sigma^2\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - x^{\frac{\sigma^2-2r}{\sigma^2}}\right)}{\left(\sigma^2 - 2r\right)H_2^{-\frac{2r}{\sigma^2}} - \tilde{\lambda}\sigma^2\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}, \quad if$$

$$\tilde{\lambda} < \frac{\left(\sigma^2 - 2r\right)H_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

and

$$E^{x}\left[\exp\left(\tilde{\lambda}L_t\right)\right] = +\infty, \quad \text{if} \quad \tilde{\lambda} \geq \frac{\left(\sigma^2 - 2r\right)H_2^{-\frac{2r}{\sigma^2}}}{\sigma^2\left(H_2^{\frac{\sigma^2-2r}{\sigma^2}} - H_1^{\frac{\sigma^2-2r}{\sigma^2}}\right)}$$

where L_t is the boundary local time of the option price.

6. Conclusion

Our novel methodology is not restricted to underlying processes which are geometric Brownian motions on S^1. Any other form of underlying process can be used, provided that there exists a transformation between the process and standard Brownian motion on S^1. Moreover, we can easily extend all the above results on spheres of higher dimensions. Our approach can be applied to the valuation of other exotic derivatives as well as to other mathematical problems. For example, Brownian motions on S^2 can be utilized for other types of derivatives' pricing in financial literature, epidemiological models and enviromental pollution models among other. Also for $n = 3$, S^3 some results appear in relativity theory [15]. We contributed in a plethora of ways. In particular, we presented new ways of estimating the expectation of the time the options seize to exist, their hitting probabilities, the exit times and their expectations, boundary local times until the first hitting and other probabilistic quantities related to the boundary local times. We deliver closed-form solutions which maybe of immense importance to traders, investors, speculators and more broadly to financial institutions.

References

[1] Luo L. Various types of double barrier options. J Comput Finance 2001;4:125–138.
[2] Merton R C. Theory of rational option pricing. Bell J Econ Manage Sci 1973;4:141–83.
[3] Rubinstein M, Reiner E. Breaking down the barriers. Risk 1991;4:28–35.
[4] Rich DR. The mathematical foundations of barrier option-pricing theory. Adv Futures Opt Res 1994;7:267–311.
[5] Kunitomo N, Ikeda M. Pricing options with curved boundaries. Math Finance 1992;2:275–98.
[6] Hui CH. One-touch double barrier binary option values. Appl Financ Econ 1996;6:343–6.
[7] Hui CH. Time-dependent barrier option values. J Futures Mark 1997;17:667–88.
[8] Roberts GO, Shortland CF. Pricing barrier options with time-dependent coefficients. Math Finance 1997;7:83–93.
[9] Dell'Era Mario MD. Pricing of double barrier options by spectral theory. MPRA Paper 17502. Germany: University Library of Munich; 2008.
[10] Do Carmo MP. Differential geometry of curves and surfaces prentice-Hall. New Jersey: Inc. Upper Saddle River; 1976.
[11] Harrison JM, Pliska SR. Martingales and stochastic integrals in the theory of continuous trading.stoch. Process Appl 1981;11:215–60.
[12] Oksendal B. Stochastic differential equations. Springer-Verlag; 1995.
[13] Dynkin EB. Markov processes. vol. 2. Berlin (1965): Springer; 1965.
[14] Papanicolaou VG. The probabilistic solutions of the third boundary value problem for second order elliptic equations. Probab Theory Relat Fields 1990;87:27–77.
[15] Kouloumpou D, Papanicolaou VG. Certain calculation regarding the Brownian motion on the sphere. J Concr Appl Math 2013;11:26–41.

J. CONCRETE AND APPLICABLE MATHEMATICS, VOL. 11, NO.'S 3-4, 303-316, 2013, COPYRIGHT 2013 EUDOXUS PRESS LLC

Certain Calculation Regarding the Brownian Motion on the Sphere

Dimitra Kouloumpou and Vassilis G. Papanicolaou
Department of Mathematics
National Technical University of Athens
Zografou Campus 157 80, Athens, GREECE.
dimkulub@mail.ntua.gr, papanico@math.ntua.gr

Abstract

We evaluate explicity certain quantities regarding the Brownian motion process on the n-dimensional sphere of radius a. First we review the transition densities of the process. Then we calculate some probabilistic quantities (e.g. moments) of the exit times of specific domains.
Key word and phrases: n-dimensional sphere, stereographic projection coordinates, Brownian motion, exit times, transition densities, reflection principle

1 Introduction

1.1 The n-Sphere

Let $n \in \mathbb{N} = \{1, 2, 3, \dots\}$. The n-dimensional sphere S^n with center $(c_1, ..., c_{n+1})$ and radius $a > 0$ is the set of all points $x \in \mathbb{R}^{n+1}$ satisfying

$$(x_1 - c_1)^2 + \cdots + (x_{n+1} - c_{n+1})^2 = a^2.$$

The most interesting case in applications is, of course, the case $n = 2$. For the sake of comparison we will also discuss the cases $n = 1$ (i.e. the circle) and $n = 3$. In some cases we will even consider the case of general n.

1.2 Stereographic Projection Coordinates

Consider the n-sphere, $n \geq 2$,

$$x_1^2 + \cdots + x_n^2 + (x_{n+1} - a)^2 = a^2.$$

To each point $(x_1, ..., x_n, x_{n+1})$ of this sphere, other than its "north pole" $N = (0, ..., 0, 2a)$ we associate the coordinates

$$\xi_1 = \frac{2ax_1}{2a - x_{n+1}}, \ \dots \ , \xi_n = \frac{2ax_n}{2a - x_{n+1}}.$$

Given the coordinates $(\xi_1, ..., \xi_n)$ of a point on the sphere with Cartesian coordinates $(x_1, ..., x_n, x_{n+1})$, we have

$$x_1 = \frac{4a^2\xi_1}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}, \ \ldots \ , x_n = \frac{4a^2\xi_n}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}, \ x_{n+1} = \frac{2a\left(\xi_1^2 + \cdots + \xi_n^2\right)}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}.$$

1.3 Spherical Coordinates

The points of the n-sphere

$$x_1^2 + \cdots + x_n^2 + x_{n+1}^2 = a^2$$

may also be described in spherical coordinates $(\theta_1, ..., \theta_{n-1}, \varphi)$ as follows:

- For $n = 1$, $x_1 = a\cos\varphi$, $x_2 = a\sin\varphi$, where $0 \leq \varphi < 2\pi$.
- In general for $n \geq 2$
  $x_1 = a\cos\theta_1 \prod_{i=1}^{n} \sin\theta_i$, $x_2 = a\prod_{i=2}^{n} \sin\theta_i$, $x_k = a\cos\theta_{k-1} \prod_{i=k}^{n} \sin\theta_i$,
  for $k = 3, 4, ..., n$
  and $x_{n+1} = a\cos\theta_n$, where $0 \leq \theta_1 < 2\pi$, $0 \leq \theta_i \leq \pi$,
  for $i = 2, 3, ..., n$,

1.4 The Laplace-Beltrami Operator

In spherical coordinates: The Laplace-Beltrami operator of a smooth function f on S^1 is

$$\Delta_1 f = \frac{1}{a^2}\frac{\partial^2 f}{\partial\varphi^2}. \tag{1.1}$$

The Laplace-Beltrami operator of a smooth function f on S^2 is

$$\Delta_2 f = \frac{1}{a^2\sin\varphi}\left(\frac{f_{\theta\theta}}{\sin\varphi} + f_\varphi\cos\varphi + f_{\varphi\varphi}\sin\varphi\right). \tag{1.2}$$

The Laplace-Beltrami operator of a smooth function f on S^3 is

$$\Delta_3 f = \frac{1}{a^2\sin^2\varphi}\left[\frac{1}{\sin^2\theta_2}\cdot\frac{\partial^2 f}{\partial\theta_1^2} + \frac{1}{\sin\theta_2}\cdot\frac{\partial}{\partial\theta_2}\left(\frac{\partial f}{\partial\theta_2}\sin\theta_2\right) + \frac{\partial}{\partial\varphi}\left(\frac{\partial f}{\partial\varphi}\sin^2\varphi\right)\right] \tag{1.3}$$

In stereographic projection coordinates: The Laplace-Beltrami operator of a smooth function f on S^n, $n \geq 2$ is

$$\Delta_n f = \frac{\left(\xi_1^2 + \cdots + \xi_n^2 + 4a^2\right)^2}{16a^4}\left[\sum_{i=1}^{n}\frac{\partial^2 f}{\partial\xi_i^2} - \frac{2(n-2)}{(\xi_1^2 + \cdots + \xi_n^2 + 4a^2)}\sum_{i=1}^{n}\xi_i\frac{\partial f}{\partial\xi_i}\right]. \tag{1.4}$$

In particular, for $n = 2$ we get

$$\Delta_2 f = \frac{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}{16a^4}\left(\frac{\partial^2 f}{\partial\xi_1^2} + \frac{\partial^2 f}{\partial\xi_2^2}\right). \tag{1.5}$$

1.5 Brownian Motion on S^n

The Brownian motion on S^n , starting from $x \in S^n$, is a diffusion (Markov) process $X_t, t \geq 0$, on S^n whose transition density is a function $P(t, x, y)$ on $(0, \infty) \times S^n \times S^n$ satisfying

$$\frac{\partial P}{\partial t} = \frac{1}{2} \Delta_n P \tag{1.6}$$

$$P\left(t, x, y\right) \to \delta_x(y) \quad \text{as} \quad t \to 0^+, \tag{1.7}$$

where Δ_n is the Laplace-Beltrami operator of S^n acting on the x-variables and $\delta_x(y)$ is the delta mass at x, i.e. $P(t, x, y)$ is the **heat kernel** of S^n. The heat kernel exists, it is unique, positive, and smooth in (t, x, y) [2].

