Geometric Model of Roots of Stochastic Matrices

by

Yelyzaveta Chetina

A Major Paper
Submitted to the faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2018

Geometric Model of Roots of Stochastic Matrices

by

Yelyzaveta Chetina

APPROVED BY:

A. A. Hussein
Department of Mathematics and Statistics

M. Hlynka, Co-Advisor
Department of Mathematics and Statistics

P. Brill, Co-Advisor
Department of Mathematics and Statistics

April 17, 2018

Declaration of Originality

I hereby certify that I am the sole author of this major paper and that no part of this major paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my major paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such materials in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my major paper, including any final revisions, as approved by my committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

Abstract

In this paper we examine the conditions under which discrete-time homogenous Markov transition matrices have probability roots. A method based on geometric interpretation of 2×2 Markov matrices is used to find regions within the unit square corresponding to probability matrices with zero, single or multiple probability roots.

Acknowledgement

First and foremost, I express my sincere gratitude to my advisor, Dr. Hlynka, without whose guidance and encouragement this paper would not be completed. His door was always open for me and I have received many valuable suggestions from him. I would also like to show my appreciation and thank my second advisor, Dr. Brill, for his kind support. Further, I would like to thank Dr. Hussein as the department reader. And last but not least, I thank my supportive and loving mother who is my inspiration.

Contents

Chapter 1

Literature Review

First, we revisit the definition of a discrete-time Markov chain. According to Ross [10], a discrete-time Markov chain is a stochastic process $\{X_n, n = 0, 1, 2, ...\}$ defined over discrete state space which satisfies the property: "given the past states $X_0, X_1, ..., X_{n-1}$ and the present state X_n, the conditional distribution of X_{n+1} is independent of the past states and depends only on the present state." This property is referred to as the Markov property and it is also known as "memoryless" property of a Markov process. The value of $P(X_{n+1} = j|X_n = i)$ is denoted as p_{ij} which corresponds to the transition probability of going from state i to state j. Therefore, a Markov chain with n states can be represented as a $n \times n$ matrix P

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{bmatrix}$$

where $p_{ij} \geq 0$ for all states i, j and $\sum_{j=1}^{n} p_{ij} = 1$ for every row $i = 0, 1, ..., n$. Such a matrix is referred to as a Markov transition matrix, or a probability matrix, or a stochastic matrix. These terms will be used interchangeably throughout this paper.

Next, we consider the case where a discrete-time Markov chain is observed on time intervals twice the duration of the original Markov chain stepsize. This implies that there are Markov transition matrices whose square roots are also transition matrices. Guerry [2012] defines a probability square root A of a transition matrix P as a probability matrix which provides a Markov chain with time unit $\frac{1}{2}$ that is also "compatible with P". According to Guerry, whether or not a given matrix P can be represented as A^2 for some probability matrix A is referred to as the embedding problem for discrete time Markov chains. Embeddability is the focal point of this paper. Both algebraic and geometric methods were employed previously to define conditions for embeddability of Markov chains.

Brill and Hlynka [2002] were the first to look at a geometric interpretation for 2×2 Markov transition matrices. They obtained a geometric method for finding powers and roots of transition matrices, and discussed the convergence of powers of a transition matrix to a common limiting probability matrix. Brill and Hlynka also included a geometric method of multiplying 2×2 transition matrices.

He and Gunn [2003] used the characteristic polynomial of a matrix to explicitly find all real root matrices of 2×2 and 3×3 stochastic matrices, and were the first to do this. Additionally, they included some numerical methods for computing square root matrices.

Higham and Lin [2011] used the theory of matrix functions for their analysis.They provided some necessary conditions for existence of p^{th} stochastic root of a stochastic matrix. Higham and Lin's paper included a geometric representation of the sets of all possible eigenvalues of 3×3 and 4×4 stochastic matrices along with powers of these sets.

Guerry [2013] provided necessary and sufficient conditions for embeddability of discrete-time two-state Markov chains. Her approach was analytic where she examined properties of row-normalized matrices. Guerry discussed the conditions for uniqueness of square roots, the concept of approximate probability roots and a method for identifying a context-sensitive root in the case of nonuniqueness. In her paper from 2012, Guerry inspected the embeddability problem in discrete-time state-wise monotone Markov chains where the point of interest was not only the existence of a square root but, further, the conditions which accommodate the properties of the original probability matrix to be reflected in its square root. Finally, in a more recent paper from 2017, Guerry extended her work to non-diagonalizable three-state transition matrices and all configurations of the signs of eigenvalues of these matrices. She stated the embedding conditions in terms of the projections and the spectral decompositions of the transition matrices.

Hence, the geometric approach first used by Brill and Hlynka indeed has no extensions in the literature which gives this paper the material to build on without being repetitive of work done by other authors.

Chapter 2

Motivation and Formulation of the Problem

Markov chain models have been commonly used in finance, management, social research, healthcare, environmental forecasting, etc. as they are a great mathematical modelling tool in the cases where the Markov property (i.e. ”memoryless” property) assumption is appropriate. We present a few applications from the literature.

Guerry [2013] refers to a manpower planning model. The model includes “internal transitions (e.g. promotions), outgoing flows (i.e. wastage) as well as incoming flows (i.e. recruitments).” Here, the transition probabilities might be such that the expected number of members in each personnel category can be computed annually. Oftentimes, there is an interest in the number of members after half a year period which in mathematical terms corresponds to a stochastic square root of the original stochastic matrix.

He and Gunn [2003] mention a stochastic model which describes the state of weather conditions at an airport. The model is constructed in a way that hourly measurements can be obtained. Given that the shorter time periods such as half an hour or fifteen minutes are frequently of interest, we seek to obtain two probability matrices which when raised to the second and fourth powers respectively will result in the original probability matrix, hence, raising an issue of embeddability.

Biritwun and Odoom [1995] use a Markov chain to model health of infants. The diagnostic records were comprised of data collected monthly. Ultimately, probabilities of state transitions for the fifteen day period were of interest which corresponds to a square root of the original Markov transition matrix.

Malik and Thomas [2012] use Markov chain based modelling to analyze credit risk of portfolios of consumer loans. The transition probabilities are based on behavioural scores on a monthly basis. Although the authors do not mention the possibility of embedding, it naturally arises in a setting where different time periods are of interest.

Given the above, we can conclude that the discrete-time Markov chain embedding problem is of high interest since it has many applications in various fields. This, therefore, is a substantial motivator for this paper.

Finally, we conclude this chapter with a formulation of the problem. We are interested in employing the geometric method first derived by Brill and Hlynka [2002] to establish conditions under which 2×2 Markov matrices have multiple, single or zero probability roots.

Chapter 3

Geometric Approach

3.1 Powers and roots of Markov matrices

We begin by summarizing results from Brill and Hlynka [2002]. Let $P = [p_{ij}]$ be an irreducible aperiodic Markov transition matrix. Let $\underline{\pi}$ be its limiting probability vector which satisfies

$$\underline{\pi} = \underline{\pi}P \text{ where } \sum_i \underline{\pi}_i = 1, 0 \leq \underline{\pi}_i \leq 1.$$

Define a 2×2 transition matrix as following

$$P = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$$

where $0 \leq x, y \leq 1$.

We can represent P as a point $[x, y]$ in the unit square. Next, set $\underline{\pi} = (a, 1-a)$. Then, from $\underline{\pi} = \underline{\pi}P$ we get

$$a = ax + (1-a)y$$
$$1 - a = a(1-x) + (1-a)(1-y)$$

The set of equations has a straight line $y = -\frac{a}{1-a}(x-1)$ as the solution which represents the collection of all transition matrices with the given limiting vector (i.e. all positive integer powers of P) and it can be represented as $[a, a]$ inside the unit square. Recall that if $\underline{\pi}$ is a limiting probability vector of a matrix P, then $\underline{\pi} = \underline{\pi}P$, which implies the following

$$\underline{\pi}P^n = \underline{\pi}P^{n-1} = ... = \underline{\pi}P = \underline{\pi} \text{ for all } n > 1.$$

Hence, all positive integer powers of P indeed have the same limiting probability vector.

According to Brill and Hlynka [2002], for any given matrix P i.e. a point $[x, y]$ within a unit square, we can instantly find its limiting vector which is the intersection of the line through $[x, y]$ and $[1, 0]$ with the line through $[a, a]$ and the origin (i.e. $y = x$). Conversely, given a limiting vector $(a, 1-a)$ i.e. a point $[a, a]$ within a unite square, we can instantly find the collection of all matrices with such limiting vector which is the line through $[1, 0]$ and $[a, a]$. Below three lines are plotted corresponding to collections of matrices with three different limiting vectors $\underline{\pi}_1 = (\frac{1}{4}, \frac{3}{4})$, $\underline{\pi}_2 = (\frac{1}{2}, \frac{1}{2})$ and $\underline{\pi}_3 = (\frac{3}{4}, \frac{1}{4})$.

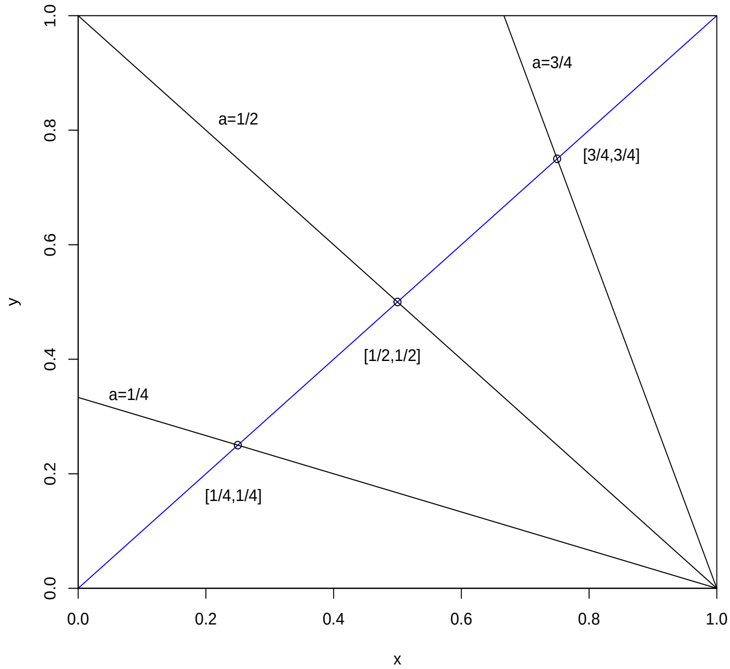


Figure 3.1

Next, given a probability matrix P, we can compute P^n. We denote P^n as a point $[x_n, y_n]$ within the unit square for all $n \geq 1$. The result of the following theorem was included in Brill and Hlynka's paper [2002] without proof. Hence, we provide the proof here.

Theorem 3.1.1. *Given a probability matrix* $P = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$, *the [1,1] entry of* P^n *is*

$$x_n = \tfrac{1}{1-x+y}(y + (1-x)(x-y)^n)$$

Proof. We can employ the diagonalization method to find P^n. First, we find eigenvalues of P. From the characteristic equation below

$$
\begin{aligned}
det(P - \lambda I) &= det \begin{bmatrix} x - \lambda & 1 - x \\ y & 1 - y - \lambda \end{bmatrix} \\
&= (x - \lambda)(1 - y - \lambda) - y(1 - x) \\
&= x - \lambda x - y + \lambda y - \lambda + \lambda^2 \\
&= x(1 - \lambda) - y(1 - \lambda) - \lambda(1 - \lambda) \\
&= (x - y - \lambda)(1 - \lambda) \\
&= 0
\end{aligned}
$$

we get $\lambda_1 = 1$ and $\lambda_2 = x - y$. Therefore, for some invertible matrix U, we get

$$P = U^{-1} \begin{bmatrix} 1 & 0 \\ 0 & x - y \end{bmatrix} U$$

which implies that

$$P^n = U^{-1} \begin{bmatrix} 1^n & 0 \\ 0 & (x - y)^n \end{bmatrix} U$$

It follows that for some constants q and k we have

$$x_n = q + k(x - y)^n$$

Note that since $x_0 = 1$ and $x_1 = x$, we obtain a set of two equations with two unknowns

$$
\begin{aligned}
q + k &= 1 \\
q + k(x - y) &= x
\end{aligned}
$$

which yields $q = \frac{y}{1-x-y}$ and $k = \frac{1-x}{1-x-y}$. Therefore, we get

$$x_n = \tfrac{1}{1-x+y}(y + (1 - x)(x - y)^n)$$

□

Note that y_n can be found by interchanging y and $(1-x)$ in the above equation. Alternatively, since the point $[x_n, y_n]$ is on the line through $[x, y]$ and $[1, 0]$, we can obtain y_n by plugging x_n into the equation $y_n = \frac{y}{x-1}(x_n - 1)$.

For every $n > 0$, $x_n = \frac{1}{1-x+y}(y + (1-x)(x-y)^n)$ can be used to locate P^n inside the unit square. For example, take

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

so that $[x, y] = [0.7, 0.2]$. Then, to obtain P^2, we calculate $x_2 = \frac{1}{1-x+y}(y + (1-x)(x-y)^2)$ and $y_2 = \frac{0.2}{0.7-1}(x_2 - 1)$ which results in $[0.55, 0.3]$ or

$$P^2 = \begin{bmatrix} 0.55 & 0.45 \\ 0.3 & 0.7 \end{bmatrix}$$

Repeating the process one more time we obtain $[x_4, y_4] = [0.4375, 0.375]$ or

$$P^4 = \begin{bmatrix} 0.4375 & 0.5625 \\ 0.375 & 0.625 \end{bmatrix}$$

The obtained matrices are labeled on the graph below

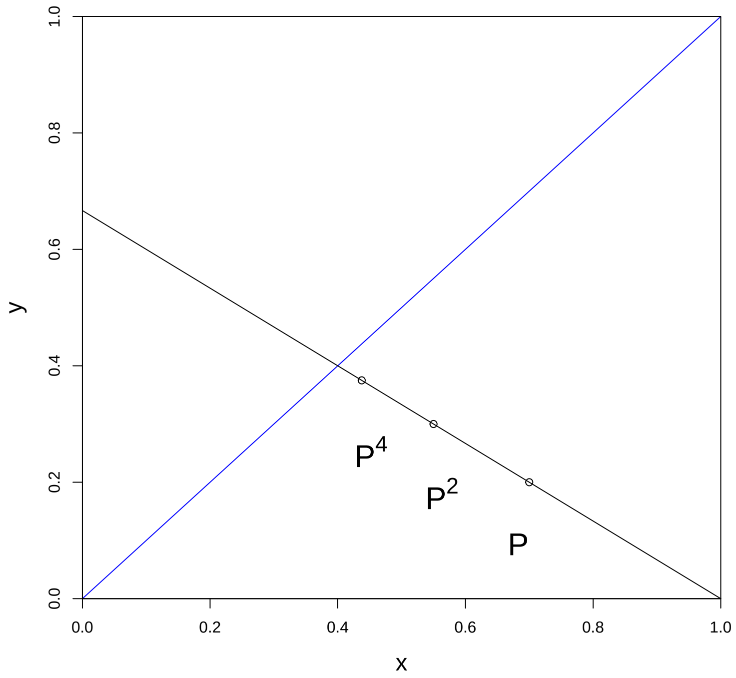


Figure 3.2

All three points lie on the same line and share the same limiting vector which is indicated as the point of intersection of the two lines in the graph.

