

# Computing MDR-Game with Applications to Real Data

<sup>1</sup>Jimbo Henri Claver, <sup>2</sup>Takeru Suzuki, <sup>3</sup>Jesus Pascal,  
<sup>4</sup>Charles Awono, <sup>5</sup>Gabriel Andjiga, <sup>6</sup>Achille Mbassi and <sup>7</sup>Pascal Foumane

<sup>1</sup>Department of Science, American University of Afghanistan and Waseda University, Tokyo, Japan

<sup>2</sup>Department of Applied Mathematics and Statistics, Waseda University, Tokyo, Japan

<sup>3</sup>Department of Science and ICT, American University of Afghanistan, Afghanistan

<sup>4,5</sup>Department of Applied Mathematics, University of Yaoundé, Yaoundé, Cameroon

<sup>6</sup>Department of Urology, Yaoundé Central Hospital, Yaoundé Cameroon

<sup>7</sup>Department of Genecology, Geneco-Obstetric Hospital, Yaoundé, Cameroon

## Article history

Received: 12-04-2018

Revised: 14-06-2018

Accepted: 12-07-2018

## Corresponding Author:

Jimbo Henri Claver  
Department of Science,  
American University of  
Afghanistan and Waseda  
University, Tokyo, Japan  
Email:jimbo.maths@gmail.com

**Abstract:** Originally developed for probability and statistic purposes, Multiple Dice Rolling (MDR) game turns out to be useful for many other applications such as complex biological systems and medical research. It has been mainly used to estimate system states that cannot be observed directly. MDR game is affected by noise and fluctuations. It is widely agreed that noise in dynamic systems is captured by using the available computational and experimental technics, which provide information on the dynamics, size and entropy. Recently, computational advances have used algorithmic strategies and focused on stochastic accuracy to predict noise level in a given systems. It is known that experimental techniques can fully characterized the noise pattern and in some cases predicted relatively good outcomes for the MDR game. In this research, we developed a stochastic theory based on the MDR game and applied real data from to the Ciona and Beer-Tavazoie datasets. We found that the Ciona data (cell biometry) was relatively stable and the Beer-Tavazoie data (gene expression) was noisy. These facts support the well-known biologists' theories that "cells are stable, but genes which are components of cells are unstable". Furthermore, we have, for the first time, demonstrated quantitatively the molecular biology dogma. This result can be further developed in a novel direction to further understand the disease prediction, control and aging monitoring at the molecular and genetic levels.

**Keywords:** Dynamic Systems, Noise, Biodynamics, Simulation, Cell Biometry, Gene Expression, Cell Dynamics, Data, Analysis, Entropy, Diseases Modelling, Predication and Control

## Introduction

Noise in Biology is a product of interdisciplinary revolution that is touching nearly every scientific field where system behavior emerges from complex interactions of inhomogeneous cellular components (Backwell and Girshick, 1954; Feistel and Ebeling, 1989; Aumann, 1989). In biodynamic systems, for example, the benefits of low noise compete with those of low population and the result is the evolution of the architecture that produce an uneven distribution of

"stochasticity" across the components and in some cases the beneficial use of noise. Therefore, the frontiers of noise Biology lie across two interconnected fronts: (i) Understand the source, processing and consequences of stochastic fluctuations in biological systems and (ii) characterize the system architecture that has evolved to produce robust function within these noisy environments. Assuming that noise in random biodynamical systems can be effectively captured by MDR games, we consider a hypothetical model describing a collection of  $N$  cells by  $N$  generated spins

on natural collisions of dice sides in a particular games. We define the MDR game as longstanding and pervasive due to the infinite richness such a game (Krasovskij, 1970; Mayerson, 1991). Quantitative understanding of dice games, both in terms of probability theory and statistical laws, remains difficult and unclear (Krasovskij, 1970). Investigations are complex due to the shortage of reliable data on the one hand and the captured randomness on the other hand. Information about outcomes of the game is often difficult to quantify and not available in large numbers (Nelson, 1985; Pearl, 1988). While for a single dice rolling game the probability mass function (p.m.f) and the statistic of the game could be easily calculated, not much is known about mechanism of formation of these probabilities in the case of MDR games (Jimbo *et al.*, 2017). With the recent advances in computer science and the appearance of extensive databases, it is now possible to design MDR games with reasonably high numbers of dice accessible to quantitative analysis even for large numbers of throws (Iverson *et al.*, 1971; Knizer, 1999). As has pointed out in (Murray, 2002), one of the major problems with the MDR game is the difficulty to simultaneously follow the dynamic of all sides of the dice over time. To overcome such a limitation in this study, we assume without loss of generality that we throw the dice and follow the dynamic separately. By studying the dynamic of the probability distribution of each side of a die over many throws, we intend to estimate the Average Probability Value (AVP) of each side of the dice at a fixed number of throws. These values may then be used to predict the future outcome of the game. The MDR game problem is subject to stochastic perturbations that may produce delays that are experienced by individual dice during throws. Those delays can be due to the existence of random collisions and/or propulsions of sides in such dynamic process. Additionally, we believe that the way the stochastic fluctuations propagate to dice is the result of the propensity allocation and parameterization of the model. In such conditions, the noise will no longer be independent across dice. On that account, the effect of stochastic fluctuations cannot simply be understood as to be redundant. As we study the distribution of dice sides subject to constant stochastic fluctuations that randomly affect the delays experienced by individual side over throws, we must not forget about parsimony in designing the model. Our model is presented in detail in the next sections of the paper. We develop a game-theoretic model by introducing a modified version of the “dynamic probability concept” of dice sides over time and examine its statistical properties. The central aim of this work is to formulate the dice game problem in the framework of Markov processes also known as stochastic games (Maitra and Sudderth, 1996). While