1.5.1 Further Properties of the Heat Kernel $P(t, x, y)$

It is well known that $P(t, x, y)$ satisfies the following properties [2]

1. Symmetry: $P(t, x, y) = P(t, y, x)$.
2. The semigroup identity: For any $s \in (0, t)$,

   $$P(t, x, y) = \int_{S^n} P(s, x, z) P(t - s, z, y) d\mu(z)$$

   where $d\mu$ is the n-th dimensional surface area.
3. For all $t > 0$ and $x \in S^n$

   $$\int_{S^n} P(t, x, y) d\mu(y) = 1.$$

4. As $t \to \infty$, $P(t, x, y)$ approaches the uniform density on S^n, i.e.

   $$\lim_{t\to\infty} P(t, x, y) = \frac{1}{A_n},$$

   where A_n is nth dimensional surface area of S^n with radius a. It is well known that [3]

   $$A_{2k+1} = \frac{2\pi^{k+1} a^{2k+1}}{(k)!}, \quad \text{and} \quad A_{2k} = \frac{2^{2k}(k-1)!\pi^k a^{2k}}{(2k-1)!}, k \in \mathbb{N}$$

Finally, the symmetry of S^n implies that $P(t, x, y)$ depends only on t and $d(x, y)$, the distance between x and y. Thus in spherical coordinates it depends on t and the angle φ between x and y. Hence

$$P(t, x, y) = p(t, \varphi),$$

where $p(t, \varphi)$ satisfies

$$\frac{\partial p}{\partial t} = \frac{1}{2} \Delta_n p = \frac{1}{2a^2} \left[(n-1) \cot \varphi \cdot \frac{\partial p}{\partial \varphi} + \frac{\partial^2 p}{\partial \varphi^2} \right] \tag{1.8}$$

and

$$\lim_{t\to 0^+} a A_{n-1} p(t, \varphi) \cdot \sin^{n-1} \varphi = \delta(\varphi). \tag{1.9}$$

Here $\delta(\cdot)$ is the standard Dirac delta function on $\mathbb{R}$.

2 Explicit Form of the Heat Kernel

Reminder (Poisson Summation Formula). Let $f(x)$ be a function in the Schwartz space $\mathcal{S}(\mathbb{R})$, where $\mathcal{S}(\mathbb{R})$ consists of the set of all infinitely differentiable functions f on $\mathbb{R}$ so that f and all its derivatives $f^{(l)}$ are rapidly decreasing, in the sense that

$$\sup_{x\in\mathbb{R}} |x|^k \left| f^{(l)}(x) \right| < \infty \quad \text{for every} \quad k, l \geq 0.$$

Then

$$\sum_{n\in\mathbb{Z}} f(x + 2\pi n) = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} F(n) \exp(inx),$$

where $F(\xi)$ is the Fourier transform of $f(x)$, i.e.

$$F(\xi) = \int_{-\infty}^{+\infty} f(x) \exp(-i\xi x) dx, \quad \xi \in \mathbb{R}.$$

2.1 The Case of S^1

Proposition 2.1 *The transition density function of the Brownian motion $X_t, t \geq 0$ on S^1 with radius a is the function*

$$p(t,\varphi) = \frac{1}{2\pi a} \sum_{n\in\mathbb{Z}} \exp\left(-\frac{n^2 t}{2a^2} + in\varphi\right) = \frac{1}{\pi a} \sum_{n\in\mathbb{N}} \left[\exp\left(-\frac{n^2 t}{2a^2}\right)\cos(n\varphi)\right] - \frac{1}{2\pi a}, \tag{2.1}$$

equivalently

$$p(t,\varphi) = \frac{1}{\sqrt{2\pi t}} \sum_{n\in\mathbb{Z}} \exp\left(-\frac{a^2}{2t}(\varphi - 2\pi n)^2\right). \tag{2.2}$$

For the proof see [5].

2.2 The Case of S^2

We remind the reader of Legendre Polynomials $P_n(x)$, $n = 0, 1, 2, \ldots$ since we are going to use them later in the paper

$$P_n(x) = \frac{1}{2^n n!} \cdot \frac{d^n}{dx^n} \left[\left(x^2 - 1\right)^n \right].$$

Proposition 2.2 *The transition density function of the Brownian motion X_t, $t \geq 0$ on S^2 with radius a is given by the function*

$$p(t,\varphi) = \frac{1}{4\pi a^2} \sum_{n\in\mathbb{N}} (2n+1) \exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right) P_n(\cos\varphi). \tag{2.3}$$

For the proof see [1] or [5].

2.3 The Case of S^3

Proposition 2.3 *Let $X_t,\ t \geq 0$ be the Brownian motion on a 3-dimensional sphere S^3 of radius a. The transition density function $p(t, \varphi)$ of X_t is given by*

$$p(t,\varphi) = \frac{\exp\left(\frac{t}{2a^2}\right)}{(2\pi t)^{3/2}\sin\varphi} \sum_{n\in\mathbb{Z}} (\varphi + 2n\pi) \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right), \tag{2.4}$$

where $\mathbb{Z}$ is the set of all integers. Equivalently,

$$p(t,\varphi) = -\frac{i}{4\pi^2 a^3 \sin\varphi} \sum_{n\in\mathbb{Z}} n \exp\left(-\frac{t(n^2-1)}{2a^2} + i\varphi n\right), \tag{2.5}$$

$$p(t,\varphi) = \frac{1}{2\pi^2 a^3 \sin\varphi} \sum_{n\in\mathbb{N}} n \sin(n\varphi) \exp\left(-\frac{t(n^2-1)}{2a^2}\right). \tag{2.6}$$

The function $p(t, \varphi)$ is analytic at $\varphi = 0$ and $\varphi = \pi$. In fact

$$p(t,0) = \lim_{\varphi\to 0^+} p(t,\varphi) = \frac{1}{2\pi^2 a^3} \sum_{n\in\mathbb{N}} n^2 \exp\left(-\frac{t(n^2-1)}{2a^2}\right)$$

and

$$p(t,\pi) = \lim_{\varphi\to \pi^-} p(t,\varphi) = \frac{1}{2\pi^2 a^3} \sum_{n\in\mathbb{N}} n^2 (-1)^n \exp\left(-\frac{t(n^2-1)}{2a^2}\right).$$

For the proof see [5].

Reminder. The ϑ_3 function of Jacobi is

$$\vartheta_3(z,r) = 1 + 2\sum_{n=0}^{\infty} \exp\left(i\pi r n^2\right) \cos(2nz),$$

where $r \in \mathbb{C}$, with $\operatorname{Im}\{r\} > 0$. It follows that

$$p(t,\varphi) = -\frac{1}{4\pi^2 a^3 \sin\varphi} \exp\left(\frac{t}{2a^2}\right) \frac{\partial}{\partial\varphi} \vartheta_3\left(\frac{\varphi}{2}, \frac{ti}{2a^2\pi}\right).$$

3 Expectations of Exit Times

Let X_t be the Brownian motion in S^n and D a Borel subset of S^n. The random variable

$$T = \inf\{t \geq 0 |\ X_t \notin D\}$$

is called the (first) exit time of D.
Reminder. If $u(x) = E^x[T]$, then $u(x)$ satisfies

$$\frac{1}{2}\Delta_n u = -1, \quad u|_{\partial D} = 0 \tag{3.1}$$

(see, e.g., [8])

Proposition 3.1 *We consider the 2-dimensional sphere S^2 of radius a. Let two circles pass through the North pole, such that in stereographic coordinates are represented by the parallel lines $\xi_1 = b$ and $\xi_2 = c$, where $b, c \in \mathbb{R}$, say $b < c$. We consider the set D in S^2, whose stereographic projection is*

$$D = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \textit{and} \quad \xi_2 \in (b, c)\}.$$

Of course

$$\partial D = \{(\xi_1, \xi_2) \,|\, \xi_1 \in \mathbb{R} \quad \textit{and} \quad \xi_2 = b \quad \textit{or} \quad \xi_2 = c\}.$$

If X_t is the Brownian motion on S^2 of radius a starting at the point A, where the stereogrpaphic projection coordinates of A are $(\xi_1, \xi_2) \in D$ and

$$T = \inf\{t \geq 0 \,|\, X_t \in D\},$$

then

$$E^A[T] = f(\xi_1, \xi_2) - 2a^2 \ln\left(\xi_1^2 + \xi_2^2 + 4a^2\right), \tag{3.2}$$

where

$$f(\xi_1,\xi_2) = \frac{1}{\pi}\int_0^\infty \left[\frac{g(\eta, c)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \eta\right)^2}\right] d\eta$$

$$+\frac{1}{\pi}\int_0^\infty \left[\frac{g(\eta, b)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right) + \left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right) - \eta\right)^2}\right] d\eta \tag{3.3}$$

and

$$g(\xi, t) = 2a^2 \ln\left[\frac{(c-b)^2 \ln^2|\xi|}{\pi^2} + t^2 + 4a^2\right]. \tag{3.4}$$

Proof

The function

$$E^A[T] = U(\xi_1, \xi_2)$$

satisfies the differential equation

$$\frac{1}{2}\Delta_2 U = -1$$

with boundary conditions

$$U(\xi_1, b) = U(\xi_1, c) = 0.$$

Here Δ_2 is the Laplace-Beltrami operator on S^2 expressed in the stereographic projection coordinates (see (1.5)). Hence the differential equation takes the form

$$\frac{\partial^2 U}{\partial \xi_1^2} + \frac{\partial^2 U}{\partial \xi_2^2} = -\frac{32a^4}{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}. \tag{3.5}$$

However the function

$$U_1(\xi_1, \xi_2) = -2a^2 \ln(\xi_1^2 + \xi_2^2 + 4a^2)$$

satisfies the differential equation (3.5). Thus

$$U(\xi_1, \xi_2) = -2a^2 \ln(\xi_1^2 + \xi_2^2 + 4a^2) + f(\xi_1, \xi_2), \tag{3.6}$$

where $f(\xi_1, \xi_2)$ satisfies

$$\frac{\partial^2 f}{\partial \xi_1^2} + \frac{\partial^2 f}{\partial \xi_2^2} = 0,$$

with boundary conditions

$$f(\xi_1, b) = 2a^2 \ln(\xi_1^2 + b^2 + 4a^2) \quad \text{and} \quad f(\xi_1, c) = 2a^2 \ln(\xi_1^2 + c^2 + 4a^2). \tag{3.7}$$

If we make the change of variables $x = \xi_1$ and $y = \xi_2 - b$ and set the function $\phi(x, y) = f(\xi_1, \xi_2)$, then $\phi(x, y)$ satisfies

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0,$$

with boundary conditions

$$\phi(x, 0) = 2a^2 \ln(x^2 + b^2 + 4a^2) \quad \text{and} \quad \phi(x, \beta) = 2a^2 \ln(x^2 + c^2 + 4a^2),$$

where $\beta = c - b$.
Now let $z = x+yi$ and $w = \exp\left(\frac{\pi z}{\beta}\right)$, i.e. $z = \frac{\beta \ln w}{\pi}$. Thus, if $w = u+vi,\ u, v \in \mathbb{R}$ then

$$u = \exp\left(\frac{\pi x}{\beta}\right) \cos\left(\frac{\pi y}{\beta}\right) \quad \text{and} \quad v = \exp\left(\frac{\pi x}{\beta}\right) \sin\left(\frac{\pi y}{\beta}\right). \tag{3.8}$$