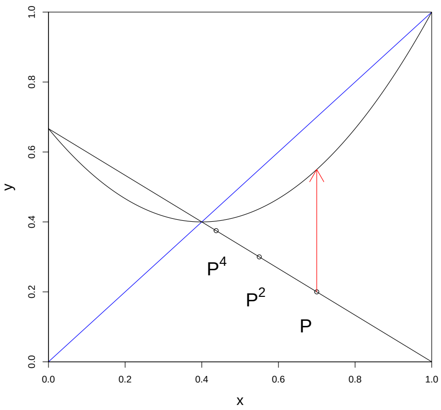
Note that we can rewrite x_n in terms of x and a, substituting $y = \frac{-a}{1-a}(x-1)$. Therefore, we get

$$x_n = (x-a)^n(1-a)^{1-n} + a.$$

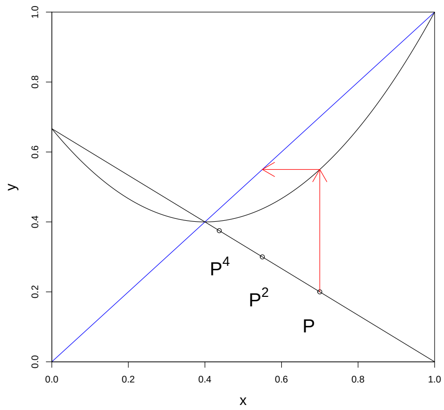
Brill and Hlynka denote this function as $f_n(x)$. Add the curve

$$f_2(x) = (x-a)^2(1-a)^{-1} + a$$

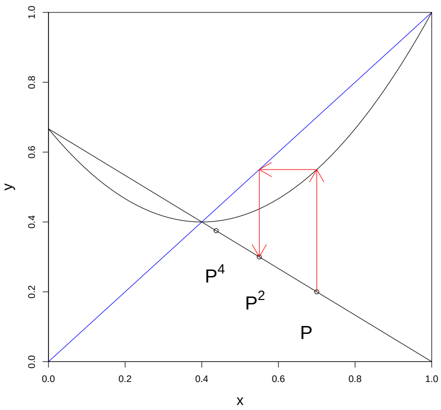
which is a parabola, to the graph and see the step by step geometric approach to finding P^2 and P^4.



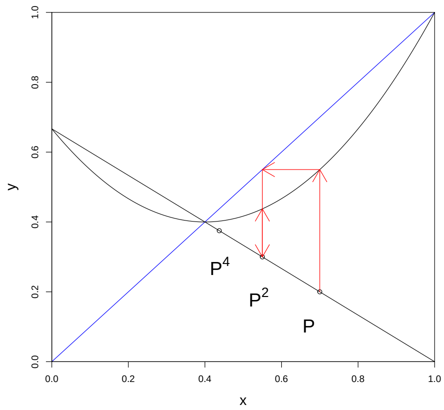
(a) sloped line to curve



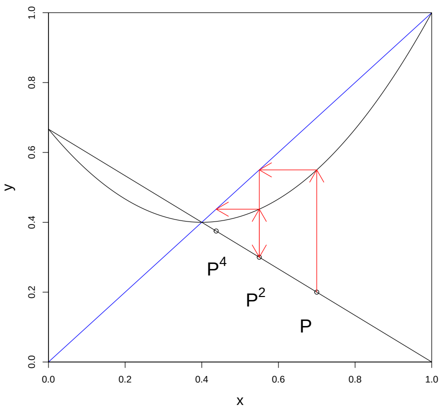
(b) curve to diagonal



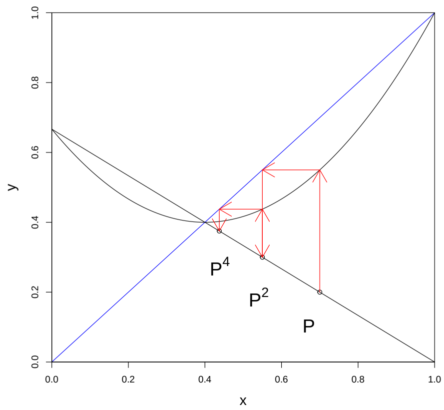
(c) diagonal to sloped line



(d) sloped line to curve



(e) curve to diagonal

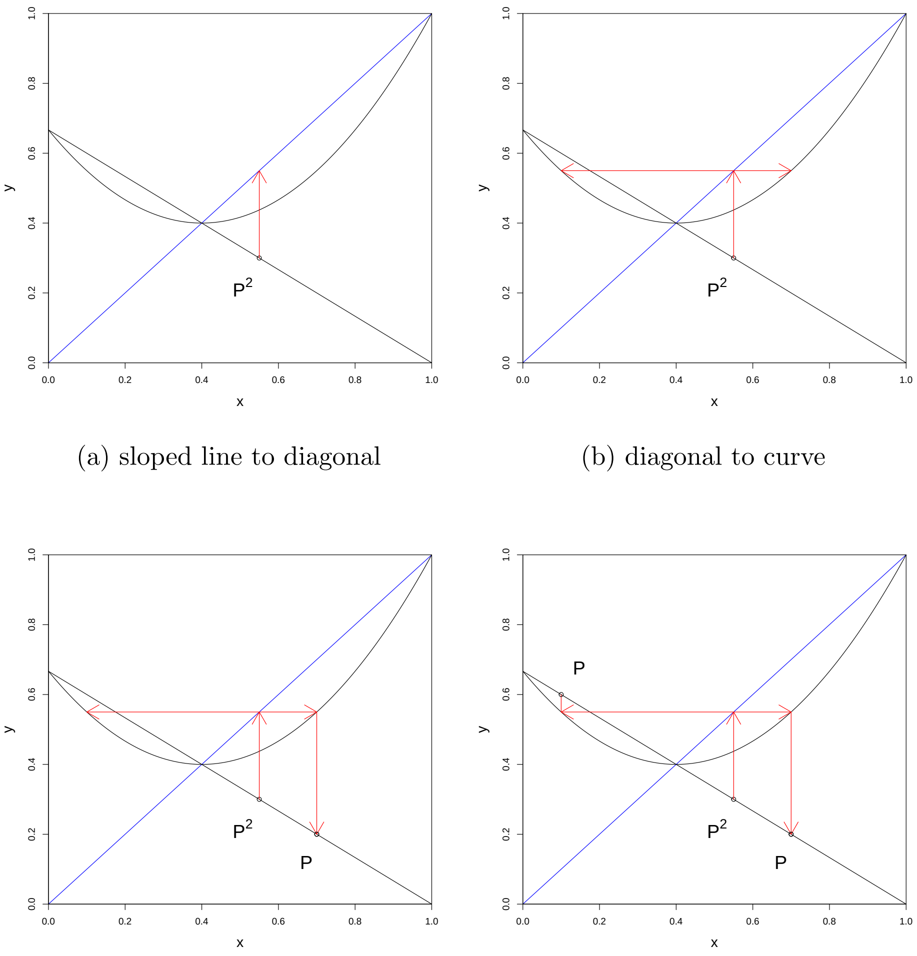


(f) diagonal to sloped line

Figure 3.3

Notice that the vertex of the parabola is exactly the intersection of the line through $[1, 0]$ and $[x, y]$ with the line $y = x$ which represents the limiting vector inside the unit square.

We can use a similar method to find a square root of P. According to Brill and Hlynka, to find $P^{\frac{1}{2}}$ we use the same geometric method but go in the opposite direction. Notice that we obtain two square roots and not one.



(a) sloped line to diagonal

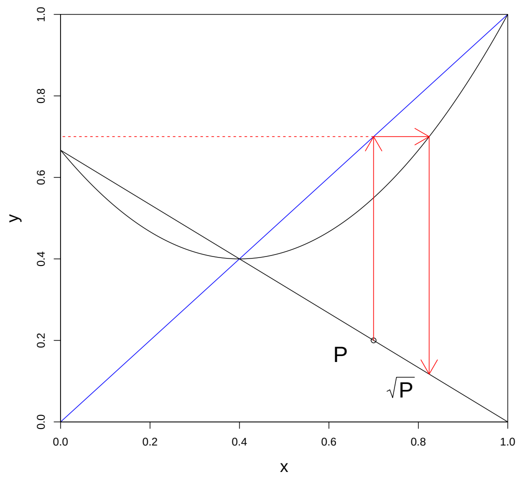
(b) diagonal to curve

(c) curve to sloped line, first root

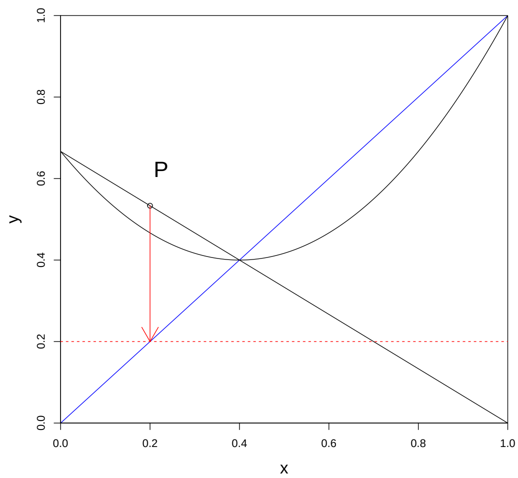
(d) curve to sloped line, second root

Figure 3.4

Further, observe that if P^2 was further down the sloped line, we would be able to obtain only a single probability square root. On the other hand, if P^2 was located up the sloped line, we would not have any probability square roots.



(a) Single stochastic square root (the second square root is outside the unit square.)



(b) No stochastic square roots.

Figure 3.5

We conclude this section by presenting an alternative geometric method for finding probability square roots for Markov transition matrices. Recall the function

$$f_n(x) = (x-a)^n(1-a)^{1-n} + a.$$

where $0 < a < 1$. Then, for $n = \frac{1}{2}$ we have

$$f_{\frac{1}{2}}(x) = \pm[(x-a)(1-a)]^{\frac{1}{2}} + a.$$

The curve $f_{\frac{1}{2}}(x)$ can also be used to locate probability square roots within the unit square. First, observe the function $f_{\frac{1}{2}}(x)$ for different values of a, $0 < a < 1$.

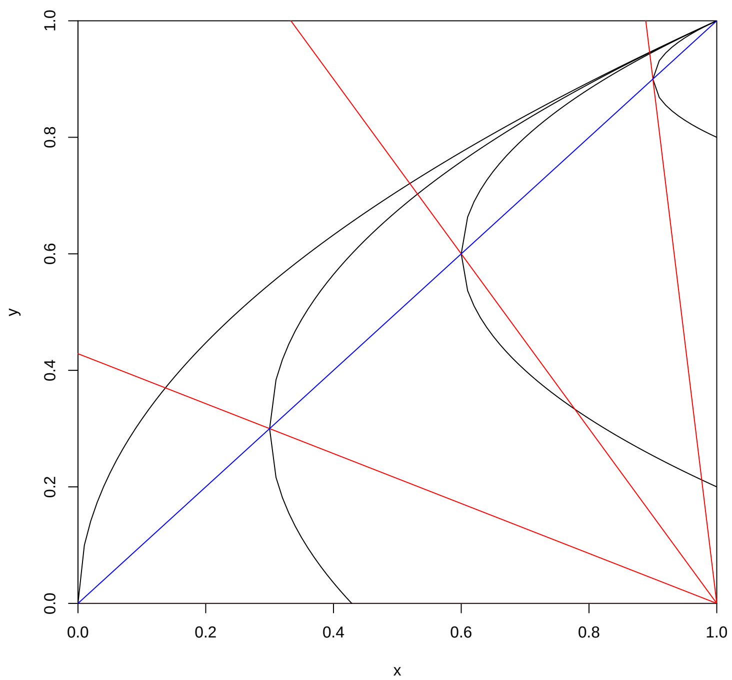
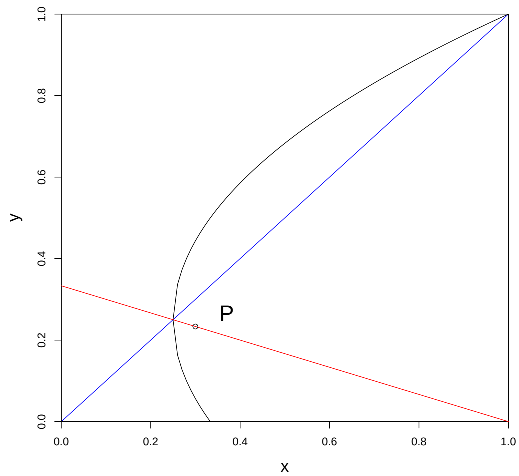
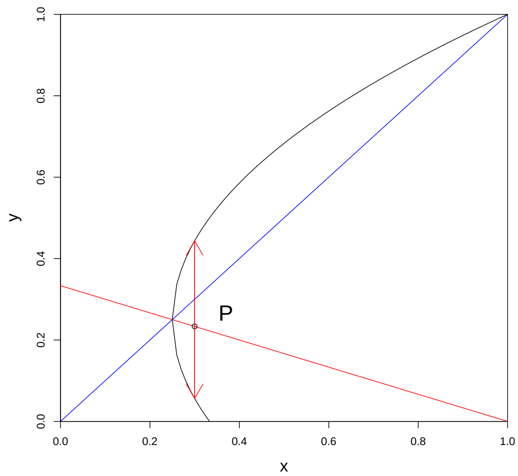


Figure 3.6

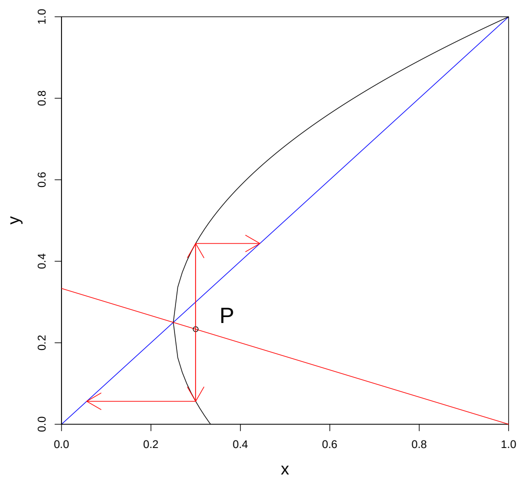
Next, given $[x, y]$ within the unit square, we use the curve $f_{\frac{1}{2}}(x)$ to find probability square roots $[x_{\frac{1}{2}}, y_{\frac{1}{2}}]$ and $[x'_{\frac{1}{2}}, y'_{\frac{1}{2}}]$ (assuming both lie within the unit square).