many methods could be applied to this task, we chose the probabilistic approach because of its natural way to deal with games and ability to encode arbitrary dependencies between variables (Backwell and Girshick, 1954). Many authors have dealt with stochasticity in game theory; (Kolmogorov, 1950) built the fundament of probability theory and developed the dynamic concept of probabilistic games. (Feller, 1957; Erdős, 1961; Ferguson, 1969; Conway, 1976) introduced, respectively, the formulation of stochastic games and developed a more general and comprehensive approach to such games. (Kac and Logan 1976; Nelson, 1985; Morris, 1994), the Bellman equation of the problem and the computation of optimal states were formulated.

In the present work, we follow the idea of Kolmogorov and develop a probabilistic approach to MDR game. We extend the Chapman-Kolmogorov concept of “dynamic probability”, which in this case includes not only the state and time dependence but also the model parameter updates (Kolmogorov, 1950). This will enable our model to adapt to the changing condition of the game. We were able to draw a significant conclusion on how often individual sides of dice may appear after a large number of throws. Another important remark is that if the initial data is less noisy, after a reasonably large number of iterations (1000, for example), the Central Limit Theorem (CLT) can be applied. But in contrast, if the initial data is unstable, the CLT is only partially achieved, or in some cases, it will never be achieved (Jimbo, 2004). Again to support the choice of our approach, it is important to mention that although in many studies, we are scientifically more concerned in obtaining the entire distribution of the stochastic game. Rather in this study, we will focus in obtaining the distribution of each random variable. While the computation of the marginal probability distribution of a die is useful in evaluating the stochastic model, we believe that information regarding such distribution may not be good enough for individual side prediction. In this respect, we need to develop an approach, which could precisely evaluate the probability of all sides at a given throw for an optimal characterization of the game (Nelson, 1985; Morris, 1994; Roth and Sotomayor, 1990; Alon and Krilievich, 2006; Jimbo *et al.*, 2017).

## Methodology

Based on the computed probabilities, the game-theoretic model can now be formulated in terms of the Chapman-Kolmogorov Equations (CKEs), also called the master equation. The transition state reflects the change in the outcome after each throw. Until now, such modelling approaches used in many problems do

not take into account the changes in the allocation of preferences over time; such changes can be efficiently captured by the propensity of the game. In this study, we proposed a new version of the CKEs. We assume that the drift term in the equation has an independent dynamic, which can be captured by a simple regression model and that each dynamic is conditioned on the initial state and time. The model is called the Extended Chapman-Kolmogorov Equations (ECKEs) and is given by:

$$\begin{cases} \frac{\partial f(x, t/x_0, t_0)}{\partial t} = \sum_{j=1}^N a_j(x - v_j) f(x - v_j, t/x_0, t_0) \\ -a_j(x) f(x, t/x_0, t_0) \\ a_j(x) = \vartheta_1 a_{j-1}(x) + \theta_1 \end{cases} \quad (1)$$

Where:

- $n$ : Total number of dice (unbiased)
- $N$ : Total number of throws (rolling)
- $k$ : Number of sides of a die
- $j$ : Number of throws in the game
- $x = x_k$ : Event "side  $k$  appears over throws"
- $f(.) = f_k(.)$ : Probability mass function of side  $k$
- $f^t(.)$ : Transpose of  $f(.)$
- $a_j(x)$ : Propensity for side  $k$  of the  $j$ -th throw
- $a_0(x)$ : Initial propensity of side  $k$ ;
- $v_j(x)$ : State change associated with a single event at the  $j$ -th throw
- $a_j(x)dt$ : Probability that  $k$ -th side appears in the interval  $[t, t + dt]$  at the  $j$ -th throw
- $\rho(x, a_j(x))$ : Correlation between an event at the  $j$ -th throw and the corresponding propensity

We assume for convenience that each side of each dice is correlated with the parameter of the game over all iterations and set the following:

$$\begin{cases} 0.5 < \rho(x, a_j(x)) < 1 \\ 0.5 < \rho(x, a_j(x - v_j)) < 1 \end{cases} \quad (2)$$