Introducing the function $\psi(u, v) = \phi(x, y)$, it follows that $\psi(u, v)$ satisfies

$$\frac{\partial^2 \psi}{\partial\, u^2} + \frac{\partial^2 \psi}{\partial\, v^2} = 0,$$

with boundary conditions

$$\psi(u, 0) = 2a^2 \ln\left(\frac{\beta^2 \ln^2 u}{\pi^2} + b^2 + 4a^2\right) \quad \text{for} \quad u > 0,$$

and

$$\psi(u, 0) = 2a^2 \ln\left(\frac{\beta^2 \ln^2 |u|}{\pi^2} + c^2 + 4a^2\right), \quad \text{for} \quad u < 0.$$

This is the standard Dirichlet boundary value problem for the half plane and it is well known [6] that its solution is given by the Poisson integral formula for the half-plane:

$$\psi(u, v) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{v\psi(\xi, 0)}{v^2 + (u - \xi)^2} d\xi,$$

or

$$\psi(u, v) = \frac{1}{\pi} \int_{-\infty}^{0} \frac{vg(\xi, c)}{v^2 + (u - \xi)^2} d\xi + \frac{1}{\pi} \int_{0}^{\infty} \frac{vg(\xi, b)}{v^2 + (u - \xi)^2} d\xi,$$

where
$$g(\xi,t)=2a^2\ln\left(\frac{\beta^2\ln^2|\xi|}{\pi^2}+t^2+4a^2\right).$$
Notice that $g(-\xi,t)=g(\xi,t)$. Hence
$$\psi(u,v)=\frac{1}{v}\pi\int_0^\infty\left(\frac{g(\xi,c)}{v^2+(u+\xi)^2}+\frac{g(\xi,b)}{v^2+(u-\xi)^2}\right)d\xi,$$
where u,v are given in (3.8). Therefore
$$\phi(x,y)=\frac{1}{\pi}\exp\left(\frac{\pi x}{\beta}\right)\sin\left(\frac{\pi y}{\beta}\right)\int_0^\infty\left(\frac{g(\eta,c)}{\exp\left(\frac{2\pi x}{\beta}\right)\sin^2\left(\frac{\pi y}{\beta}\right)+\left(\exp\left(\frac{\pi x}{\beta}\right)\cos\left(\frac{\pi y}{\beta}\right)+\eta\right)^2}\right)$$
$$+\frac{1}{\pi}\exp\left(\frac{\pi x}{\beta}\right)\sin\left(\frac{\pi y}{\beta}\right)\int_0^\infty\left(\frac{g(\eta,b)}{\exp\left(\frac{2\pi x}{\beta}\right)\sin^2\left(\frac{\pi y}{\beta}\right)+\left(\exp\left(\frac{\pi x}{\beta}\right)\cos\left(\frac{\pi y}{\beta}\right)-\eta\right)^2}\right)d\eta,$$
i.e.
$$f(\xi_1,\xi_2)=\frac{1}{\pi}\int_0^\infty\left[\frac{g(\eta,c)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right)+\left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right)+\eta\right)^2}\right]d\eta$$
$$+\frac{1}{\pi}\int_0^\infty\left[\frac{g(\eta,b)\exp\left(\frac{\pi\xi_1}{c-b}\right)\sin\left(\frac{\pi(\xi_2-b)}{c-b}\right)}{\exp\left(\frac{2\pi\xi_1}{c-b}\right)\sin^2\left(\frac{\pi(\xi_2-b)}{c-b}\right)+\left(\exp\left(\frac{\pi\xi_1}{c-b}\right)\cos\left(\frac{\pi(\xi_2-b)}{c-b}\right)-\eta\right)^2}\right]d\eta.$$
Therefore
$$E^A[T]=f(\xi_1,\xi_2)-2a^2\ln\left(\xi_1^2+\xi_2^2+4a^2\right).$$
■

4 Hitting Probabilities

Let X_t be the Brownian motion in S^n, $D\subset S^n$, and T its exit time.
Reminder. If $\Gamma\subset\partial D$ and $u(x)=P^x\{X_T\in\Gamma\}$ then [4] $u(x)$ satisfies
$$\Delta_n u=0,\quad u|_\Gamma=1,\quad u|_{\partial D\setminus\Gamma}=0.$$

Proposition 4.1 *We consider the 2-dimensional sphere S^2 of radius a. Let two circles passing through the North Pole, such that in stereographic coordinates are represented by the parallel lines $\xi_2=b$ and $\xi_2=c$, where $b,c\in\mathbb{R}$, with $b<c$. We consider the sets D_1,D_2 in S^2, whose stereographic projection are*
$$D_1=\{(\xi_1,\xi_2)\mid\xi_1\in\mathbb{R},\xi_2\in(b,+\infty)\}\quad\text{and}\quad D_2=\{(\xi_1,\xi_2)\mid\xi_1\in\mathbb{R},\xi_2\in(-\infty,c)\}.$$
Of course,
$$\partial D_1=\{(\xi_1,\xi_2)\mid\xi_1\in\mathbb{R},\xi_2=b\}\quad\text{and}\quad\partial D_2=\{(\xi_1,\xi_2)\mid\xi_1\in\mathbb{R},\xi_2=c\}.$$

Let X_t is the Brownian motion on S^2 of radius a starting at the point A, where the stereographic projection coordinate of A are

$$(\xi_1, \xi_2) \in D_1 \cap D_2.$$

If

$$T_1 = \inf \{t \geq 0 \,|\, X_t \notin D_1\}, T_2 = \inf \{t \geq 0 \,|\, X_t \notin D_2\}$$

and

$$T = \inf \{t \geq 0 \,|\, X_t \notin D_1 \cap D_2\},$$

then

$$P^A \{T = T_1\} = \frac{c - \xi_2}{c - b} \quad \text{and} \quad P^A \{T = T_2\} = \frac{\xi_2 - b}{c - b}. \tag{4.1}$$

Proof
The function

$$u(\xi_1, \xi_2) = P^A \{T = T_1\}$$

is the unique solution of the differential equation

$$\frac{1}{2}\Delta_2 u = 0,$$

or (see (1.5))

$$\frac{\partial^2 u}{\partial \xi_1^2} + \frac{\partial^2 u}{\partial \xi_2^2} = 0, \tag{4.2}$$

with boundary conditions

$$u(\xi_1, b) = 1 \quad \text{and} \quad u(\xi_1, c) = 0. \tag{4.3}$$

Since (4.2)-(4.3) has a unique solution, (4.1) follows immediately. ■

Remark 4.1 *In stereographic coordinates a function is harmonic with respect to Δ_2, (the Laplace-Beltrami operator of S^2), if and only if it is harmonic with respect to the standard Euclidean Laplacian. This fact is not true for $S^n, n \geq 3$.*

5 Reflection Principle on S^n and Applications

The following notation will be used in **Theorem 5.1**.

Definition 5.1 *For every $A = (x_1, x_2, \ldots, x_{n+1}) \in S^n$ we denote by $\hat{A}$ the point $(x_1, x_2, \ldots, -x_{n+1}) \in S^n$, namely the symmetric of A with respect to the $(x_1, x_2, \ldots, x_n)$-hyperplane.*

Theorem 5.1 *Let X_t, $t \geq 0$ be the Brownian motion on a n-dimensional sphere $(n \geq 2)$, S^n of radius a starting at the point*

$$A = (\theta_1, \ldots, \theta_n, \varphi) \in D,$$

where

$$D = \{(\theta_1, \ldots, \theta_{n-1}, \varphi) \in S^n |\ \theta_1 \in [0, 2\pi), \theta_i \in [0, \pi] \quad for \quad i = 2, \ldots, n-1 \quad and \quad \varphi \in \left(\frac{\pi}{2}, \pi\right]\right\}.$$

If

$$T = \inf \{ t \geq 0 |\ X_t \notin D \},$$

then

$$P^A \{T < t\} = 2P^A \{X_t \notin D\}. \tag{5.1}$$

Proof.

$$P^A \{T < t\} = P^A \{T < t, X_t \notin D\} + P^A \{T < t, X_t \in D\}. \tag{5.2}$$

However, if $X_t \notin D$ then of course $T < t$. Thus,

$$P^A \{T < t, X_t \notin D\} = P^A \{X_t \notin D\}. \tag{5.3}$$

On the other hand, if we set

$$\tilde{X}_t = \begin{cases} X_t, & \text{if} \quad t \leq T; \\ \hat{X}_t, & \text{if} \quad t > T, \end{cases}$$

where $\hat{X}_t$ is given by **Definition 5.1**, then by the strong Markov property of X_t, X_t and $\tilde{X}_t$ have the same law. Hence,

$$P^A \{T < t, X_t \in D\} = P^A \left\{T < t, \tilde{X}_t \in D\right\},$$

but if $\tilde{X}_t \in D$ then $X_t \notin D$. Hence,

$$P^A \left\{T < t, \tilde{X}_t \in D\right\} = P^A \{T < t, X_t \notin D\},$$

or

$$P^A \left\{T < t, \tilde{X}_t \in D\right\} = P^A \{X_t \notin D\}. \tag{5.4}$$

Therefore from (5.2), (5.3) and (5.4) we obtain that

$$P^A \{T < t\} = 2P^A \{X_t \notin D\}.$$

■

In the case of S^1 we can prove the next result in a similar manner.

Theorem 5.2 *Let X_t, $t \geq 0$ be the Brownian motion on a 1-dimensional sphere S^1 of radius a starting at the point $\varphi \in D$, where*

$$D = (\pi, 2\pi).$$

If

$$T = \inf \{ t \geq 0 |\ X_t \notin D \},$$

then

$$P^\varphi \{T < t\} = 2P^\varphi \{X_t \notin D\}. \tag{5.5}$$

5.0.1 Applications of the Reflection Principle

The reflection principle can help to calculate the distribution functions of certain exit times.

The case of S^1
Let X_t be the Brownian motion on a 1-dimensional sphere S^1 of radius a starting at the point φ. If $D = (\pi, 2\pi)$, then

$$P\{X_t \notin D\} = \int_0^{\pi} a \cdot p(t, x - \varphi)dx = \int_{-\varphi}^{\pi-\varphi} a \cdot p(t, y)dy,$$

where $p(t,\varphi)$ is the transition density function of the Brownian motion on S^1 of radius a. Hence, form (2.1)

$$P\{X_t \notin D\} = \int_{-\varphi}^{\pi-\varphi} a \left[\frac{1}{\pi a} \sum_{n\in\mathbb{N}} \left(\exp\left(-\frac{n^2 t}{2a^2}\right) \cos(ny) \right) - \frac{1}{2\pi a} \right] dy,$$

or

$$P\{X_t \notin D\} = -\frac{1}{2} + \frac{1}{\pi} \sum_{n\in\mathbb{N}} \left[\exp\left(-\frac{n^2 t}{2a^2}\right) \int_{-\varphi}^{\pi-\varphi} \cos(ny)dy \right].$$

Therefore,

$$P\{X_t \notin D\} = \frac{1}{2} + \frac{1}{\pi} \sum_{n\in\mathbb{N}^*} \left[\exp\left(-\frac{n^2 t}{2a^2}\right) \frac{\sin(n\pi - n\varphi) + \sin(n\varphi)}{n} \right],$$

i.e.

$$P\{X_t \notin D\} = \frac{1}{2} + \frac{1}{\pi} \sum_{n\in\mathbb{N}^*} \left[\exp\left(-\frac{n^2 t}{2a^2}\right) \frac{\sin(n\varphi)\left(1 - (-1)^n\right)}{n} \right].$$

Thus

$$P\{X_t \notin D\} = \frac{1}{2} + \frac{2}{\pi} \sum_{n \text{ odd}} \left(\frac{\exp\left(-\frac{n^2 t}{2a^2}\right) \sin(n\varphi)}{n} \right).$$

It follows (by using **Theorem 5.2**) that, if $T = \inf\{t \geq 0 | \; X_t \notin D\}$, then

$$P^{\varphi}\{T < t\} = 1 + \frac{4}{\pi} \sum_{n \text{ odd}} \frac{1}{n} \exp\left(-\frac{n^2 t}{2a^2}\right) \sin(n\varphi). \tag{5.6}$$

for every $\varphi \in (\pi, 2\pi)$.