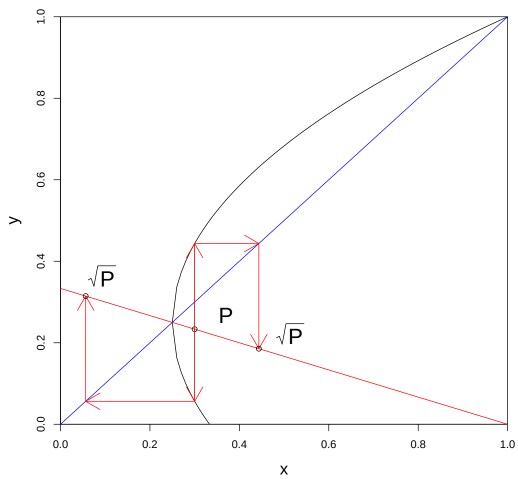
(a) probability matrix P



(b) sloped line to curve



(c) curve to diagonal



(d) diagonal to sloped line

Figure 3.7

Note that the same conclusion about the existence of probability square roots can be drawn: if P was further down the sloped line, we would be able to obtain only a single probability square root. Alternatively, if P was located up the sloped line, there would not be any probability square roots. This leads to the question: what are the conditions for a stochastic matrix to have zero, one and two probability square roots?

3.2 Stochastic roots: Regions within the unit square

Now our goal is to find regions within the unit square which will represent the collections of stochastic matrices with zero, one and two stochastic square roots respectively. As was briefly mentioned in the preceding section, we are aware that the location of P within the unit square determines the number of stochastic square roots that it has. Specifically, we are concerned with how far up or down the sloped line point $[x, y]$ is located.

First, graph parabolas for different values of a to observe their behaviour. Recall that $f_2(x) = (x - a)^2(1 - a)^{-1} + a$.

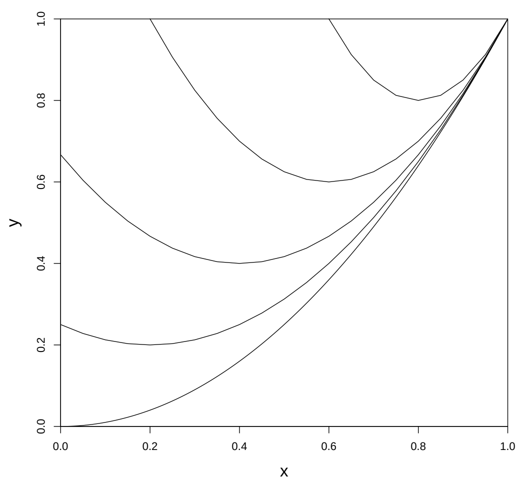


Figure 3.8: Parabolas for $a = 0, a = 0.2, a = 0.4, a = 0.6, a = 0.8$

From the graph above we see that parabolas with $a < \frac{1}{2}$ and parabolas with $a > \frac{1}{2}$ behave differently in terms of where they intersect the boundaries of the unit square: for $a < \frac{1}{2}$, they intersect $x = 0$ and for $a > \frac{1}{2}$, they intersect $y = 1$. Therefore, in this section cases for $a < \frac{1}{2}$ and $a > \frac{1}{2}$ will be considered separately.

We start with the case $a < \frac{1}{2}$. In this case, each parabola intersects $y = 0$ at a point $[0, \frac{a}{1-a}]$ where $[a, a]$ represents the limiting vector for the collection of matrices on the sloped line.

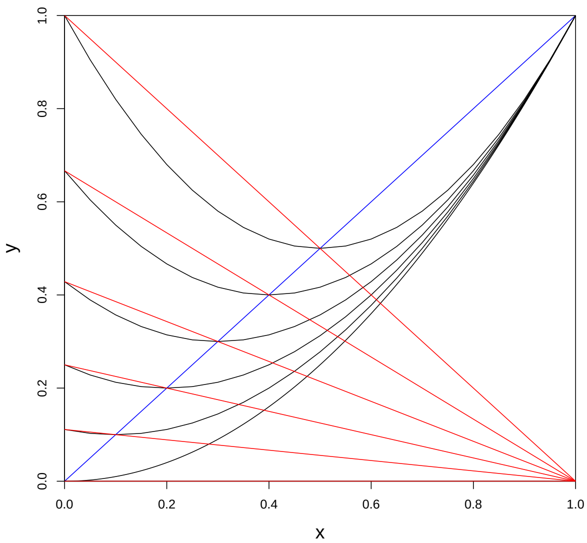


Figure 3.9: Sloped lines (collections of probability matrices) with their corresponding parabolas

Next, we observe that if P projects onto the diagonal below the parabola's vertex, then P has zero square roots. This occurs for every matrix P (point $[x, y]$ within the unit square) on the top right hand side of the diagonal within the unit square. Next, if P projects onto the diagonal above the point where the parabola intersects $x = 0$, then P has exactly one probability square root. And lastly, if P projects onto the diagonal between the vertex and the point where parabola intersects $x = 0$, then P has two probability square roots. Hence, the height from the point of intersection to the vertex of the parabola determines the segment on the sloped line which corresponds to the collection of probability matrices with two probability square roots.

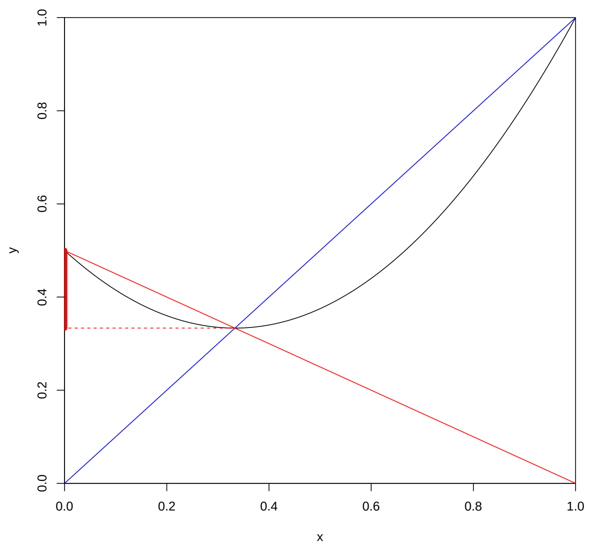


Figure 3.10: Height from the point of intersection of a parabola with $y = 0$ to the vertex.

In other words, we are looking at the segment of the sloped line which projects onto the diagonal within that height. Projections below and above result in a segments representing the collections of matrices with zero and one probability square roots respectively.

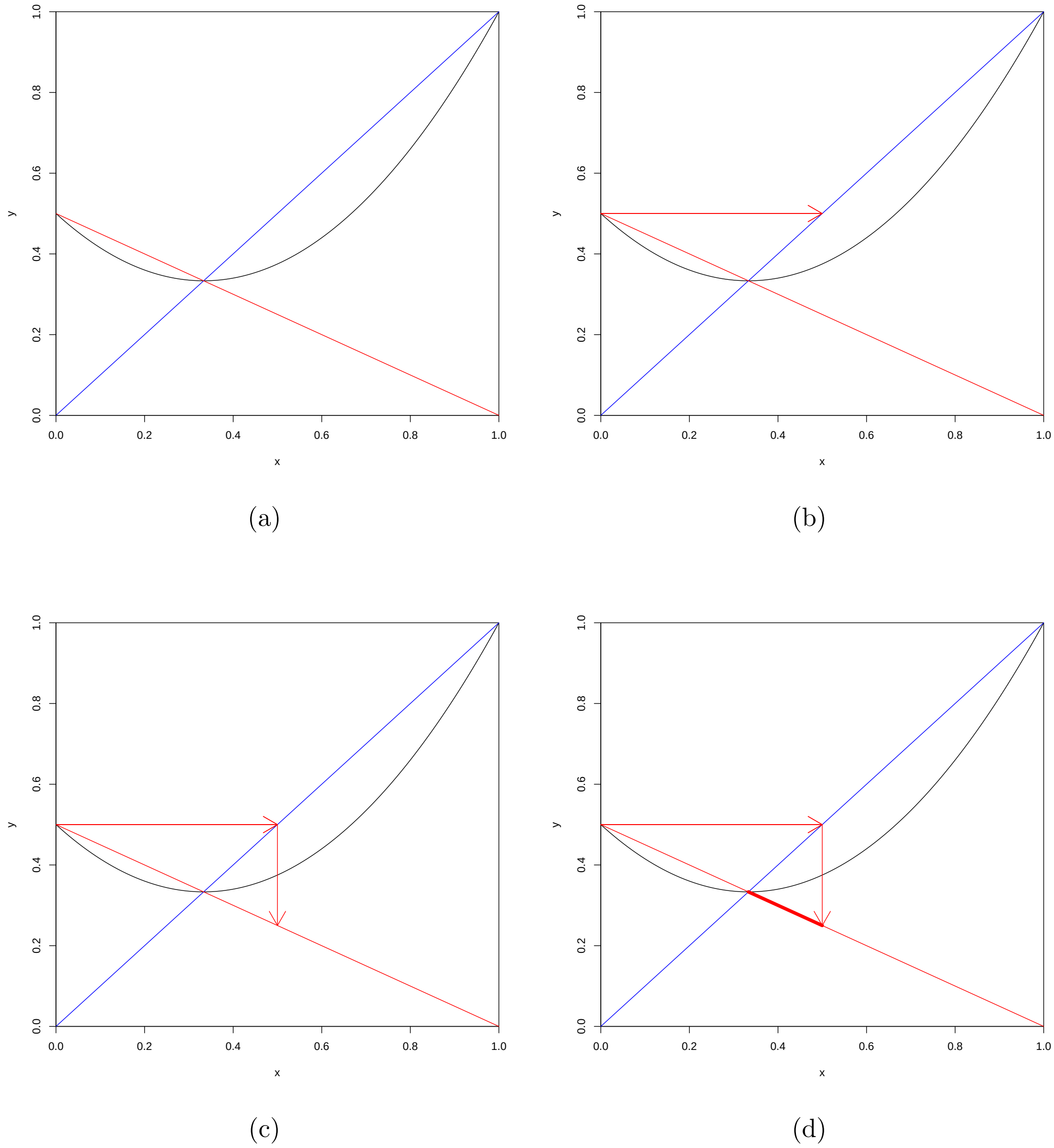


Figure 3.11: Process of finding the segment on the sloped line which corresponds to a collection of probability matrices with two probability square roots

We can repeat this procedure for multiple parabolas with a from 0 to $\frac{1}{2}$.

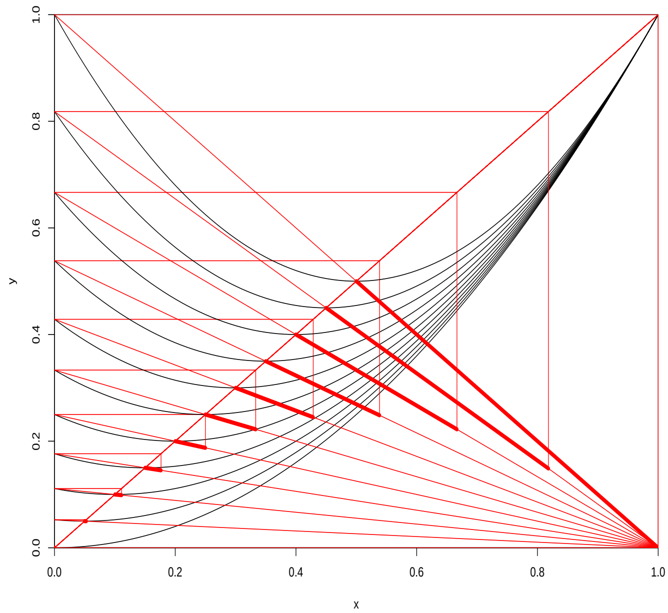


Figure 3.12

The bold segments in the graph indicate the region within the unit square representing the collection of probability matrices with two probability square roots. As was indicated previously, the region on the left hand side of the diagonal represents the collection of probability matrices with zero probability square roots. Next, the region below the bold segments represents the collection of probability matrices with a unique probability square root. Finally, the triangular region on the right hand side requires further investigation for the case where $\frac{1}{2} < a < 1$.

Note that the sloped line $y = \frac{a}{1-a}(1 - x)$ intersects the vertical line $x = \frac{a}{1-a}$ in $3.11(d)$ so that (x, y) satisfies $y = x(1-x)$ on the boundary of zero roots. Hence, the boundary curve is an upside-down parabola which passes through points $[0, 0]$, $[1, 0]$ and $[0.5, 0.25]$ and has the following equation

$$y = -x^2 + x.$$

Now, consider the case where $\frac{1}{2} < a < 1$. In this case parabolas intersect the line $y = 1$.

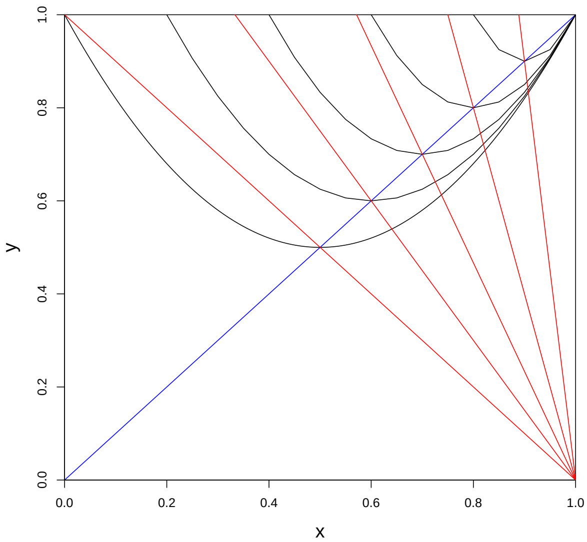
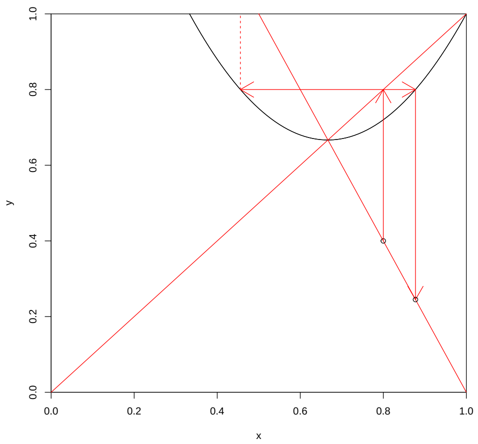
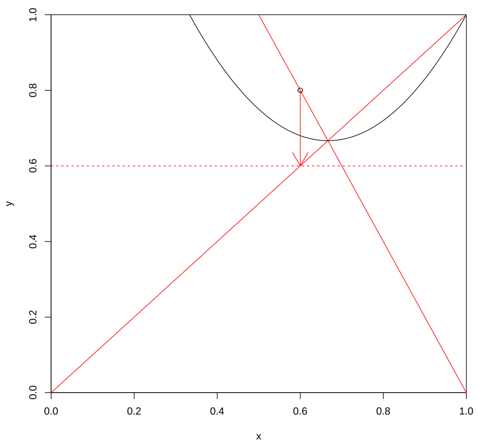


Figure 3.13

Since to obtain a square root we are projecting from the curve onto the diagonal, it is easy to notice that the tails of the parabolas will project onto the diagonal outside the unit square resulting in one non-stochastic square root. Further, as mentioned before, any matrix represented as a point on the upper left hand side of the diagonal has no square roots.



(a) Single stochastic square root (the second square root is outside the unit square.)



(b) No stochastic square roots.

Figure 3.14

Hence, we are interested in finding the segment on the sloped line which projects onto the tails of the parabola within the boundaries of the unit square resulting in two probability square roots.

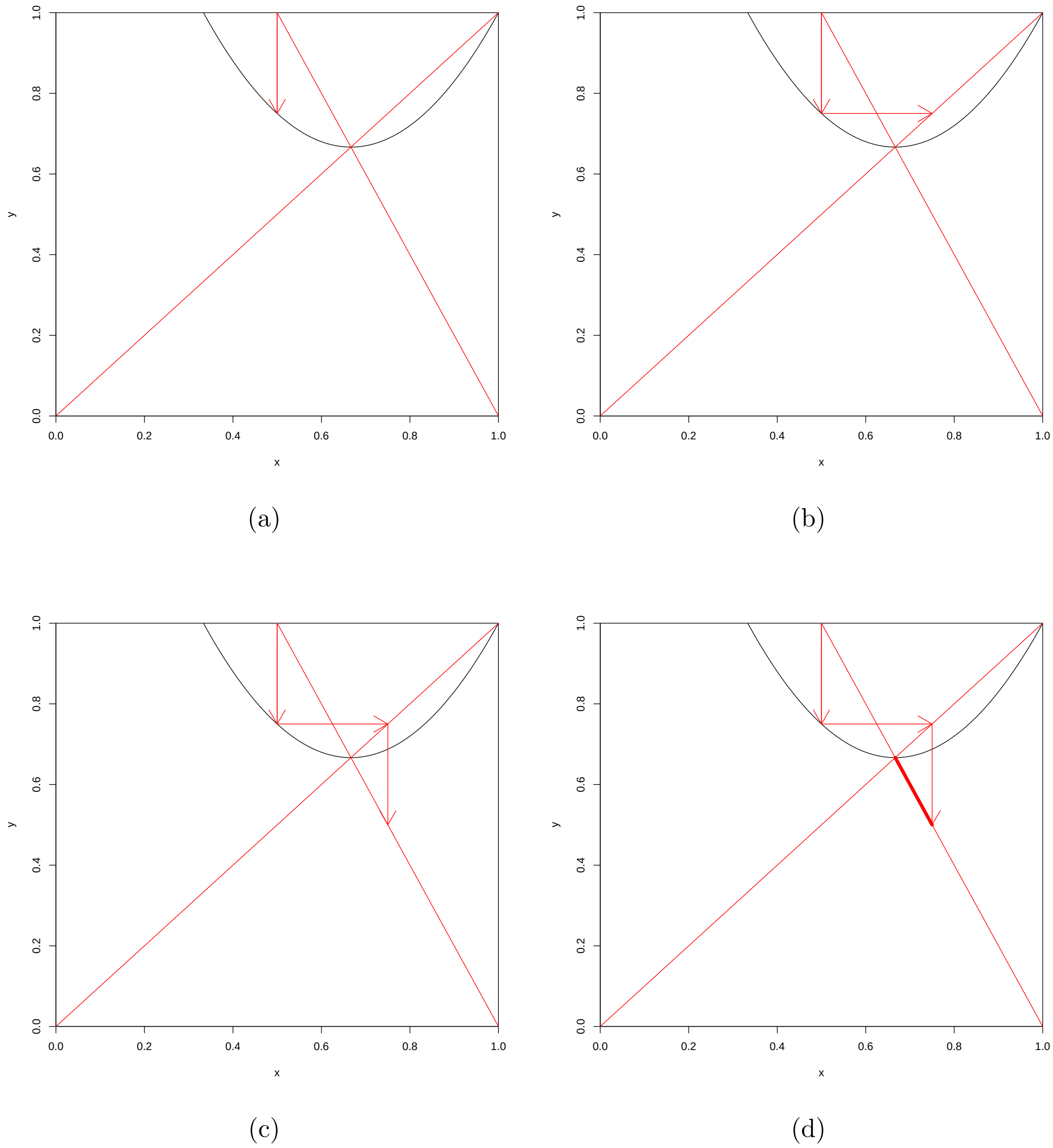


Figure 3.15: Process of finding the segment on the sloped line which corresponds to a collection of probability matrices with two probability square roots

We repeat this procedure for multiple parabolas with a from $\frac{1}{2}$ to 1.

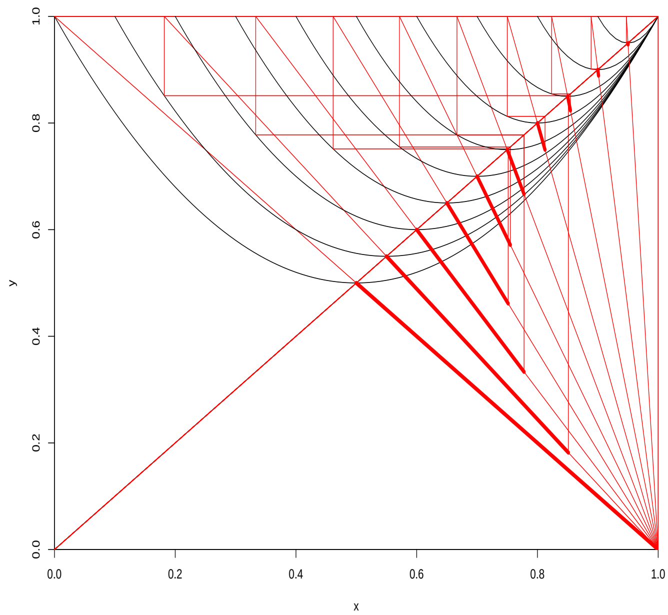


Figure 3.16

The bold segments in the graph illustrate the region within the unit square which represents the collection of probability matrices with two probability square roots. The boundary curve for $\frac{1}{2} < a < 1$ is identical to the one for $a < \frac{1}{2}$ but reflected against $y = 1 - x$. It is a parabola passing through the points $[1, 0]$, $[1, 1]$ and $[0.75, 0.5]$ i.e.

$$x = y^2 - y + 1.$$

Note that the symmetry of the boundary curves occurs due to the fact that the states in a two-state Markov chain are interchangeable i.e. relabeling state 1 into state 2 does not have any affect on the analysis.

Therefore, we have obtained the regions within the unit square corresponding to the collections of probability matrices with zero, one and two probability square roots respectively. The regions are illustrated in the graph below.

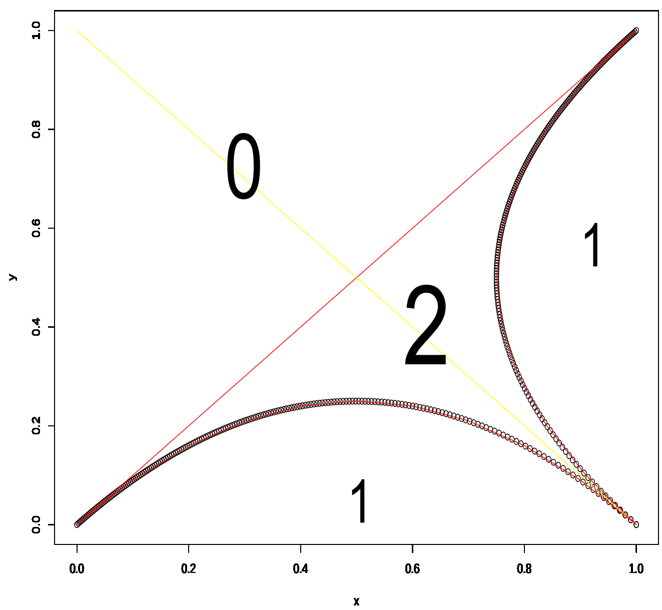


Figure 3.17: 0, 1 and 2 indicate the number of probability square roots in the region

In case of nonuniqueness of probability square roots, we are faced with a few questions: which probability square root should we select and why one root could be preferred over another? According to Guerry [2013], nonuniqueness of $\sqrt{P}$ corresponds to the case where "the observations are consistent with more than one discrete-time Markov chain with time unit 0.5" [4]. Hence, according to Guerry [2013], we can select a more appropriate probability square root examining the following criterion: which root results in monotonic evolutions of the expected outcome? Therefore, the root that produces more logical fluctuations of the expected outcome in a half-time period will be given the preference. In our geometric analysis, the selection of a more appropriate probability square root is visually intuitive since the root that occurs in the lower right diagonal and is in closer proximity to P within the unit square is obviously preferred.

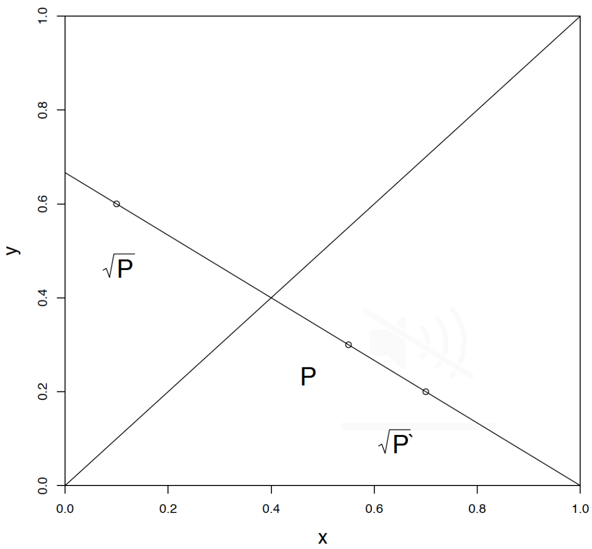
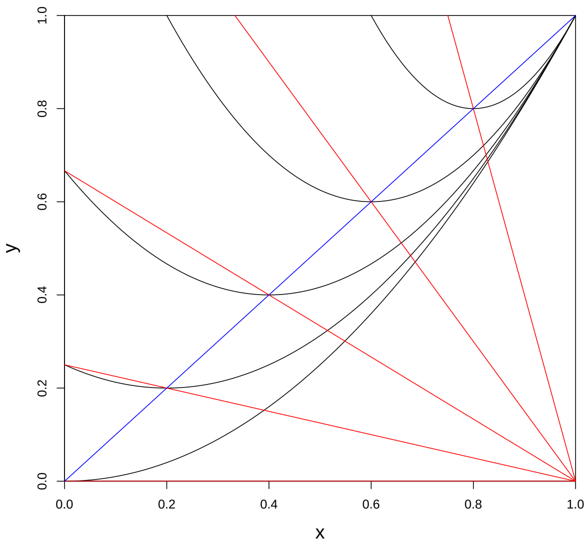


Figure 3.18: $\sqrt{P'}$ is the preferred probability square root.