The case of S^2
Let X_t be the Brownian motion on a 2-dimensional sphere S^2 of radius a starting at the point $N(0,0)$ in spherical coordinates. If

$$D = \left\{ (\theta,\varphi) \in S^2 \;\middle|\; \theta \in [0, 2\pi), \varphi \in \left(\frac{\pi}{2}, \pi\right] \right\}$$

then

$$P^N\{X_t \notin D\} = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} p(t,\varphi) a^2 \sin(\varphi) d\theta d\varphi,$$

i.e.
$$P^N\{X_t \notin D\} = 2\pi a^2 \int_0^{\frac{\pi}{2}} p(t,\varphi)\sin(\varphi)d\varphi,$$

where $p(t,\varphi)$ is the transition density function of the Brownian motion on S^2 of radius a. Hence from (2.3)

$$P^N\{X_t \notin D\} = 2\pi a^2 \int_0^{\frac{\pi}{2}} \frac{1}{4\pi a^2}\sin\varphi \sum_{n\in\mathbb{N}}(2n+1)\exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right)P_n(\cos\varphi)d\varphi,$$

or

$$P^N\{X_t \notin D\} = \frac{1}{2}+\frac{1}{2}\sum_{n\in\mathbb{N}^*}(2n+1)\exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right)\int_0^{\frac{\pi}{2}} P_n(\cos\varphi)\sin(\varphi)d\varphi. \tag{5.7}$$

However for every $n \in \mathbb{N}^*$

$$I = \int_0^{\frac{\pi}{2}} P_n(\cos\varphi)\sin(\varphi)d\varphi = \int_0^1 P_n(x)dx.$$

It is known that (see [7])

$$P_n(x) = \frac{1}{2n+1}\left[P_{n+1}(x) - P_{n-1}(x)\right].$$

Thus
$$I = \frac{1}{2n+1}\left(P_{n+1}(1) - P_{n-1}(1) - P_{n+1}(0) + P_{n-1}(0)\right),$$
or
$$I = \frac{1}{2n+1}\left(P_{n-1}(0) - P_{n+1}(0)\right).$$
It is also known that for every $n \in \mathbb{N}^*$

$$P_{2n}(0) = (-1)^n\frac{(2n)!}{2^{2n}(n!)^2} \quad \text{and} \quad P_{2n+1}(0) = 0.$$

Thus, if n is even then $I = 0$.
If n is odd, i.e. $n = 2k+1$, then

$$I = \frac{1}{4k+3}\left(P_{2k}(0) - P_{2(n+1)}(0)\right),$$

i.e.
$$I = \frac{(-1)^n(2k)!(2k+3)}{(4k+3)2^{2k+1}k!}. \tag{5.8}$$

From (5.7) and (5.8) we obtain that

$$P^N\{X_t \notin D\} = \frac{1}{2}+\frac{1}{2}\sum_{n\in\mathbb{N}}(-1)^n\exp\left(-\frac{(2n+1)(2n+1)\sqrt{t}}{a}\right)\cdot\frac{(2n)!(2n+3)}{2^{2n+1}n!}. \tag{5.9}$$

Furthermore, if $S(0,\pi)$ namely the south pole of S^2, then

$$P^S\{X_t \notin D\} = P^N\{\hat{X}_t \notin D\} = P^N\{X_t \in D\} = 1 - P^N\{X_t \notin D\}.$$

Therefore,

$$P^S\{X_t \notin D\} = \frac{1}{2} - \frac{1}{2}\sum_{n\in\mathbb{N}}(-1)^n \exp\left(-\frac{(2n+1)(2n+1)\sqrt{t}}{a}\right)\cdot\frac{(2n)!(2n+3)}{2^{2n+1}n!}. \tag{5.10}$$

Theorem 5.1 implies that, if $T = \inf\{\, t > 0 \mid X_t \notin D\}$, then

$$P^S\{T < t\} = 1 - \sum_{n\in\mathbb{N}}(-1)^n \exp\left(-\frac{(2n+1)(2n+1)\sqrt{t}}{a}\right)\cdot\frac{(2n)!(2n+3)}{2^{2n+1}n!}. \tag{5.11}$$

The case of S^3
Let X_t be the Brownian motion on a 3-dimensional sphere S^3 of radius a starting at the point $N(0,0,0)$ in spherical coordinates. If

$$D = \left\{(\theta_1,\theta_2,\varphi)\in S^3 \;\middle|\; \theta_1\in[0,2\pi),\theta_2\in[0,\pi],\varphi\in\left(\frac{\pi}{2},\pi\right]\right\}$$

then

$$P^N\{X_t\notin D\} = \int_0^{\frac{\pi}{2}}\int_0^{\pi}\int_0^{2\pi} p(t,\varphi)a^3\sin\theta_2\sin^2(\varphi)d\theta_1 d\theta_2 d\varphi,$$

i.e.

$$P^N\{X_t\notin D\} = 4\pi a^3\int_0^{\frac{\pi}{2}} p(t,\varphi)\sin^2(\varphi)d\varphi,$$

where $p(t,\varphi)$ is the transition density function of the Brownian motion on S^3 of radius a. Hence from (2.6)

$$P^N\{X_t\notin D\} = 4\pi a^3\int_0^{\frac{\pi}{2}}\sin^2(\varphi)\frac{1}{2\pi^2a^3\sin(\varphi)}\sum_{n\in\mathbb{N}} n\sin(n\varphi)\exp\left(-\frac{t(n^2-1)}{2a^2}\right)d\varphi,$$

or

$$P^N\{X_t\notin D\} = \frac{2}{\pi}\sum_{n\in\mathbb{N}} n\exp\left(-\frac{t(n^2-1)}{2a^2}\right)\int_0^{\frac{\pi}{2}}\sin(\varphi)\sin(n\varphi)d\varphi. \tag{5.12}$$

Let us call

$$I = \int_0^{\frac{\pi}{2}}\sin(\varphi)\sin(n\varphi)d\varphi.$$

If $n = 1$, then $I = \frac{\pi}{4}$. If $n > 1$, then

$$I = -\frac{n\cos\left(\frac{n\pi}{2}\right)}{n^2-1}.$$

Thus from (5.12),

$$P^N\{X_t\notin D\} = \frac{1}{2} - \frac{2}{\pi}\sum_{n=2}^{\infty} n^2\exp\left(-\frac{t(n^2-1)}{2a^2}\right)\cos\left(\frac{n\pi}{2}\right).$$

However, $\cos\left(\frac{n\pi}{2}\right) = 0$ for every n odd, hence

$$P^N\{X_t\notin D\} = \frac{1}{2} - \frac{2}{\pi}\sum_{n\ even} n^2\exp\left(-\frac{t(n^2-1)}{2a^2}\right)\cos\left(\frac{n\pi}{2}\right),$$

or

$$P^N\{X_t \notin D\} = \frac{1}{2} - \frac{8}{\pi}\sum_{n\in\mathbb{N}^*}(-1)^n n^2 \exp\left(-\frac{t(4n^2-1)}{2a^2}\right). \tag{5.13}$$

Furthermore, if $S = (0, 0, \pi)$ then,

$$P^S\{X_t \notin D\} = P^N\{\hat{X}_t \notin D\} = P^N\{X_t \in D\} = 1 - P^N\{X_t \notin D\}.$$

Therefore,

$$P^S\{X_t \notin D\} = \frac{1}{2} + \frac{8}{\pi}\sum_{n\in\mathbb{N}^*}(-1)^n n^2 \exp\left(-\frac{t(4n^2-1)}{2a^2}\right). \tag{5.14}$$

Theorem 5.1 implies that, if $T = \inf\{t > 0 \mid X_t \notin D\}$, then

$$P^S\{T < t\} = 1 + \frac{16}{\pi}\sum_{n\in\mathbb{N}^*}(-1)^n n^2 \exp\left(-\frac{t(4n^2-1)}{2a^2}\right). \tag{5.15}$$

Acknowledgment. This work was partially supported by a Π.Ε.Β.Ε. grant of the National Technical University of Athens.

References

[1] Chung M.K., Heat Kernel Smoothing On Unit Sphere *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 992–995 (2006).

[2] Dodziuk J., Maximum Principle for Parabolic Inequalities and the Heat Flow on Open Manifolds, *Indiana Univ. Math. J.*, **32** (no.5), 115–142 (1983).

[3] John F., *Partial Differential Equations*, Springer, USA (1982).

[4] Klebaner F.C., *Introduction to Stochastic Calculus with Applications,* Imperial College Press, Melbourne (2004).

[5] Kouloumpou D. and Papanicolaou V.G., The Random Motion on the Sphere Generated by the Laplace-Beltrami Operator, *Journal of Applied Functional Analysis,* **7** (1-2) 26-41 (2012).

[6] Spiegel M. R., *Complex Variables With an Introduction to Conformal Mapping and Its Applications,* Schaum's Outline Series (1999).

[7] Strauss W.A., *Partial Differential Equations,* John Wiley and Sons, Inc (1992).

[8] Øksendal B., *Stochastic Differential Equations,* Springer-Verlag (1995).