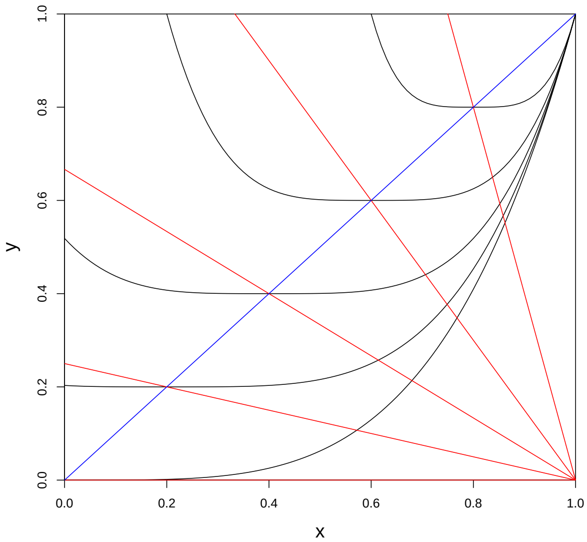
To conclude this section, we briefly compare the regions within the unit square corresponding to the collections of probability matrices with zero, one and two probability roots when $n = 2$ and $n = 4$ (i.e. probability square roots and probability forth roots). Note that when $n = 4$, the function $f_n(n)$ has the following form

$$f_4(x) = (x - a)^4(1 - a)^{-3} + a.$$

We plot $f_4(x)$ for different values of a $(0 < a < 1)$ and compare the results to $f_2(x)$:



(a) $f_2(x)$



(b) $f_4(x)$

Figure 3.19

From the plots above we observe that the curves are similar; however, $f_4(x)$ is wider near the vertex compared to a parabola. Hence, we can anticipate that the region containing the collection of probability matrices with two probability *forth* roots will be more narrow compared to the corresponding region with two probability *square* roots. Applying the same techniques for finding boundaries within the unit square, our assumption is confirmed and we can see the comparison of the boundaries below:

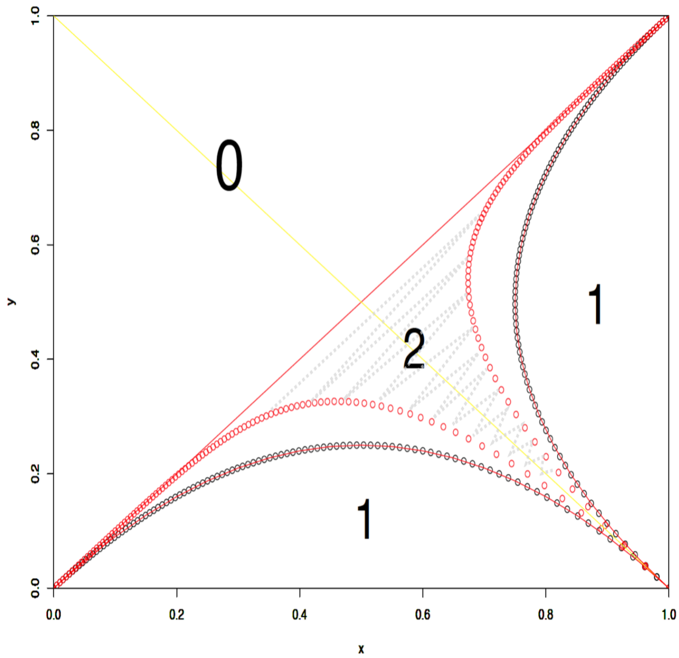


Figure 3.20: 0, 1 and 2 indicate the number of probability square/forth roots in the region. The shaded region corresponds to the case when $n = 4$.

The parametric equations of the bottom boundary and the right-hand side boundary of the shaded region are

$$\begin{cases} x = \frac{t^4}{(1-t)^3} + t \\ y = \frac{-t^5}{(1-t)^4} + t \end{cases}$$

and

$$\begin{cases} x = \frac{(t-1)^5}{t^4} + t \\ y = \frac{(t-1)^5+t^5-t^4}{(t-1)t^3} \end{cases}$$

respectively, which are symmetric about the line $y = 1 - x$.

3.3 Comparing geometric and numeric results

So far we have focused on the geometric interpretation of finding probability square roots. In this section we aim to compare our results obtained with geometric approach to the results of other authors who used analytic and/or numeric methods.

We start with Guerry [2013] where she discusses the embedding problem for discrete-time Markov chains. Guerry defines a 2×2 probability matrix P as

$$P = \begin{bmatrix} c & 1-c \\ d & 1-d \end{bmatrix}.$$

Then, according to Guerry, a probability matrix A is a stochastic square root of P if and only if

$$a_{11}^2 + (1 - a_{11})a_{21} = c$$
$$a_{11}a_{21} + (1 - a_{21})a_{21} = d$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

This results in a quadratic equation in a_{11}

$$(1 - c + d)a_{11}^2 - 2da_{11} + c^2 - c + d = 0$$

with the discriminant $D = 4(1-c)^2(c-d)$. Then, according to Guerry, we consider three cases.

First, if $c < d$, no stochastic square roots exist for P. This is mirrored in our geometric approach as the case when the matrix P lies in the left upper diagonal of the unit square and, hence, has no probability square roots.

Next, if $c = d$, then P has exactly one stochastic square root $A = [a_{ij}]$ where $a_{11} = a_{21} = c = d$. The preceding is equivalent to the case when the matrix P lies on the diagonal of the unit square and, therefore, has exactly one probability square root.

Finally, if $c > d$ and $1 - c + d \neq 0$, then the above quadratic equation in a_{11} has the following solutions

$$a_{11} = \frac{\sqrt{c-d}(1-c)+d}{1-c+d} \text{ and } a'_{11} = \frac{\sqrt{c-d}(c-1)+d}{1-c+d}$$

Substituting both into $a_{21} = \frac{c-a_{11}^2}{1-a_{11}}$ we get

$$a_{21} = d\frac{1-\sqrt{c-d}}{1-c+d} \text{ and } a'_{21} = d\frac{1+\sqrt{c-d}}{1-c+d}.$$

Note that a_{11} and a_{21} always belong to $[0, 1]$. Hence, we have at least one stochastic root A of P. When a'_{11} and a'_{21} also belong to $[0, 1]$, two probability roots exist. This case is equivalent to the instance when P lies in the lower right diagonal of the unit square and there exists at least one probability square root. The second probability square root exists in the region of the unit square bounded by the curves $y = -x^2 + x$, $x = y^2 - y + 1$ and $y = x$.

Next, we present He and Gunn [2003] results. They defined the stochastic root problem as following: given a $m \times m$ matrix A such that $Ae = e$, find a matrix B that satisfies

$$A = B^n, \text{ where } B \neq 0 \text{ and } Be = e$$

for a positive integer $n > 1$. The explicit solutions to the above problem (all stochastic root matrices) were found for the 2×2 case where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

According to He and Gunn, if $a_{11} + a_{22} < 2$, A has exactly one stochastic square root defined as following:

$$b_{11} = \frac{1 - a_{22} + (1 - a_{11})\sqrt{a_{11} + a_{22} - 1}}{2 - a_{11} - a_{22}}, b_{12} = 1 - b_{11};$$
$$b_{22} = \frac{1 - a_{11} + (1 - a_{22})\sqrt{a_{11} + a_{22} - 1}}{2 - a_{11} - a_{22}}, b_{21} = 1 - b_{22}.$$

Further, if $a_{11} + a_{22} < 2$ and $min\{\frac{\sqrt{1-a_{11}}}{\sqrt{1-a_{22}}}, \frac{\sqrt{1-a_{22}}}{\sqrt{1-a_{11}}}\} \geq Tr(A) - 1 \geq 0$, then A has another stochastic square root

$$b_{11} = \frac{1 - a_{22} - (1 - a_{11})\sqrt{a_{11} + a_{22} - 1}}{2 - a_{11} - a_{22}}, b_{12} = 1 - b_{11};$$
$$b_{22} = \frac{1 - a_{11} - (1 - a_{22})\sqrt{a_{11} + a_{22} - 1}}{2 - a_{11} - a_{22}}, b_{21} = 1 - b_{22}.$$

He and Gunn claimed that for the 3×3 case all solutions cannot be found explicitly. Alternatively, they focused on all real root matrices that are functions of the original stochastic matrix A i.e.

$$B = \sum_i d_i A^i$$

for some complex numbers $d_0, d_1, ...$ (see Lancaster and Tismenetsky [1985]). According to He and Gunn, when $m = 2$, all stochastic roots are functions of A (unless $A = I$), whereas, when $m = 3$, there are infinitely many stochastic roots that are not functions of A (non-function roots).

Consider the following example provided by He and Gunn:

$$A = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

$$B(\epsilon) = A + \epsilon \begin{bmatrix} 0 & 0 & 0 \\ -2 & 1 & 1 \\ 2 & -1 & -1 \end{bmatrix}.$$

It is easily confirmed that $(B(\epsilon))^n = A$ for any $n > 1$ and $B(\epsilon)e = e$ for all small enough ϵ, where e is a vector of 1's. Hence, we can define a set $\mathcal{B}_\epsilon = \{B(\epsilon), |\epsilon| < 0.1\}$ which is an infinite set of non function stochastic nth roots of a stochastic matrix A for any $n > 1$.

3.4 Eigenvalues and trace of a stochastic matrix

We begin by proving two theorems which illustrate some properties of eigenvalues of a stochastic matrix. Although in this paper we are mostly dealing with 2×2 stochastic matrices, both theorems are proved for a general case of an $n \times n$ matrix.

Theorem 3.4.1. *One of the eigenvalues of a stochastic matrix is always equal to 1.*

Proof. Given a $n \times n$ stochastic matrix $P = [p_{ij}]$, we have $\sum_{j=1}^{n} p_{ij} = 1$ for all $i = 1, 2, ..., n$. This is equivalent to $Pv = 1v$ where $v = [1, 1, ..., 1]^T$ is a column vector of length n. Hence, 1 is always an eigenvalue of a stochastic matrix. □

Theorem 3.4.2. *For every eigenvalue λ of a stochastic matrix, $|\lambda| \leq 1$.*

Proof. Given a $n \times n$ stochastic matrix $P = [p_{ij}]$, let λ be one of the eigenvalues of P and let v be its corresponding eigenvector. Then,

$$Av = \lambda v.$$

Next, select $k, 1 \leq k \leq n$, such that $|v_k| \geq |v_i|$ for every $i = 1, 2, ..., n$. Then, extract the k^{th} row of the equation $Av = \lambda v$ to get

$$\sum_{j=1}^{n} p_{kj} v_j = \lambda v_k.$$

Taking absolute value of both sides and using the triangle inequality, we get

$$|\sum_{j=1}^{n} p_{kj} v_j| \leq \sum_{j=1}^{n} p_{kj} |v_j| \leq \sum_{j=1}^{n} p_{kj} |v_k| \leq |v_k|$$

Therefore,

$$|\sum_{j=1}^{n} p_{kj} v_j| = |\lambda v_k| = |\lambda||v_k| \leq |v_k|$$

Hence, we can conclude that $|\lambda| \leq 1$. □

Now, recall the equation for diagonalization of matrix P

$$P = U^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} U$$

which is valid for any 2×2 probability matrix P.

Guerry [2013] provided a brief proof of the following lemma:

Lemma 3.4.1. *Each 2×2 probability matrix is diagonalizable.*

Proof. Given a 2×2 matrix $P = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$, the characteristic equation is a quadratic equation in λ:

$$\lambda^2 - (1 + x - y)\lambda + x - y = 0$$

with the discriminant $D = (1 + x - y)^2 - 4(x - y) = (1 - x + y)^2 \geq 0$.

When $D > 0$, P has two distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 \neq 1$. Hence,

$$P = U^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \lambda_2 \end{bmatrix} U.$$

When $D = 0$, we have: $x - y = 1$. This is possible for a unique case where

$$P = \begin{bmatrix} 1 & 0 \\ \\ 0 & 1 \end{bmatrix}$$

which is diagonalizable.

Therefore, the proof is complete. □

Now, for P^n we have

$$P^n = U^{-1} \begin{bmatrix} \lambda_1^n & 0 \\ \\ 0 & \lambda_2^n \end{bmatrix} U.$$

Taking $n = \frac{1}{2}$ and $\lambda_1 = 1$, we get

$$\sqrt{P} = U^{-1} \begin{bmatrix} \pm 1 & 0 \\ \\ 0 & \pm\sqrt{\lambda_2} \end{bmatrix} U.$$

Therefore, if $\lambda_2 < 0$, P does not have a real-valued probability square root.

Recall that the trace of a matrix is the sum of its entries on the main diagonal. Hence, for a 2×2 probability matrix P we obtain

$$tr(P) = x + (1 - y) = 1 + (x - y) = \lambda_1 + \lambda_2$$

Theorem 3.4.2 implies that for any 2×2 probability matrix P we have

$$0 \leq tr(P) \leq 2$$

Further we have

$$\lambda_2 < 0 \iff x - y < 0 \iff 1 + x - y < 1 \iff tr(P) < 1$$

Hence, we can conclude that if $tr(P) < 1$, then P does not have a real-valued probability square root.

Simulating traces of 2×2 probability matrices over 10 million iterations gives us the following plot which confirms that $0 \leq tr(P) \leq 2$ and implies that $E[tr(P)] = 1$.

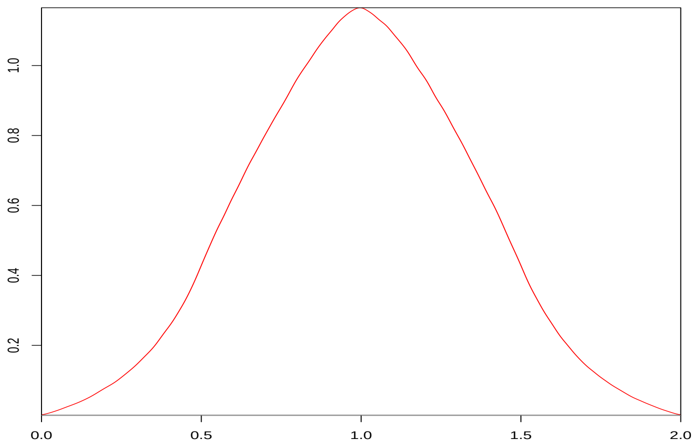


Figure 3.21: Density function of the trace for 2×2 probability matrices

Next we prove that $E[tr(P)] = 1$ for any $n \times n$ probability matrix P.

Theorem 3.4.3. *Given uniformly distributed random variables $X_1, X_2, ..., X_n$, we have*

$$E[\tfrac{X_1}{X_1+X_2+...+X_n}] = \tfrac{1}{n}.$$

Proof.

$$\begin{aligned}1 &= E[\frac{X_1 + X_2 + ... + X_n}{X_1 + X_2 + ... + X_n}] \\ &= E[\frac{X_1}{X_1 + X_2 + ... + X_n} + \frac{X_2}{X_1 + X_2 + ... + X_n} + ... + \frac{X_n}{X_1 + X_2 + ... + X_n}] \\ &= nE[\frac{X_1}{X_1 + X_2 + ... + X_n}].\end{aligned}$$

Hence,

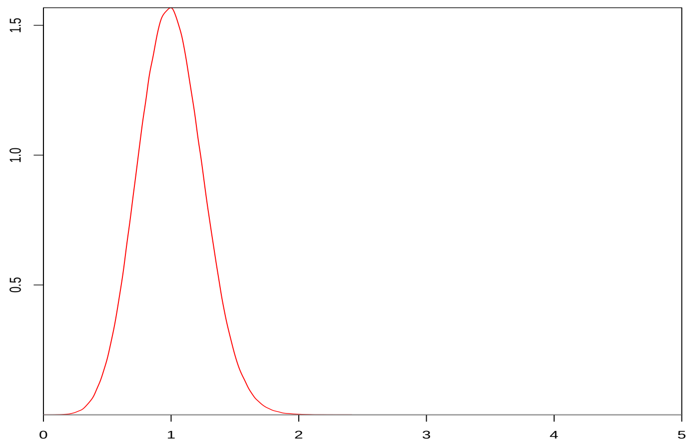
$$E[\tfrac{X_1}{X_1+X_2+...+X_n}] = \tfrac{1}{n}.$$

□

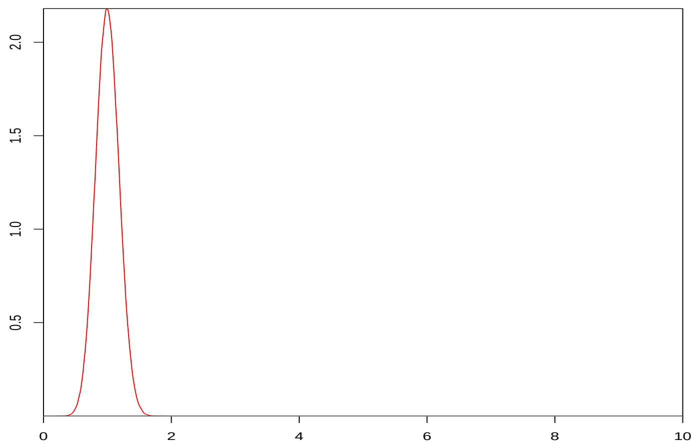
It follows from the theorem above, that for any $n \times n$ row-normalized matrix $P = [p_{ij}]$ derived by normalizing a matrix $A = [a_{ij}]$, we have

$$
\begin{aligned}
E[tr(P)] &= E[p_{11} + p_{22} + ... + p_{nn}] \\
&= E[\frac{a_{11}}{a_{11} + a_{12} + ... + a_{1n}} + \frac{a_{22}}{a_{21} + a_{22} + ... + a_{2n}} + ... + \frac{a_{nn}}{a_{n1} + a_{n2} + ... + a_{nn}}] \\
&= nE[\frac{a_{11}}{a_{11} + a_{12} + ... + a_{1n}}] \\
&= n\frac{1}{n} \\
&= 1
\end{aligned}
$$

Below observe distribution functions for traces of $(a) 5 \times 5$ and $(b) 10 \times 10$ Markov matrices



(a)



(b)

Figure 3.22

As mentioned previously $tr(P) \geq 1$ is a necessary and sufficient condition for a 2×2 probability matrix P to have square roots. According to Guerry [2013], this condition is not sufficient for 3×3 and 4×4 probability matrices. Guerry provided the following example where a matrix P has a stochastic root A and $tr(P) < 1$

$$P = \begin{bmatrix} 0.17 & 0.66 & 0.17 \\ 0.17 & 0.17 & 0.66 \\ 0.66 & 0.17 & 0.17 \end{bmatrix} \text{ and } A = \begin{bmatrix} 0.1 & 0.1 & 0.8 \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}.$$

According to Guerry, the condition $tr(P) \geq 1$ remains necessary for a probability matrix P to have a probability square root A for both 3×3 and 4×4 probability matrices. Guerry presented this result as a theorem in her paper [2012] where she considered each matrix size separately. Guerry proved each case of the theorem by showing that

$$tr(P) = tr(A \times A) = 1 + [\text{nonnegative term}] \geq 1.$$

3.5 Stochastic cube roots: Regions within the unit square

In this section we aim to find the regions within the unit square which correspond to the collections of 2×2 probability matrices with zero and one cube root. The curve $f_n(x)$ when $n = 3$ is of the following form

$$f_3(x) = (x - a)^3(1 - a)^{-2} + a.$$

As previously, we consider two cases: $0 < a < \frac{1}{2}$ and $\frac{1}{2} < a < 1$.

We start with the first case. Below is the plot of $f_3(x)$ for different values of a when $0 < a < \frac{1}{2}$.

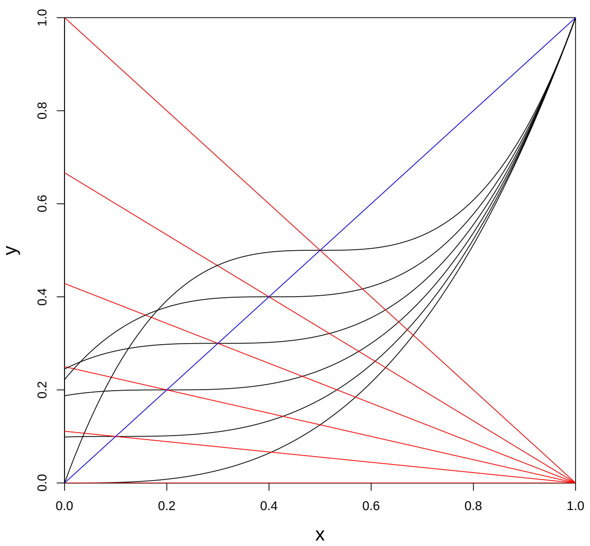
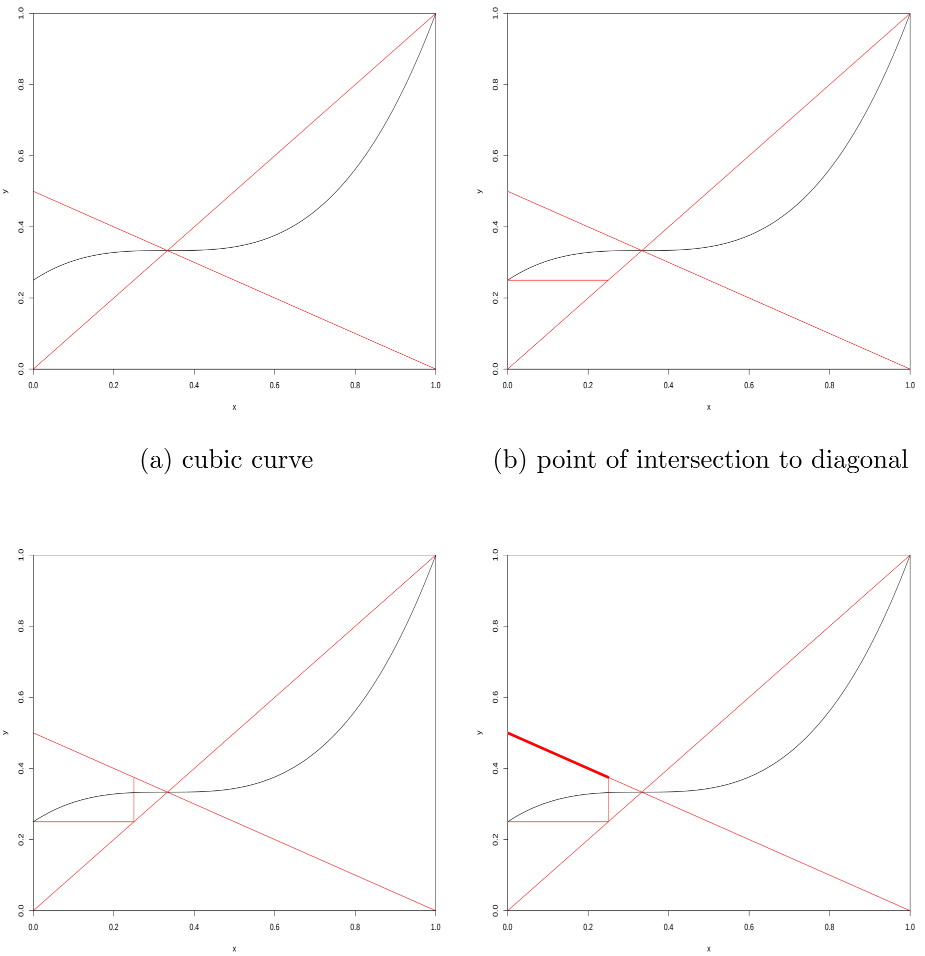


Figure 3.23

Observe that each cubic curve intersects y-axis at a point

$$[0, -a^3(1 - a)^{-2} + a] \text{ where } 0 < a < \tfrac{1}{2}.$$

The segment on the sloped line which projects onto the diagonal and then onto the cubic curve below this point of intersection contains matrices with single cube roots beyond the boundaries of the unit square (i.e. no probability cube roots). Alternatively, projections from the sloped line outside the said segment result in exactly one probability cube root. Below the process of finding the segment on the sloped line which corresponds to the collection of probability matrices with zero probability cube roots is shown.



(a) cubic curve

(b) point of intersection to diagonal

(c) diagonal to sloped line

(d) the obtained segment

Figure 3.24

Repeating this process for multiple cubic curves (multiple values of a), we get the following:

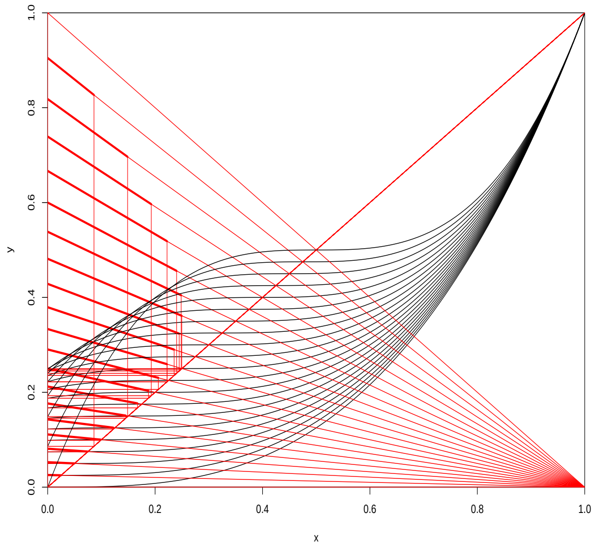


Figure 3.25

The bold segments on the plot correspond to the region within the unit square representing the collection of probability matrices with zero probability cube roots.

Next, we move onto the second case. Observe below the plot of $f_3(x)$ for different values of a when $\frac{1}{2} < a < 1$.

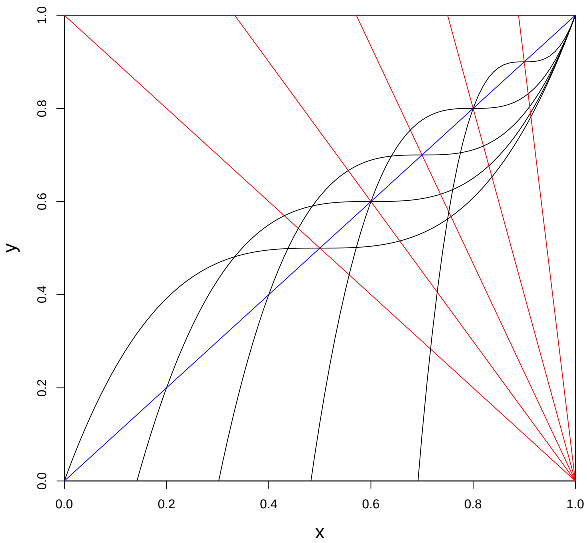
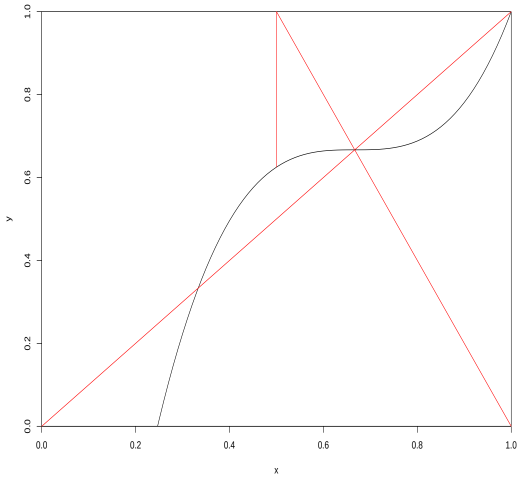


Figure 3.26

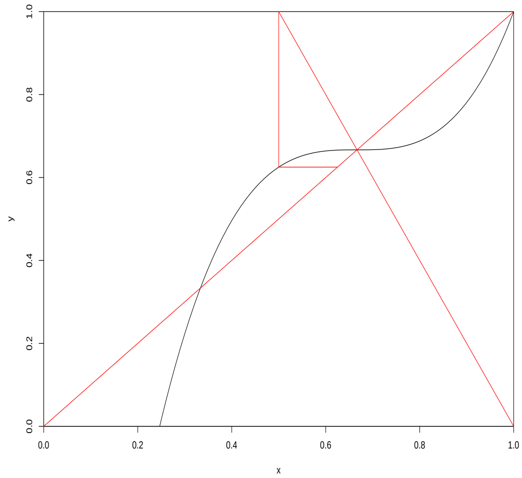
Here, we observe that each sloped line intersects $y = 1$ at a point

$$[2 - \frac{1}{a}, 1] \text{ where } \frac{1}{2} < a < 1.$$

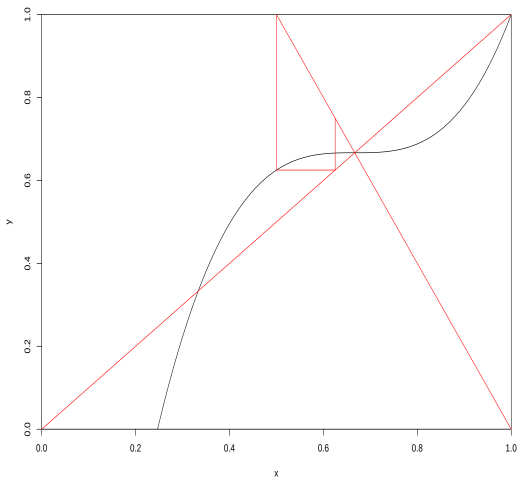
This point of intersection, when projected onto the curve, then the diagonal and then onto the sloped line, will determine the length of the segment on the sloped line which corresponds to the collection of probability matrices with zero probability cube roots. Observe that any matrix on the segment has a cube root beyond the boundaries of the unit square therefore resulting in zero probability cube roots.