The random motion on the sphere generated by the Laplace-Beltrami operator

Article · January 2012

2 authors:


Dimitra Kouloumpou
Technological Educational Institute of Peloponnese


Vassilis George Papanicolaou
National Technical University of Athens

JOURNAL OF APPLIED FUNCTIONAL ANALYSIS, VOL.7, NO'S 1-2,26-41,COPYRIGHT 2012 EUDOXUS PRESS,LLC

The Random Motion on the Sphere Generated by the Laplace-Beltrami Operator

Dimitra Kouloumpou
Department of Mathematics
National Technical University of Athens
Zografou Campus 157 80, Athens, GREECE.
dimkulub@mail.ntua.gr
and
Vassilis G. Papanicolaou
Department of Mathematics
National Technical University of Athens
Zografou Campus 157 80, Athens, GREECE.
papanico@math.ntua.gr

Abstract

Using the Laplace-Beltrami operator we construct the Brownian motion process on the n-dimensional sphere, $n = 1, 2, 3$. Then we evaluate explicitly certain quantities for this process. We start with the transition density and continue with the calculation of some probabilistic quantities regarding the exit times of specific domains possessing certain symmetries.

Key words and phrases: n-dimensional sphere, Laplace-Beltrami operator, Brownian motion, transition densities, exit times.

1

1 Introduction

1.1 The n-Sphere

Let $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$. The n-dimensional sphere S^n with center $(c_1, ..., c_{n+1})$ and radius $a > 0$ is the set of all points $x \in \mathbb{R}^{n+1}$ satisfying

$$(x_1 - c_1)^2 + \cdots + (x_{n+1} - c_{n+1})^2 = a^2$$

The most interesting case in applications is, of course, the case $n = 2$. For the sake of comparison we will also discuss the cases $n = 1$ (i.e. the circle) and $n = 3$. In some cases we will even consider the case of general n.

1.2 Stereographic Projection Coordinates

Consider the n-sphere, $n \geq 2$,

$$x_1^2 + \cdots + x_n^2 + (x_{n+1} - a)^2 = a^2$$

To each point $(x_1, ..., x_n, x_{n+1})$ of this sphere, other than its "north pole" $N = (0, ..., 0, 2a)$ we associate the coordinates

$$\xi_1 = \frac{2ax_1}{2a - x_{n+1}}, \ \ldots \ , \xi_n = \frac{2ax_n}{2a - x_{n+1}}$$

Given the coordinates $(\xi_1, ..., \xi_n)$ of a point on the sphere with Cartesian coordinates $(x_1, ..., x_n, x_{n+1})$, we have

$$x_1 = \frac{4a^2 \xi_1}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}, \ \ldots \ , x_n = \frac{4a^2 \xi_n}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}, \ x_{n+1} = \frac{2a\left(\xi_1^2 + \cdots + \xi_n^2\right)}{\xi_1^2 + \cdots + \xi_n^2 + 4a^2}.$$

1.3 Spherical Coordinates

The points of the n-sphere

$$x_1^2 + \cdots + x_n^2 + x_{n+1}^2 = a^2$$

may also be described in spherical coordinates $(\theta_1, ..., \theta_{n-1}, \varphi)$ as follows:

- For $n = 1$, $x_1 = a\cos\varphi$, $x_2 = a\sin\varphi$, where $0 \leq \varphi < 2\pi$.
- For $n = 2$, $(\theta_1 = \theta)$ $x_1 = a\cos\theta\sin\varphi$, $x_2 = a\sin\theta\sin\varphi$, $x_3 = a\cos\varphi$, where $0 \leq \theta < 2\pi$ and $0 \leq \varphi \leq \pi$.
- For $n = 3$, $x_1 = a\cos\theta_1\sin\theta_2\sin\varphi$, $x_2 = a\sin\theta_1\sin\theta_2\sin\varphi$, $x_3 = a\cos\theta_2\sin\varphi$, $x_4 = a\cos\varphi$, where $0 \leq \theta_1 < 2\pi$, $0 \leq \theta_2 \leq \pi$, and $0 \leq \varphi \leq \pi$.
- In general for $n \geq 4$
  $x_1 = a\cos\theta_1\sin\theta_2\sin\theta_3 ... \sin\theta_{n-1}\sin\varphi$, $x_2 = a\sin\theta_1\sin\theta_2\sin\theta_3 ... \sin\theta_{n-1}\sin\varphi$,
  $x_k = a\cos\theta_{k-1}\sin\theta_k ... \sin\theta_{n-1}\sin\varphi$, for $k = 3, 4, ..., n$
  and $x_{n+1} = a\cos\varphi$, where $0 \leq \theta_1 < 2\pi$, $0 \leq \theta_i \leq \pi$,
  for $i = 2, 3, ..., n-1$, and $0 \leq \varphi \leq \pi$.

1.4 The Laplace-Beltrami Operator

In spherical coordinates: The Laplace-Beltrami operator of a smooth function f on S^1 is

$$\Delta_1 f = \frac{1}{a^2} \cdot \frac{\partial^2 f}{\partial \varphi^2}$$

The Laplace-Beltrami operator of a smooth function f on S^2 is

$$\Delta_2 f = \frac{1}{a^2 \sin\varphi} \left(\frac{f_{\theta\theta}}{\sin\varphi} + f_\varphi \cos\varphi + f_{\varphi\varphi} \sin\varphi \right)$$

In the case where f is independent of θ we have

$$\Delta_2 f = \frac{1}{a^2} \left(f_{\varphi\varphi} + f_\varphi \cot\varphi \right)$$

The Laplace-Beltrami operator of a smooth function f on S^3 is

$$\Delta_3 f = \frac{1}{a^2 \sin^2\varphi} \left[\frac{1}{\sin^2\theta_2} \cdot \frac{\partial^2 f}{\partial \theta_1^2} + \frac{1}{\sin\theta_2} \cdot \frac{\partial}{\partial\theta_2} \left(\frac{\partial f}{\partial \theta_2} \sin\theta_2 \right) + \frac{\partial}{\partial\varphi} \left(\frac{\partial f}{\partial\varphi} \sin^2\varphi \right) \right]$$

and if f is independent of θ_1 and θ_2,

$$\Delta_3 f = \frac{1}{a^2} \left(f_{\varphi\varphi} + 2 f_\varphi \cot\varphi \right)$$

In stereographic projection coordinates: The Laplace-Beltrami operator of a smooth function f on S^n, $n \geq 2$ is

$$\Delta_n f = \frac{\left(\xi_1^2 + \cdots + \xi_n^2 + 4a^2\right)^2}{16 a^4} \left[\sum_{i=1}^{n} \frac{\partial^2 f}{\partial \xi_i^2} - \frac{2(n-2)}{(\xi_1^2 + \cdots + \xi_n^2 + 4a^2)} \sum_{i=1}^{n} \xi_i \frac{\partial f}{\partial \xi_i} \right]$$

In particular, for $n = 2$ we get

$$\Delta_2 f = \frac{\left(\xi_1^2 + \xi_2^2 + 4a^2\right)^2}{16 a^4} \left(\frac{\partial^2 f}{\partial \xi_1^2} + \frac{\partial^2 f}{\partial \xi_2^2} \right)$$

1.5 Brownian motion on S^n (starting at $x \in S^n$)

The Brownian motion on S^n is a diffusion (Markov) process $X_t, t \geq 0$, on S^n whose transition density is a function $P(t, x, y)$ on $(0, \infty) \times S^n \times S^n$ satisfying

$$\frac{\partial P}{\partial t} = \frac{1}{2} \Delta_n P,$$

$$P(t, x, y) \to \delta_x(y) \quad \text{as} \quad t \to 0^+$$

where Δ_n is the Laplace-Beltrami operator of S^n acting on the x-variables and $\delta_x(y)$ is the delta mass at x, i.e. $P(t, x, y)$ is the **heat kernel** of S^n. The heat kernel exists, it is unique, positive, and smooth in (t, x, y) [4].

1.5.1 Further Properties of the Heat Kernel $P(t,x,y)$

It is well known that $p(t,x,y)$ satisfies the following properties [4]

1. Symmetry: $P(t,x,y) = P(t,y,x)$
2. The semigroup identity: For any $s \in (0,t)$,
$$P(t,x,y) = \int_{S^n} P(s,x,z)P(t-s,z,y)d\mu(z)$$
where $d\mu$ is the n-th dimensional surface area.
3. For all $t > 0$ and $x \in S^n$
$$\int_{S^n} P(t,x,y)d\mu(y) = 1$$
4. As $t \to \infty$, $P(t,x,y)$ approaches the uniform density on S^n, i.e.
$$\lim_{t\to\infty} P(t,x,y) = \frac{1}{A_n}$$
where A_n is nth dimensional surface area of S^n with radius a. It is well known that [8]
$$A_n = \frac{2\pi^{\frac{n+1}{2}}a^n}{(\frac{n-1}{2})!}, \qquad \text{for } n \text{ odd}$$
$$A_n = \frac{2^n(\frac{n}{2}-1)!\pi^{\frac{n}{2}}a^n}{(n-1)!}, \qquad \text{for } n \text{ even.}$$

Finally, the symmetry of S^n implies that $P(t,x,y)$ depends only on t and $d(x,y)$, the distance between x and y. Thus in spherical coordinates it depends on t and the angle φ between x and y. Hence
$$P(t,x,y) = p(t,\varphi),$$
where $p(t,\varphi)$ satisfies
$$\frac{\partial p}{\partial t} = \frac{1}{2}\Delta_n p = \frac{1}{2a^2}\left[(n-1)\cot\varphi \cdot \frac{\partial p}{\partial \varphi} + \frac{\partial^2 p}{\partial \varphi^2}\right]$$
and
$$\lim_{t\to 0^+} aA_{n-1}p(t,\varphi)\cdot \sin^{n-1}\varphi = \delta(\varphi).$$
Here $\delta(\cdot)$ is the standard Dirac delta function on $\mathbb{R}$.