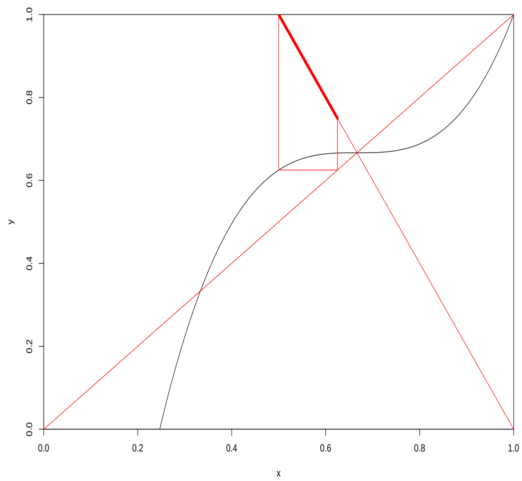
(a) point of intersection to curve



(b) curve to diagonal



(c) diagonal to sloped line



(d) the obtained segment

Figure 3.27

Repeating this process for multiple cubic curves, we get the following:

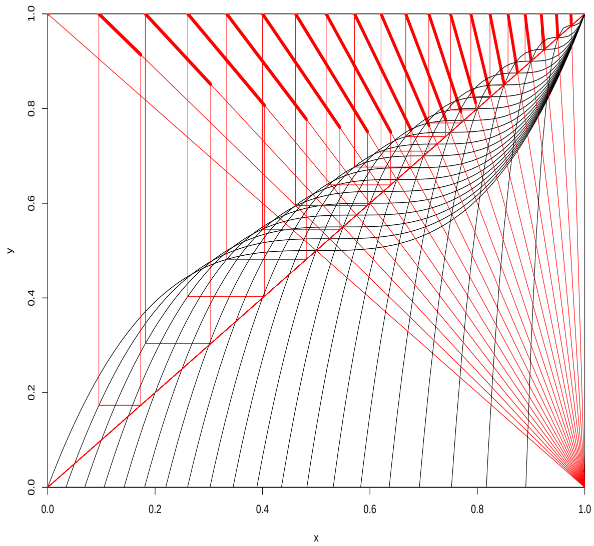


Figure 3.28

Hence, we have obtained the regions within the unit square which correspond to the collections of probability matrices with zero and one probability cube root. The regions are illustrated in the graph below.

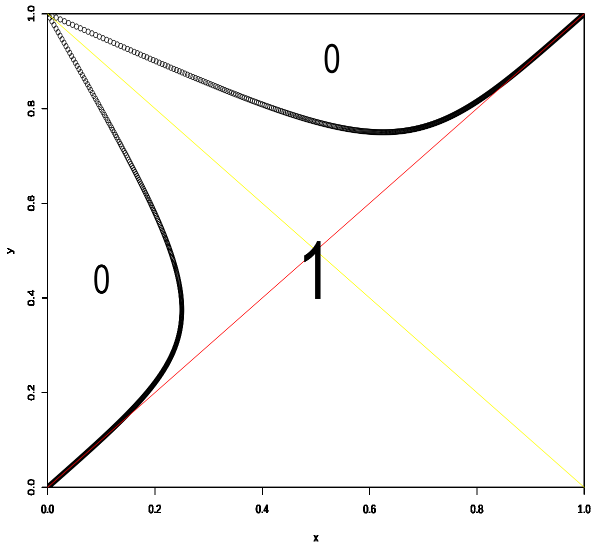


Figure 3.29

The parametric equations of the left-hand side boundary and the top boundary are

$$\begin{cases} x = \frac{-t^3}{(1-t)^2} + t \\ y = \frac{t^4}{(1-t)^3} + t \end{cases}$$

and

$$\begin{cases} x = \frac{(t-1)^4}{t^3} + t \\ y = \frac{(t-1)^4+t^4-t^3}{(t-1)t^2} \end{cases}$$

respectively, which are symmetric about the line $y = 1 - x$.

Chapter 4

A note on the algorithms for randomly generating probability matrices

In the previous chapter we used R to generate traces of probability matrices. The following algorithm for randomly generating an $n \times n$ probability matrix was used:

Algorithm 1

1. Use uniform distribution with support $(0, 1)$ to populate an $n \times n$ matrix.
2. Compute the row sums.
3. Divide each entry of the matrix by the corresponding row sum.
4. Obtain a row-normalized matrix i.e. a probability matrix.

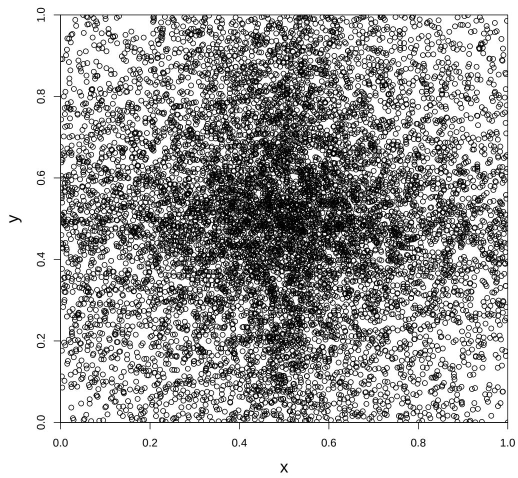
For a 2×2 case, consider an alternative algorithm:

Algorithm 2

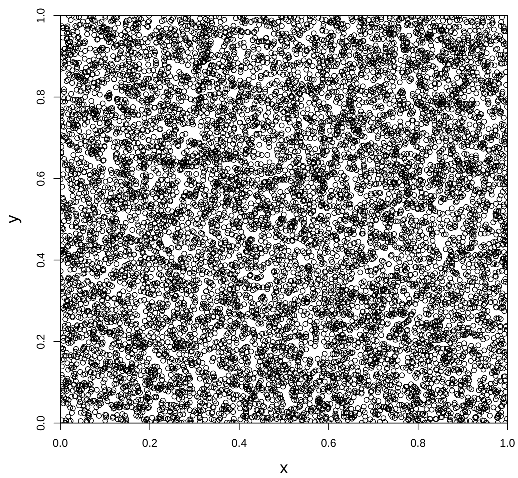
1. Use uniform distribution with support $(0, 1)$ to generate two values $[x, y]$.

2. Populate the matrix as following $\begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}$

Below observe the scatterplots within the unit squares produced in R for both algorithms for generating 2×2 probability matrices.



(a) Algorithm 1



(b) Algorithm 2

Figure 4.1

From the scatterplots above, it is clear that the two algorithms are not equivalent. The second algorithm generates 2×2 probability matrices which are uniformly distributed within the unit square. The scatterplot for the first algorithm, on the other hand, indicates that the 2×2 probability matrices are generated more densely in the centre of the unit square in the shape of a cross.

The nonequivalence of the two algorithms prompts us to ask the following question: which algorithm is more natural. We can argue that for the 2×2 case the second algorithm produces uniform results and, hence, is more appropriate. Note that for the $n \times n$ cases where $n \geq 3$ we are limited to the first algorithm. Hence, we are forced to use the algorithm which does not uniformly generate matrices within a unit hypercube.

We complete this chapter by presenting a theorem which defines the probability density function of a single diagonal entry of a 3×3 row-normalized matrix (probability matrix).

Theorem 4.0.1. *A random variable $U = \frac{X}{X+Y+Z}$ such that X, Y and Z are independent uniformly distributed random variables on $[0,1]$ has the following probability density function*

$$f(u) = \begin{cases} \frac{1}{(1-u)^2}, & \textit{if } 0 < u < \frac{1}{3} \\ \frac{1}{3}\left(\frac{1}{u^2} - \frac{1}{(1-u)^2}\right) + \frac{3u-1}{3u^3}, & \textit{if } \frac{1}{3} \leq u < \frac{1}{2} \\ \frac{1-u}{3u^3}, & \textit{if } \frac{1}{2} \leq u < 1. \end{cases}$$

Proof. Given $U = \frac{X}{X+Y+Z}$ where $X, Y, Z \sim unif(0,1)$, define

$$V = \frac{Y}{X+Y+Z} \text{ and } W = X + Y + Z.$$

Then,

$$x = uw$$
$$y = vw$$
$$z = w(1 - u - v)$$

and the Jacobian is the following

$$\begin{aligned}
||J|| &= det \begin{bmatrix} w & 0 & u \\ 0 & w & v \\ -w & -w & 1 - u - v \end{bmatrix} \\
&= w(w(1 - u - v) + vw) + uw^2 \\
&= w^2
\end{aligned}$$

Hence, we have

$$f_{U,V,W}(u, v, w) = w^2 f_{X,Y,Z}(x, y, z)$$

where

$$f_{X,Y,Z}(x, y, z) = \begin{cases} 1, & \text{if } 0 < x, y, z < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$f_{U,V,W}(u, v, w) = w^2$$

and we are interested in finding the marginal p.d.f. of U, $f_U(u)$. Note that since

$$0 < x = uw < 1$$
$$0 < y = vw < 1$$
$$0 < z = w(1 - u - v) < 1$$

we have

$$
\begin{aligned}
w &< \frac{1}{u}\\
w &< \frac{1}{v}\\
w &< \frac{1}{1-u-v}\\
v &< 1-u
\end{aligned}
$$

Next, we consider three cases: $0 < u < \frac{1}{3}$, $\frac{1}{3} \leq u < \frac{1}{2}$ and $\frac{1}{2} \leq u < 1$.

Case I: $0 < u < \frac{1}{3}$ implies

$$
\begin{aligned}
\frac{2}{3} &< 1-u < 1\\
\frac{1}{3} &< \frac{1-u}{2} < \frac{1}{2}\\
1 &< \frac{1}{1-u} < \frac{3}{2}\\
3 &< \frac{1}{u} < \infty.
\end{aligned}
$$

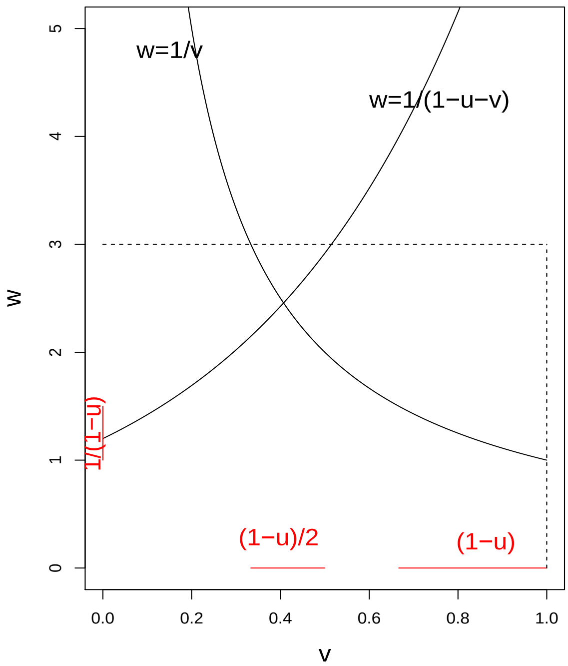


Figure 4.2

Then, the area under the curves $w = \frac{1}{1-u-v}$ and $w = \frac{1}{v}$ is:

$$
\begin{aligned}
f_U(u) &= \int_0^{\frac{1-u}{2}} \int_0^{\frac{1}{1-u-v}} w^2 dw dv + \int_{\frac{1-u}{2}}^{1-u} \int_0^{\frac{1}{v}} w^2 dw dv \\
&= \int_0^{\frac{1-u}{2}} \frac{1}{3} \frac{1}{(1-u-v)^3} dv + \int_{\frac{1-u}{2}}^{1-u} \frac{1}{3} \frac{1}{v^3} dv \\
&= \frac{1}{3}\frac{1}{2}(1-u-v)^{-2}\Big|_0^{\frac{1-u}{2}} + \frac{1}{3}\frac{1}{2}v^{-2}\Big|_{\frac{1-u}{2}}^{1-u} \\
&= \frac{1}{6}\Big(\frac{1-u}{2}\Big)^{-2} - \frac{1}{6}(1-u)^{-2} - \frac{1}{6}(1-u)^{-2} + \frac{1}{6}\Big(\frac{1-u}{2}\Big)^{-2} \\
&= \frac{1}{(1-u)^2}
\end{aligned}
$$

Case II: $\frac{1}{3} \leq u < \frac{1}{2}$ implies

$$\frac{1}{2} < 1 - u < \frac{2}{3}$$
$$\frac{1}{4} < \frac{1-u}{2} < \frac{1}{3}$$
$$\frac{3}{2} < \frac{1}{1-u} < 2$$
$$2 < \frac{1}{u} < 3.$$

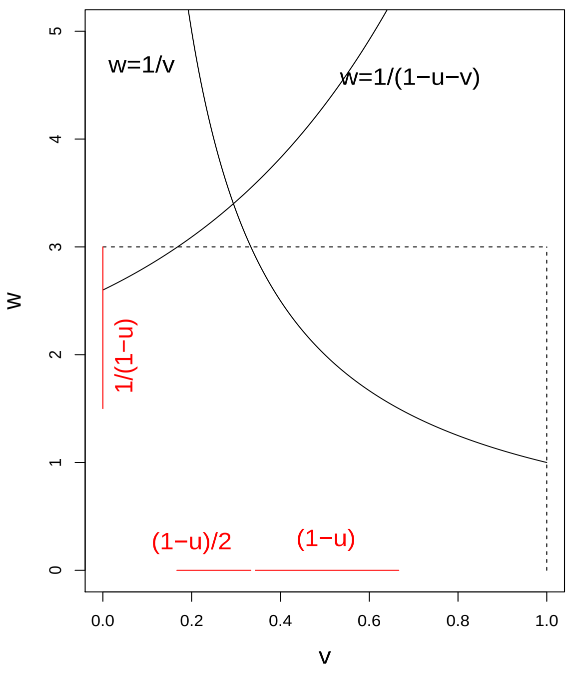


Figure 4.3

The area under $w = \frac{1}{1-u-v}$, $w = 3$ and $w = \frac{1}{v}$ is:

$$
\begin{aligned}
f_U(u) &= \int_0^{1-2u} \int_0^{\frac{1}{1-u-v}} w^2 dw dv + \int_{1-2u}^{u} \int_0^{\frac{1}{u}} w^2 dw dv + \int_u^{1-u} \int_0^{\frac{1}{v}} w^2 dw dv \\
&= \int_0^{1-2u} \frac{1}{3} \frac{1}{(1-u-v)^3} dv + \int_{1-2u}^{u} \frac{1}{3} \frac{1}{u^3} dv + \int_u^{1-u} \frac{1}{3} \frac{1}{v^3} dv \\
&= \frac{1}{3} \frac{1}{2} (1-u-v)^{-2} \Big|_0^{1-2u} + \frac{1}{3} \frac{1}{u^3} (u - 1 + 2u) + \frac{1}{3} \frac{1}{2} (v)^{-2} \Big|_u^{1-u} \\
&= \frac{1}{6} \frac{1}{(1-u-1+2u)^2} - \frac{1}{6} \frac{1}{(1-u)^2} + \frac{3u-1}{3u^3} - \frac{1}{6} \frac{1}{(1-u)^2} + \frac{1}{6} \frac{1}{u^2} \\
&= \frac{1}{3} \Big(\frac{1}{u^2} - \frac{1}{(1-u)^2} \Big) + \frac{3u-1}{3u^3}
\end{aligned}
$$

Case III: $\frac{1}{2} \leq u < 1$ implies

$$
\begin{aligned}
0 < 1 - u < \frac{1}{2} \\
0 < \frac{1-u}{2} < \frac{1}{4} \\
2 < \frac{1}{1-u} < \infty \\
1 < \frac{1}{u} < 2.
\end{aligned}
$$

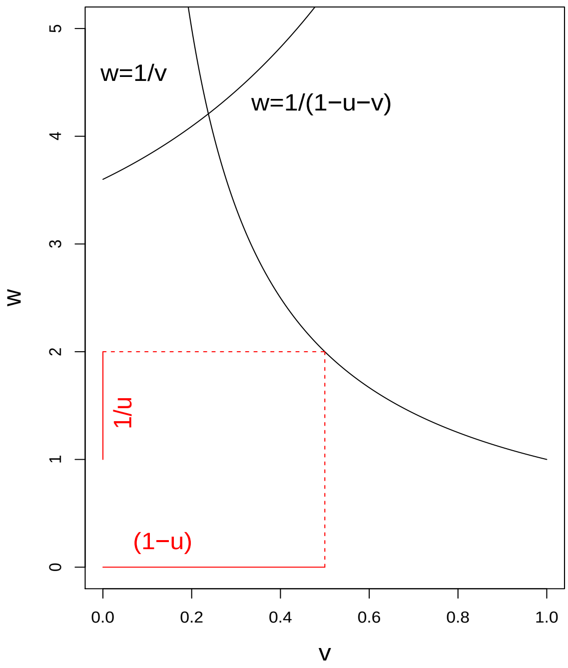


Figure 4.4

The area within the rectangle is:

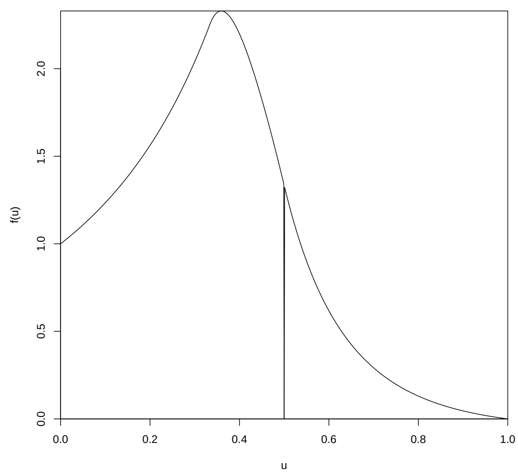
$$f_U(u) = \int_0^{1-u} \int_0^{\frac{1}{u}} w^2 dw dv = \int_0^{1-u} \frac{1}{3} \frac{1}{u^3} dv = \frac{1-u}{3u^3}.$$

Therefore, the density function of $U = \frac{X}{X+Y+Z}$ such that X, Y, Z are independent uniformly distributed random variables is the following

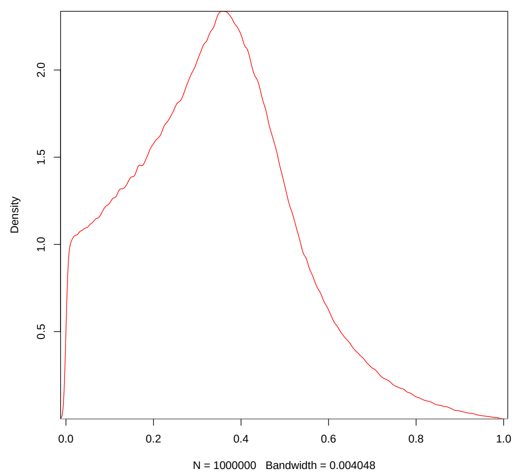
$$f(u) = \begin{cases} \frac{1}{(1-u)^2}, & \text{if } 0 < u < \frac{1}{3} \\ \frac{1}{3}\left(\frac{1}{u^2} - \frac{1}{(1-u)^2}\right) + \frac{3u-1}{3u^3}, & \text{if } \frac{1}{3} \leq u < \frac{1}{2} \\ \frac{1-u}{3u^3}, & \text{if } \frac{1}{2} \leq u < 1. \end{cases}$$

□

Below observe the plots of the density function $f(u)$ and the density function obtained by generating a million single entries of 3×3 row-normalized matrices (i.e. $\frac{a_{11}}{a_{11}+a_{12}+a_{13}}$) in R.



(a) Plot of $f(u)$



(b) Plot of the density function generated in R

Figure 4.5

Using R we confirm that $f(u)$ integrates to 1 and that $E[U] = \frac{1}{3}$.

Conclusions

In this paper we examined the embeddability problem in discrete-time homogenous Markov chains which corresponds to the case where a discrete-time Markov chain is observed on time intervals of a fraction (e.g. $\frac{1}{2}, \frac{1}{3}$ etc.) of the duration of the original Markov chain. The geometric method first derived by Brill and Hlynka was employed to establish the conditions under which 2×2 probability matrices have multiple, single or zero probability square or cube roots. The results were compared to numeric results derived by other authors. Finally, algorithms for generating $n \times n$ Markov transition matrices were discussed. All geometric derivations for roots of stochastic matrices are new.

Appendix A

R code

```
power2<-function(n){      ###curves for finding square roots, method 1
  for (a in seq(0,1,n)) {
    eq = function(x){((x-a)^2)/(1-a)+a}
    curve(eq, from=0, to=1, add=TRUE, xlim=c(0,1), ylim=c(0,1),
            xlab="x", ylab="y", xaxs="i", yaxs="i",main="",cex.lab=1.5)
    x<-seq(0,1,0.05)
    lines(x,-a/(1-a)*(x-1),type= "l", col="red")
  }
  #grid(nx = 20, ny = 20, col = "lightgray", lty = "dotted")
   lines(x,x,type= "l", col="blue")
  return()
}

sqroot<-function(n){       ###curves for finding square roots, method 2
  for (a in seq(0,1,n)) {
    eq1 = function(x){sqrt(x-a)*sqrt(1-a)+a}
    eq2 = function(x){-sqrt(x-a)*sqrt(1-a)+a}

    curve(eq1, add=TRUE, xlim=c(0,1), ylim=c(0,1),
            xaxs="i", yaxs="i",main="",xlab="x", ylab="y")
    curve(eq2, add=TRUE, xlim=c(0,1), ylim=c(0,1),
            xaxs="i", yaxs="i",main="",xlab="x", ylab="y")

    x<-seq(0,1,0.001)
    lines(x,-a/(1-a)*(x-1),type= "l", col="red")
  }
  lines(x,x,type= "l", col="blue")
  return()
}
```

```
cubic<-function(n){     ###curves for finding cube roots
  for (a in seq(0,1,n)) {
    eq = function(x){((x-a)^3)/(1-a)^2+a}
    curve(eq, from=0, to=1, add=TRUE, xlab="x", ylab="y",
          xaxs="i", yaxs="i",main="",cex.lab=1.5)
    lines(x,-a/(1-a)*(x-1),type= "l", col="red")
  }
  #grid(nx = 20, ny = 20, col = "lightgray", lty = "dotted")
  lines(x,x,type= "l", col="blue")
  return()
}

power4<-function(n){     ###curves for finding 4th roots
  for (a in seq(0,1,n)) {
    eq = function(x){((x-a)^4)/(1-a)^3+a}
    curve(eq, from=0, to=1, add=TRUE, xlab="x", ylab="y",
          xaxs="i", yaxs="i",main="",cex.lab=1.5)
    lines(x,-a/(1-a)*(x-1),type= "l", col="red")
  }
  #grid(nx = 20, ny = 20, col = "lightgray", lty = "dotted")
  lines(x,x,type= "l", col="blue")
  return()
}


boundfull<-function(n){

  ##### square root boundaries

  b1 <- numeric()
  b2 <- numeric()
  c1 <- numeric()
  c2 <- numeric()
  i=1
  for (a in seq(0,0.5,n)) {
    b1[i]=a/(1-a)
    b2[i]=b1[i]*(1-b1[i])
    i=i+1
  }
  i=1
  for (a in seq(0.5,1,n)) {
    #c1[i]=(2-a-1/a)^2/(1-a)+a
    #c2[i]=-a/(1-a)*(c1[i]-1)
    c1[i]=((1-a)^3)/a^2+a
    c2[i]=2-1/a
    i=i+1
  }
```

```
max1=which.max(b2)
min1=which.min(c1)

l1=length(b1)
l2=length(c1)
for (i in 1:l1) {
  plot(b1[[i]], b2[[i]], xlim=c(0,1), ylim=c(0,1),
          xlab = "x", ylab = "y",xaxs="i", yaxs="i")
  par(new=T)
}
for (i in 1:l2) {
  plot(c1[[i]], c2[[i]], xlim=c(0,1), ylim=c(0,1),
          xlab = "x", ylab = "y", xaxs="i", yaxs="i")
  par(new=T)
}

##### 4th root boundaries

d1 <- numeric()
d2 <- numeric()
e1 <- numeric()
e2 <- numeric()
i=1
for (a in seq(0,0.5,n)) {
  d1[i]=(a^4)/(1-a)^3+a
 #d2[i]=-a/(1-a)*(d1[i]-1)
  d2[i]=a-a^5/(1-a)^4
  i=i+1
}
i=1
for (a in seq(0.5,1,n)) {
  e1[i]=(2-a-1/a)^4/(1-a)^3+a
  e2[i]=-a/(1-a)*(e1[i]-1)
  i=i+1
}

l1=length(d1)
l2=length(e1)
for (i in 1:l1) {
  plot(d1[[i]], d2[[i]], xlim=c(0,1), col="red", ylim=c(0,1),
          xlab = "x", ylab = "y",xaxs="i", yaxs="i")
  par(new=T)
}
for (i in 1:l2) {
  plot(e1[[i]], e2[[i]], xlim=c(0,1), col="red", ylim=c(0,1),
          xlab = "x", ylab = "y", xaxs="i", yaxs="i")
  par(new=T)
```

```
  }
  max2=which.max(d2)
  min2=which.min(e1)

  lines(x,x,type= "l", col="red")
  lines(x,-x+1,type= "l", col="yellow")
  lines(x,-x^2+x,type= "l", col="red")
  lines(x^2-x+1,x,type= "l", col="red")
  #return(c(b1[max1],b2[max1],c1[min1],c2[min1],
        d1[max2],d2[max2],e1[min2],e2[min2]))
  return()
}


boundcubic<-function(n){

  ##### cube root boundaries

  b1 <- numeric()
  b2 <- numeric()
  c1 <- numeric()
  c2 <- numeric()
  i=1
  for (a in seq(0,0.5,n)) {
    b1[i]=-a^3/(1-a)^2+a
    b2[i]=(-a/(1-a))*(b1[i]-1)
    i=i+1
  }
  i=1
  for (a in seq(0.5,1,n)) {
    c1[i]=(2-a-1/a)^3/(1-a)^2+a
    c2[i]=(-a/(1-a))*(c1[i]-1)
    i=i+1
  }

  max=which.max(b1)
  min=which.min(c2)

  l1=length(b1)
  l2=length(c1)
  for (i in 1:l1) {
    plot(b1[[i]], b2[[i]], xlim=c(0,1), ylim=c(0,1),
            xlab = "x", ylab = "y",xaxs="i", yaxs="i")
    par(new=T)
  }
  for (i in 1:l2) {
    plot(c1[[i]], c2[[i]], xlim=c(0,1), ylim=c(0,1),
```

```
            xlab = "x", ylab = "y", xaxs="i", yaxs="i")
    par(new=T)
  }

  lines(x,x,type= "l", col="red")
  lines(x,-x+1,type= "l", col="yellow")
}


#matrix with random uniformly dstributed entries
myMat<-matrix(runif(n*n), ncol = n)
sums<-rowSums(myMat)
#normalized matrix
for (i in 1:n){
  for (j in 1:n){
    myMat[i,j]<-myMat[i,j]/sums[i]
  }
}

sum(diag(myMat)) #trace

### plot traces for multiple matrices
tr<-numeric()
for (k in 1:1000000){
  myMat<-matrix(runif(n*n), ncol=n)
  sums<-rowSums(myMat)
  for (i in 1:n){
    for (j in 1:n){
      myMat[i,j]<-myMat[i,j]/sums[i]
    }
  }
  tr[k]=sum(diag(myMat))
  k=k+1
}

d<-density(tr) #trace density
plot(d, xlim=c(0,n),col="red", xlab ="", ylab ="",
        xaxs="i", yaxs="i", main = "")
```

Bibliography

[1] R. B. Biritwum, S. I. Odoom. (1995). Application of Markov process modelling to health status switching behaviour of infants. *International Journal of Epidemiology*, 24(1), 177-182, 1995.

[2] P. H. Brill, M. Hlynka. A geometric interpretation of 2×2 Markov transition matrices. *Missouri Journal of Mathematical Sciences*,14, 159-174, 2002.

[3] M.-A. Guerry. Necessary conditions for the embeddability of discrete-time state-wise monotone Markov chains. Technical Report. Vrije University Brussels, 2012.

[4] M.-A. Guerry. On the embedding problem for discrete-time Markov chains. *Journal of Applied Probability*, 50, 918-930, 2013.

[5] M.-A. Guerry. Sufficient embedding conditions for three-state discrete-time Markov chains with real eigenvalues. *Linear and Multilinear Algebra*, 65(8), 1529-1539, 2017.

[6] Q.-M. He, E. Gunn. A note on the stochastic roots of stochastic matrices. *Journal of Systems Science and Systems Engineering*, 12(2), 210-223, 2003.

[7] J. H. Higham, L. Lin. On pth roots of stochastic matrices. *Linear Algebra and its Applications*, 435(3), 448-463, 2011.

[8] P. Lancaster, M. Tismenetsky. *The Theory of Matrices*, 2nd edition. Academic Press, 1985.

[9] M. Malik, L. C. Thomas. Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28, 261-272, 2012.

[10] S. M. Ross. *Introduction to Probability Models*, 10th edition. Academic Press, 2009.