2 Explicit Form of the Heat Kernel

Reminder (Poisson Summation Formula)**.** Let $f(x)$ be a function in the Schwartz space $\mathcal{S}(\mathbb{R})$, where $\mathcal{S}(\mathbb{R})$ consists of the set of all infinitely differentiable functions f on $\mathbb{R}$ so that f and all its derivatives $f^{(l)}$ are rapidly decreasing, in the sense that
$$\sup_{x\in\mathbb{R}} |x|^k \left|f^{(l)}(x)\right| < \infty \quad \text{for every} \quad k,l \geq 0.$$

Then

$$\sum_{n\in\mathbb{Z}} f(x+2\pi n) = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} F(n)\exp(inx),$$

where $F(\xi)$ is the Fourier transform of $f(x)$, i.e.

$$F(\xi) = \int_{-\infty}^{+\infty} f(x)\exp(-i\xi x)dx, \quad \xi\in\mathbb{R}.$$

For example, if

$$f(x) = \exp(-Ax^2+Bx), \quad A>0, \quad B\in\mathbb{C},$$

then

$$F(\xi) = \sqrt{\frac{\pi}{A}}\exp\left(\frac{(i\xi-B)^2}{4A}\right)$$

2.1 The Case of S^1

Proposition 2.1 *The transition density function of the Brownian motion $X_t, t\geq 0$ on S^1 with radius a is the function*

$$p(t,\varphi) = \frac{1}{2\pi a}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{n^2t}{2a^2}+in\varphi\right).$$

Equivalently

$$p(t,\varphi) = \frac{1}{\pi a}\sum_{n\in\mathbb{N}}\left[\exp\left(-\frac{n^2t}{2a^2}\right)\cos(n\varphi)\right] - \frac{1}{2\pi a}$$

and

$$p(t,\varphi) = \frac{1}{\sqrt{2\pi t}}\sum_{n\in\mathbb{Z}}\exp\left(-\frac{a^2}{2t}\left(\varphi-2\pi n\right)^2\right).$$

Proof. If

$$p(t,\varphi) = \frac{1}{\pi a}\sum_{n\in\mathbb{N}}\left[\exp\left(-\frac{n^2t}{2a^2}\right)\cos(n\varphi)\right] - \frac{1}{2\pi a},$$

then

$$\frac{\partial p(t,\varphi)}{\partial t} = -\frac{1}{2\pi a^3}\sum_{n\in\mathbb{N}} n^2\cos(n\varphi)\exp\left(-\frac{n^2t}{2a^2}\right) \tag{2.1}$$

and

$$\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2} = -\frac{1}{\pi a}\sum_{n\in\mathbb{N}} n^2\cos(n\varphi)\exp\left(-\frac{n^2t}{2a^2}\right). \tag{2.2}$$

Therefore

$$\frac{\partial p(t,\varphi)}{\partial t} = \frac{1}{2a^2}\frac{\partial^2 p(t,\varphi)}{\partial\varphi^2}.$$

We will now show that

$$\lim_{t\to 0^+} ap(t,\varphi) = \delta(\varphi).$$

If $\varphi \in (0, 2\pi)$, then

$$\lim_{t \to 0^+} ap(t, \varphi) = 0. \tag{2.3}$$

Next we observe that

$$\int_0^{2\pi} ap(t, \varphi) d\varphi = \frac{1}{\pi} \int_0^{2\pi} \sum_{n \in \mathbb{N}} \left[\exp\left(-\frac{n^2 t}{2a^2} \right) \cos(n\varphi) \right] d\varphi - 1. \tag{2.4}$$

For $t > 0$ let us consider the functions

$$f_n : [0, 2\pi] \to \mathbb{R}, \quad n \in \mathbb{N},$$

with

$$f_n(\varphi) = \cos(n\varphi) \exp\left(-\frac{n^2 t}{2a^2} \right).$$

Notice that $f_n(\varphi)$ are integrable functions on $[0, 2\pi]$. Furthermore

$$\sum_{n=1}^{+\infty} f_n(\varphi)$$

converges uniformly on $[0, 2\pi]$ because

$$|f_n(\varphi)| \leq \exp\left(-\frac{n^2 t}{2a^2} \right)$$

and the series

$$\sum_{n=1}^{\infty} \exp\left(-\frac{n^2 t}{2a^2} \right)$$

converges. Therefore (2.4) gives

$$\int_0^{2\pi} ap(t, \varphi) d\varphi = -1 + \frac{1}{\pi} \sum_{n \in \mathbb{N}} \exp\left(-\frac{n^2 t}{2a^2} \right) \int_0^{2\pi} \cos(n\varphi) d\varphi,$$

thus

$$\int_0^{2\pi} ap(t, \varphi) d\varphi = 1, \text{ for every } t > 0. \tag{2.5}$$

Therefore from (2.4) and (2.5)

$$\lim_{t \to 0^+} ap(t, \varphi) = \delta(\varphi)$$

and this complete the proof. ∎

2.2 The Case of S^2

Let $X_t,\ t \geq 0$ be the Brownian motion on a 2-dimensional sphere S^2 of radius a. The transition density function $p(t, \varphi)$ of X_t is the unique solution of

$$\frac{\partial p}{\partial t} = \frac{1}{2a^2 \sin \varphi} \left(\frac{\partial^2 p(t, \varphi)}{\partial \varphi^2} \sin \varphi + \frac{\partial p}{\partial \varphi} \cos \varphi \right) \tag{2.6}$$

and

$$\lim_{t\to 0^+} 2\pi a^2 \sin(\varphi)p(t,\varphi) = \delta(\varphi). \tag{2.7}$$

The solution to the diffusion equation

$$\frac{\partial K(t,\varphi)}{\partial t} = \frac{1}{\sin\varphi}\left(\cos\varphi\frac{\partial K(t,\varphi)}{\partial\varphi} + \sin\varphi\frac{\partial^2 K(t,\varphi)}{\partial\varphi^2}\right) \tag{2.8}$$

with initial condition

$$\lim_{t\to 0^+} 2\pi\sin(\varphi)K(t,\varphi) = \delta(\varphi) \tag{2.9}$$

is given by the function (see [3])

$$K(t,\varphi) = \frac{1}{4\pi}\sum_{n\in\mathbb{N}}(2n+1)\exp\left(-n(n+1)\sqrt{2t}\right)P_n^0(\cos\varphi). \tag{2.10}$$

Here $P_n^0(\cdot)$ is the associated Legendre polynomials of order zero, i.e.

$$P_n^0(x) = \frac{1}{2^n n!}\cdot\frac{d^n}{dx^n}\left[(x^2-1)^n\right]. \tag{2.11}$$

This fact implies the following

Proposition 2.2 *The transition density function of the Brownian motion X_t, $t\geq 0$, on S^2 with radius a it is given by the function*

$$p(t,\varphi) = \frac{1}{4\pi a^2}\sum_{n\in\mathbb{N}}(2n+1)\exp\left(-\frac{n(n+1)\sqrt{t}}{a}\right)P_n^0(\cos\varphi). \tag{2.12}$$

2.3 The Case of S^3

Proposition 2.3 *Let X_t, $t\geq 0$ be the Brownian motion on a 3-dimensional sphere S^3 of radius a. The transition density function $p(t,\varphi)$ of X_t is given by*

$$p(t,\varphi) = \frac{\exp\left(\frac{t}{2a^2}\right)}{(2\pi t)^{3/2}\sin\varphi}\sum_{n\in\mathbb{Z}}(\varphi+2n\pi)\exp\left(-\frac{(\varphi+2n\pi)^2a^2}{2t}\right),$$

where $\mathbb{Z}$ is the set of all integers. Equivalently

$$p(t,\varphi) = -\frac{i}{4\pi^2a^3\sin\varphi}\sum_{n\in\mathbb{Z}}n\exp\left(-\frac{t(n^2-1)}{2a^2}+i\varphi n\right)$$

and

$$p(t,\varphi) = \frac{1}{2\pi^2a^3\sin\varphi}\sum_{n\in\mathbb{N}}n\sin(n\varphi)\exp\left(-\frac{t(n^2-1)}{2a^2}\right).$$

Furthermore $p(t,\varphi)$ is analytic about $\varphi=0$ and $\varphi=\pi$. In fact

$$p(t,0) = \lim_{\varphi\to 0^+}p(t,\varphi) = \frac{1}{2\pi^2a^3}\sum_{n\in\mathbb{N}}n^2\exp\left(-\frac{t(n^2-1)}{2a^2}\right)$$

and

$$p(t,\pi) = \lim_{\varphi\to\pi^-} p(t,\varphi) = \frac{1}{2\pi^2 a^3}\sum_{n\in\mathbb{N}} n^2(-1)^n \exp\left(-\frac{t(n^2-1)}{2a^2}\right).$$

Reminder. The ϑ_3 function of Jacobi is

$$\vartheta_3(z,r) = 1 + 2\sum_{n=0}^{\infty} \exp\left(i\pi r n^2\right)\cos(2nz),$$

where $r \in \mathbb{C}$ with $\operatorname{Im}\{r\} > 0$. It follows that

$$p(t,\varphi) = -\frac{1}{4\pi^2 a^3 \sin\varphi}\exp\left(\frac{t}{2a^2}\right)\frac{\partial}{\partial\varphi}\vartheta_3\left(\frac{\varphi}{2},\frac{ti}{2a^2\pi}\right).$$

Sketch of Proof. First we will prove that $p(t,\varphi)$ satisfies the differential equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\Delta_3 p.$$

After that, we will show that

$$\lim_{t\to 0^+} 4\pi a^3 \sin^2(\varphi) p(t,\varphi) = \delta(\varphi).$$

For arbitrarily small $\epsilon > 0$, let

$$I_\epsilon = \int_0^\epsilon 4\pi a^3 \sin^2(\varphi) p(t,\varphi) d\varphi.$$

We have

$$\lim_{t\to 0^+} I_\epsilon = \lim_{t\to 0^+}\frac{4\pi a^3 \exp\left(\frac{t}{2a^2}\right)}{(2t\pi)^{3/2}}\left(\int_0^\epsilon \varphi\sin(\varphi)\exp\left(-\frac{\varphi^2 a^2}{2t}\right)d\varphi\right.$$
$$\left. + \sum_{n\in\mathbb{Z}^*}\int_0^\epsilon (\varphi+2n\pi)\sin(\varphi)\exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right)d\varphi\right),$$

where $\mathbb{Z}^* = \mathbb{Z} - \{0\}$. However

$$\left|\sum_{n\in\mathbb{Z}^*}\int_0^\epsilon (\varphi+2n\pi)\sin(\varphi)\exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right)d\varphi\right| \le \sum_{n\in\mathbb{Z}^*}\int_0^\epsilon (2|n|+1)\pi\exp\left(-\frac{n^2\pi^2 a^2}{2t}\right)d\varphi$$

and

$$\sum_{n\in\mathbb{Z}^*}\int_0^\epsilon (2|n|+1)\pi\exp\left(-\frac{n^2\pi^2 a^2}{2t}\right)d\varphi = \epsilon\sum_{n\in\mathbb{Z}^*}(2|n|+1)\pi\exp\left(-\frac{n^2\pi^2 a^2}{2t}\right),$$

which converges to 0 as $t \to 0^+$, by Lebesgue's Dominated Convergence Theorem. Therefore

$$\lim_{t\to 0^+} I_\epsilon = \lim_{t\to 0^+}\frac{4\pi a^3 \exp\left(\frac{t}{2a^2}\right)}{(2t\pi)^{3/2}}\int_0^\epsilon \varphi\sin\varphi\exp\left(-\frac{\varphi^2 a^2}{2t}\right)d\varphi.$$

By using the Laplace method for integrals [1]

$$\int_0^{\epsilon} \varphi \sin(\varphi) \exp\left(-\frac{\varphi^2 a^2}{2t}\right) d\varphi \sim \int_0^{\epsilon} \varphi^2 \exp\left(-\frac{\varphi^2 a^2}{2t}\right) d\varphi \sim \int_0^{\infty} \varphi^2 \exp\left(-\frac{\varphi^2 a^2}{2t}\right) d\varphi,$$

as $t \to 0^+$. Here $A \sim B$ means that $\frac{A}{B} \to 1$. Hence

$$\lim_{t\to 0^+} I_\epsilon = \lim_{t\to 0^+} \frac{4\pi a^3 \exp\left(\frac{t}{2a^2}\right)}{(2t\pi)^{3/2}} \int_0^{\infty} \varphi^2 \exp\left(-\frac{\varphi^2 a^2}{2t}\right) d\varphi,$$

or, for

$$u = \frac{\varphi a}{\sqrt{t}}$$

$$\lim_{t\to 0^+} I_\epsilon = \lim_{t\to 0^+} 2\exp\left(\frac{t}{2a^2}\right) \int_0^{\infty} \frac{u^2}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$$

i.e.

$$\lim_{t\to 0^+} I_\epsilon = 1. \tag{2.13}$$

Furthermore, for every $t > 0$, we have

$$I = \int_0^{\pi} 4\pi a^3 \sin^2(\varphi) p(t,\varphi) d\varphi, \tag{2.14}$$

hence,

$$I = \frac{4\pi a^3 \exp\left(\frac{t}{2a^2}\right)}{(2t\pi)^{3/2}} \int_0^{\pi} \sum_{n\in\mathbb{Z}} (\varphi + 2n\pi) \sin(\varphi) \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right) d\varphi.$$

The series

$$\sum_{n\in\mathbb{Z}} (\varphi + 2n\pi) \sin(\varphi) \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right)$$

converges uniformly on $[0,\pi]$ for every $t > 0$, because

$$\left|(\varphi + 2n\pi) \sin(\varphi) \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right)\right| \leq 2|n|\pi \exp\left(-\frac{n^2\pi^2 a^2}{2t}\right)$$

and the series

$$\sum_{n\in\mathbb{Z}} M_n,$$

where

$$M_n = 2|n|\pi \exp\left(-\frac{n^2\pi^2 a^2}{2t}\right)$$

converges. Therefore (2.14), implies that

$$I = \frac{4\pi a^3 \exp\left(\frac{t}{2a^2}\right)}{(2t\pi)^{3/2}} \sum_{n\in\mathbb{Z}} \int_0^{\pi} (\varphi + 2n\pi) \sin(\varphi) \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right) d\varphi.$$

Hence

$$I = \frac{a \exp\left(\frac{t}{2a^2}\right)}{\sqrt{2t\pi}} \sum_{n\in\mathbb{Z}} \int_0^{\pi} [\exp(i\varphi) + \exp(-i\varphi)] \exp\left(-\frac{(\varphi+2n\pi)^2 a^2}{2t}\right) d\varphi. \tag{2.15}$$

Let $u = \varphi + 2n\pi$, then (2.15) gives

$$I = \frac{a}{\sqrt{2t\pi}} \left(\frac{\sqrt{2t\pi}}{2a} + \frac{\sqrt{2t\pi}}{2a} \right) = 1$$

for every $t > 0$. In particular

$$\lim_{t \to 0^+} \int_0^\pi 4\pi a^3 \sin^2(\varphi) p(t,\varphi) d\varphi = 1. \tag{2.16}$$

From (2.13) and (2.16) we have that that

$$\lim_{t \to 0^+} 4\pi a^3 \sin^2(\varphi) p(t,\varphi) d\varphi = \delta(\varphi)$$

and this complete the proof. ■

3 Stochastic Differential Equation (SDE) of X_t in Local Coordinates

In spherical coordinates:

The Brownian motion on S^1 satisfies the SDE

$$dX_t = \frac{1}{a} dB_t.$$

The Brownian motion on S^2 satisfies the SDE

$$dX_t = \left(0, \frac{\cos\varphi}{2a^2 \sin\varphi}\right) dt + \begin{bmatrix} \frac{1}{a\sin\varphi} & 0 \\ 0 & \frac{1}{a} \end{bmatrix} \begin{bmatrix} dB_1(t) \\ dB_2(t) \end{bmatrix}.$$

The Brownian motion on S^3 satisfies the SDE

$$dX_t = \left(0, \frac{\cos\theta_2}{2a^2 \sin\theta_2 \sin^2\varphi}, \frac{\cos\varphi}{a^2 \sin\varphi}\right) dt + \begin{bmatrix} \frac{1}{a\sin\theta_2\sin\varphi} & 0 & 0 \\ 0 & \frac{1}{a\sin\varphi} & 0 \\ 0 & 0 & \frac{1}{a} \end{bmatrix} \begin{bmatrix} dB_1(t) \\ dB_2(t) \\ dB_3(t) \end{bmatrix}.$$

In stereographic projection coordinates: The Brownian motion on S^2 satisfies the SDE

$$dX_t = \frac{\xi_1^2 + \xi_2^2 + 4a^2}{4a^2} \begin{bmatrix} dB_1(t) \\ dB_2(t) \end{bmatrix}.$$

The Brownian motion on S^3 satisfies the SDE

$$dX_t = -\frac{\left(\xi_1^2 + \xi_2^2 + \xi_3^2 + 4a^2\right)}{16a^4} (\xi_1, \xi_2, \xi_3)\, dt + \frac{\left(\xi_1^2 + \xi_2^2 + \xi_3^2 + 4a^2\right)}{4a^2} \begin{bmatrix} dB_1(t) \\ dB_2(t) \\ dB_3(t) \end{bmatrix}.$$

The Brownian motion on $S^n, n \geq 2$ satisfies the SDE

$$dX_t = (2-n) \frac{\left(\xi_1^2 + \cdots + \xi_n^2 + 4a^2\right)}{16a^4} (\xi_1, \cdots, \xi_n)\, dt + \frac{\left(\xi_1^2 + \ldots + \xi_n^2 + 4a^2\right)}{4a^2} \begin{bmatrix} dB_1(t) \\ dB_2(t) \\ \vdots \\ dB_n(t) \end{bmatrix}.$$

4 Expectations of exit times

Let X_t be the Brownian motion in S^n and $D \subset S^n$. The random variable

$$T = \inf\{t \geq 0 |\ X_t \notin D\}$$

is called the (first) exit time D.
Reminder. If

$$u(x) = E^x[T],$$

(here the superscript x indicates that $X_0 = x$) then $u(x)$ satisfies

$$\frac{1}{2}\Delta_n u = -1$$

$$u|_{\partial D} = 0$$

Let $\varphi_1, \varphi_2 \in [0, 2\pi)$, $\varphi_1 < \varphi_2$. Consider the set

$$D = (\varphi_1, \varphi_2)\,.$$

If X_t is the Brownian motion on S^1 starting at the point $\varphi \in D$, then

$$E^{\varphi}[T] = a^2\,(\varphi - \varphi_1)\,(\varphi_2 - \varphi)$$

Let $\varphi_0 \in (0, \pi)$ be fixed. We consider the set D in S^n, $n \geq 2$, such that

$$D = \{\,(\theta_1, \ldots, \theta_{n-1}, \varphi) |\ \varphi \in [0, \varphi_0)\}\,.$$

If X_t is the Brownian motion on S^n starting at the point

$$A = (\theta_1, \ldots, \theta_{n-1}, \varphi) \in D$$

then

$$E^A[T] = u(\varphi) = 2a^2 \int_{\varphi}^{\varphi_0} \frac{\int_0^x (\sin\omega)^{n-1} d\omega}{(\sin x)^{n-1}} dx$$

Notice that $u(\varphi)$ is an elementary function.

For $\underline{n = 2}$ we obtain

$$E^A[T] = 2a^2 \ln\left(\frac{1 + \cos\varphi}{1 + \cos\varphi_0}\right).$$

For $\underline{n = 3}$ we obtain

$$E^A[T] = a^2\,(\varphi \cot\varphi - \varphi_0 \cot\varphi_0)\,.$$

Let $\varphi_1, \varphi_2 \in (0, \pi)$, $\varphi_1 < \varphi_2$. Consider the set D in $S^n, n \geq 2$,

$$D = \{\,(\theta_1, \ldots, \theta_{n-1}, \varphi) |\ \varphi \in (\varphi_1, \varphi_2)\}\,.$$

If X_t is the Brownian motion on S^n starting at the point

$$A = (\theta_1, \ldots, \theta_{n-1}, \varphi) \in D$$

then

$$E^A[T] = 2a^2 \left[\int_{\varphi}^{\varphi_1} \frac{\int_0^x (\sin\omega)^{n-1} d\omega}{(\sin x)^{n-1}} dx + \frac{\int_{\varphi_1}^{\varphi_2} \frac{\int_0^x (\sin\omega)^{n-1} d\omega}{(\sin x)^{n-1}} dx}{\int_{\varphi 1}^{\varphi_2} \frac{1}{(\sin x)^{n-1}} dx} \cdot \int_{\varphi_1}^{\varphi} \frac{1}{(\sin x)^{n-1}} dx \right].$$

For $\underline{n=2}$ we obtain

$$E^A[T] = \frac{4a^2}{\ln\left(\frac{\tan(\varphi_2/2)}{\tan(\varphi_1/2)}\right)} \left[\ln\left(\frac{\cos(\varphi_1/2)}{\cos(\varphi_2/2)}\right) \ln\left(\frac{\sin(\varphi/2)}{\sin(\varphi_1/2)}\right) - \ln\left(\frac{\cos(\varphi_1/2)}{\cos(\varphi/2)}\right) \ln\left(\frac{\sin(\varphi_2/2)}{\sin(\varphi_1/2)}\right) \right].$$

For $\underline{n=3}$ we obtain

$$E^A[T] = \frac{a^2 \left[(\varphi - \varphi_1) \cot\varphi \cot\varphi_1 + (\varphi_1 - \varphi_2) \cot\varphi_1 \cot\varphi_2 + (\varphi_2 - \varphi) \cot\varphi_2 \cot\varphi \right]}{\cot\varphi_1 - \cot\varphi_2}.$$

Notice that the formulas for $n = 2$ and $n = 3$ are quite different.

5 Hitting Probabilities

Let X_t be the Brownian motion in S^n, $D \subset S^n$, and T its exit time.
Reminder. Let $\Gamma \subset D$ and

$$u(x) = P^x\{X_T \in \Gamma\},$$

then $u(x)$ satisfies

$$\Delta_n u = 0$$

$$u|_{\Gamma} = 1, \qquad u|_{\partial D \setminus \Gamma} = 0$$

Consider the subset $D = (\varphi_1, \varphi_2)$ of S^1, $0 < \varphi_1 < \varphi_2 < 2\pi$. If $\Gamma_1 = \{\varphi_1\}$, then

$$P^{\varphi}\{X_T \in \Gamma_1\} = \frac{\varphi_2 - \varphi}{\varphi_2 - \varphi_1}$$

Let $\varphi_1, \varphi_2 \in (0, \pi)$, $\varphi_1 < \varphi_2$. Consider the set D in $S^n, n \geq 2$,

$$D = \{ (\theta_1, \ldots, \theta_{n-1}, \varphi) |\ \varphi \in (\varphi_1, \varphi_2)\}$$

and the point

$$A = (\theta_1, \ldots, \theta_{n-1}, \varphi) \in D.$$

If $\Gamma_1 = \{(\theta_1, \ldots, \theta_{n-1}, \varphi_1)\}$, then

$$P^A\{X_T \in \Gamma_1\} = \frac{\int_{\varphi}^{\varphi_2} \frac{1}{(\sin x)^{n-1}} dx}{\int_{\varphi_1}^{\varphi_2} \frac{1}{(\sin x)^{n-1}} dx}.$$

For $\underline{n=2}$ we obtain

$$P^A\{X_T \in \Gamma_1\} = \frac{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi}{2}\right)}\right)}{\ln\left(\frac{\tan\left(\frac{\varphi_2}{2}\right)}{\tan\left(\frac{\varphi_1}{2}\right)}\right)}.$$

For <u>$n = 3$</u> we obtain

$$P^A\{X_T \in \Gamma_1\} = \frac{\cot\varphi - \cot\varphi_2}{\cot\varphi_1 - \cot\varphi_2} = \frac{\sin\varphi_1 \sin(\varphi_2 - \varphi)}{\sin\varphi \sin(\varphi_2 - \varphi_1)}.$$

Let D be domain on S^2 whose stereographic coordinate description is

$$D = \{(\xi_1, \xi_2) |\ b < \xi_2 < c\},$$

i.e. D is the domain bounded by two circles passing through the north pole. If $A = (\xi_1, \xi_2) \in D$ and

$$\Gamma_1 = \{(\xi_1, b) |\ \xi_1 \in \mathbb{R}\},$$

then

$$P^A\{X_T \in \Gamma_1\} = \frac{c - \xi_2}{c - b}. \tag{5.1}$$

6 The Moment Generating Function of T

Reminder. Assume that $\lambda > -\lambda_1/2$, where λ_1 is the first Dirichlet eigenvalue of $D \subset S^n$. If

$$u(x) = E^x[e^{-\lambda T}],$$

then $u(x)$ satisfies

$$\frac{1}{2}\Delta_n u = \lambda u$$
$$u|_{\partial D} = 1$$

Suppose $D \subset S^1$ is the domain

$$D = (\varphi_1, \varphi_2), \qquad 0 \leq \varphi_1 < \varphi_2 < 2\pi.$$

Then, for $\varphi \in (\varphi_1, \varphi_2)$

$$E^\varphi[e^{-\lambda T}] = \frac{\sinh\left(a\sqrt{2\lambda}(\varphi_2 - \varphi)\right) + \sinh\left(a\sqrt{2\lambda}(\varphi - \varphi_1)\right)}{\sinh\left(a\sqrt{2\lambda}(\varphi_2 - \varphi_1)\right)}$$

provided

$$\lambda > -\frac{\pi^2}{2a^2(\varphi_2 - \varphi_1)^2}$$

Let X_t be the Brownian motion on S^2 starting at the point

$$A = (\theta, \varphi) \in D,$$

where D is the domain

$$D = \{(\theta, \varphi) |\ \theta \in [0, 2\pi), \quad \text{and} \quad \varphi \in [0, \varphi_0)\}.$$

Then

$$E^A[\exp(-\lambda T)] = \frac{P_\nu(\cos\varphi)}{P_\nu(\cos\varphi_0)},$$

where ν is such that $\nu(\nu+1) = -2a^2\lambda$ and $P_\nu(\cdot)$ is the Legendre function

$$P_\nu(z) = P_{-\nu-1}(z) = \frac{1}{\pi}\int_0^\pi \left(z + \sqrt{z^2-1}\,\cos\phi\right)^\nu d\phi,$$

where the multiple-valued function $\left(z + \sqrt{z^2-1}\,\cos\phi\right)^\nu$ is to be determined in such a way that for $\phi = \pi/2$ it is equal to (the principal value of) z^ν (which is, in particular, real for positive z and real ν).

Let X_t be the Brownian motion on S^n starting at the point $A \in D$, where

$$D = \{(\theta_1, \dots, \theta_{n-1}, \varphi) |\ \theta_1 \in [0, 2\pi), \theta_i \in [0, \pi] \quad \text{for} \quad i = 2, \dots, n-1 \quad \text{and} \quad \varphi \in [0, \varphi_0)\}.$$

Then

$$E^A[\exp(-\lambda T)] = \frac{(\sin\varphi)^{1-\frac{n}{2}}\, P_\nu^\mu(\cos\varphi)}{(\sin\varphi_0)^{1-\frac{n}{2}}\, P_\nu^\mu(\cos\varphi_0)}, \tag{6.1}$$

where

$$\nu = \frac{1}{2}\left(\sqrt{(n-1)^2 - 8a^2\lambda} - 1\right) \quad \text{and} \quad \mu = \frac{1}{2}(n-2).$$

The function $P_\nu^\mu(\cdot)$ is the associated Legendre function

$$P_\nu^\mu(z) = \frac{1}{\Gamma(-\nu)\Gamma(\nu+1)}\left(\frac{1+z}{1-z}\right)^{\mu/2} \sum_{n=0}^{\infty} \frac{\Gamma(n-\nu)\Gamma(n+\nu+1)}{\Gamma(n+1-\mu)n!}\left(\frac{1-z}{z}\right)^n.$$

Here $\Gamma(\cdot)$ denotes the Gamma function.

7 The Reflection Principle

We will discuss the reflection principle on S^2. Everything extends easily to S^n.

Notation. For every point $A = (x_1, x_2, x_3) \in S^2$ we denote by $\hat{A}$ the symmetric of A with respect to the x_1x_2-plane. In other words

$$\hat{A} = (x_1, x_2, -x_3) \in S^2$$

Theorem 7.1 *Let X_t, $t \geq 0$, be the Brownian motion on S^2 starting at the point $A = (\theta, \varphi)$ (in spherical coordinates). We assume that $A \in D$, where D is the lower hemisphere, i.e.*

$$D = \{(\theta, \varphi) |\ \theta \in [0, 2\pi) \quad \textit{and} \quad \varphi \in (\pi/2, \pi]\}$$

If

$$T = \inf\{t \geq 0 |\ X_t \notin D\},$$

then

$$P^A\{T < t\} = 2P^A\{X_t \notin D\}.$$

Sketch of Proof.

$$P^A\{T < t\} = P^A\{T < t, X_t \notin D\} + P^A\{T < t, X_t \in D\}.$$

However, if $X_t \notin D$, then, of course, $T < t$. Thus

$$P^A\{T < t, X_t \notin D\} = P^A\{X_t \notin D\}.$$

On the other hand, if we set

$$\tilde{X}_t = \begin{cases} X_t, & \text{if} \quad t \leq T \\ \hat{X}_t, & \text{if} \quad t > T \end{cases}$$

then, by the strong Markov property of X_t

$$P^A\{T < t, X_t \in D\} = P^A\left\{T < t, \tilde{X}_t \in D\right\},$$

but $\tilde{X}_t \in D$ if and only if $X_t \notin D$. Hence,

$$P^A\left\{T < t, \tilde{X}_t \in D\right\} = P^A\{T < t, X_t \notin D\} = P^A\{X_t \notin D\}$$

and

$$P^A\{T < t, X_t \in D\} = P^A\{X_t \notin D\}.$$

Therefore $P^A\{T < t\} = 2P^A\{X_t \notin D\}$. ■

7.1 Applications of the Reflection Principle

The reflection principle can help to calculate the distribution functions of certain exit times.
Let X_t be the Brownian motion on S^2 starting at the south pole S, where $S = (0, \pi)$ in spherical coordinates. If D is the lower hemisphere and T its exit time, then

$$P^S\{T < t\} = 1 - \sum_{n=0}^{\infty}(-1)^n \exp\left(-\frac{(2n+1)^2\sqrt{t}}{a}\right) \cdot \frac{(2n)!(2n+3)}{2^{2n+1}n!}$$

<u>The case of S^1</u>:
Let X_t be the Brownian motion on S^1 starting at $\varphi \in D = (\pi, 2\pi)$. If T is the exit time of D, then

$$P^{\varphi}\{T < t\} = 1 + \frac{4}{\pi}\sum_{n \text{ odd}} \frac{1}{n}\exp\left(-\frac{n^2 t}{2a^2}\right)\sin(n\varphi)$$

<u>The case of S^3</u>:
Let X_t be the Brownian motion on S^3 starting at the south pole S, where $S = (0, 0, \pi)$ in spherical coordinates. If D is the lower hemisphere, namely

$$D = \left\{(\theta_1, \theta_2, \varphi) \in S^3 \;\middle|\; \theta_1 \in [0, 2\pi), \theta_2 \in [0, \pi], \varphi \in (\pi/2, \pi]\right\}$$

and T the exit time of D, then

$$P^S\{T < t\} = 1 + \frac{16}{\pi}\sum_{n=1}^{\infty}(-1)^n n^2 \exp\left(-\frac{(4n^2-1)t}{2a^2}\right).$$

Acknowledgment. This work was partially supported by a Π.Ε.Β.Ε. Grant of the National Technical University of Athens.

References

[1] Bender M. and Orszag S., *Advanced Mathematical Methods for Scientists and Engineers* McGraw-Hill Book Company, USA (1978)

[2] Camporesi R., Harmonic Analysis and Propagators on Homogeneous Spaces, *Phisycs Reports* **196** (7) 1–134 (1990).

[3] Chung M.K., Heat Kernel Smoothing On Unit Sphere *in Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 992–995 (2006).

[4] Dodziuk J., Maximum Principle for Parabolic Inequalities and the Heat Flow on Open Manifolds, *Indiana Oniv. Math. J.* **32** no.5 115–142 (1983).

[5] Dynkin E.B., *Markov Processes vol 2* Springer, Berlin (1965).

[6] Grigor'yan A. and Saloff-Coste L., Hitting Probabilities for Brownian Motion on Riemannian Manifolds, *J. Math. Pures et Appl.* **81** 115–142 (2002) Company, Malabar, Florida, (1987).

[7] Hsu E.P., A Brief Introduction to Brownian Motion On A Riemannian Manifold, *Lecture notes.*

[8] John F., *Partial Differential Equations* Springer, USA (1982)

[9] Karatzas I. and Shreve S.E., *Brownian Motion and Stochastic Calculus* Springer, USA (1991).

[10] Klebaner F.C., *Introduction to Stochastic Calculus with Applications* Imerial College Press, Melbourne (2004).

[11] Oksendal B., *Stochastic Differential Equations.* Springer-Verlag (1995).