As the choice of the propensity will closely depend on the related event, we assume that the absence of such boundary conditions on the correlations will affect the probability values, but yet we cannot confirm whether they will decrease or increase. This will be investigated in our next work. Furthermore, if for certain values of time  $t$  the function  $f$  is too hard to obtain, we propose a simple approximation by assuming the random variable  $x = x_k$  has approximately an exponential distribution:  $x \sim \exp(-\lambda_k)$ , or *p.m.f* of  $f(x, t) \sim e^{-\lambda_k t}$ , for every  $t$ , where we  $\lambda_k > 0$  is

the rate parameter of the distribution. Since  $\lambda_k$  is not constant over time, we can also approximate the first derivative of  $f(x, .)$  to  $\lambda_k e^{-\lambda_k t}$ . Based on some consideration of the *Fréchet Derivatives* (Fréchet derivative is the derivative of a function in a given direction) (Pearl, 1988), the Equation 1 becomes:

$$\begin{cases} \lambda_k \exp(-\lambda_k) = \sum_{j=1}^N a_j(x - v_j) \exp(-\lambda_k - v_j) - a_j(x) \exp(-\lambda_k) \\ a_j(x) = \vartheta_1 a_{j-1}(x) + \theta_1 \end{cases} \quad (3)$$

After simplification, this system of equations becomes:

$$\begin{cases} \lambda_k = \sum_{j=1}^N a_j(x - v_j) \exp(-v_j) - a_j(x) \\ a_j(x) = \vartheta_1 a_{j-1}(x) + \theta_1 \end{cases} \quad (4)$$

From the above formula, we compute  $\lambda_k$  as a function of the side number  $k$  and the iteration period  $j$ ; that is  $\lambda_k = \lambda_k(j)$ . We have that  $\vartheta_1$  and  $\theta_1$  are the constant parameters of the game, which are assumed to be known. In this particular case, we set  $\vartheta_1 = -0.05$  and  $\theta_1 = +0.05$  for convenience. Also we assume that the initial propensity  $a_k(x)$  is known and  $v_j$  is a vector with a finite number of elements. Since Equation 1 cannot be solved analytically for  $f(x, t)$ , we propose a numerical approach using (4); thus, the algorithm in explained in the next section.

## Algorithm

As the analytical solution to (1) is hard to obtain even for a moderate number of throws, a numerical algorithm using adapted stochastic simulation approach on Equation 4 is proposed in this study. In our algorithm, two random variables determine the temporal evolution of the game. The variable  $\tau_k$  is the time for the next event to occur (next side appearance of dice) and  $k$ , as before, is the side of the dice.

Furthermore, the probability density of an event  $x = x_k$  is then evaluated based on the propensity of the game on the event involved. Overall, this will give a better flexibility and applicability of the algorithm in some sense. The main purpose of creating such an algorithm is to simultaneously simulate the game and predict the online probability mass functions of each event in the game process using model 1.

One important remark here is that the probability mass function  $f$  at each throw is a vector because we throw many dice at once and each side has a certain probability of appearance. We use  $f$  transpose instead of  $f$  to ensure better readability of our outputs, but all properties remain the same. In addition, our algorithm

facilitates evaluation of the Average Probability Value (APV) of each side after a certain number of throws; this is important if we want to study the general trend of the game. Our algorithm is presented as follows:

**Start.** Input: Initial data  $x_0$

Outputs:  $f(x, t)$ ,  $\lambda$

1. Set  $a_0 := x_0$ .
2. For  $t = 1: n_t$  ( $n_t$  = number of iterations) do
  - a. Let  $\tau \sim Po(10)$  be the time until the next event, which changes over time
  - b. Let  $v_t \sim N(2, 0.05)$  be the change associated to a single event
  - c. Compute  $x_t = \theta x_{t-1} + \vartheta$  and  $a_t = \vartheta a_{t-1} + \theta$ .
  - d. Compute the value of the *pmf*  $f(x, t)$  based upon (4)
  - e. Compute  $\lambda_t = |a(x-v)e_{-v}-a(x)|$
3. Output  $f(x, t)\lambda$

**End**

In the next section, we will present some MDR game limitations and weaknesses.

## Limitations

The main problem with the MDR game model is the lack of robust test data in which the dynamic interactions are known in advance. We need the data in which the true correlations are known in advance if we want to compare our approach with the existing methodology. It is not possible to find the exact values for the probabilities first because the propensity is not unique and secondly because of the lack of clear error measurements in such a dynamic system. One way to relax these requirements is to introduce additional parameters, which will influence the generation of probabilities over time. Afterward, we build a new generation of flexible algorithms, which create test data sets and accurately predict outcomes of the game over time.

Furthermore, we propose a propensity allocation based approach using liner regression in order for our model to generate and predict the state of the game, but other researchers may prefer more complex models. Finally, we present the test datasets created by our technique, which can be used to assess future performance of algorithms that attempt to determine the underlying predicted states. The above algorithm is available and can be provided under requests. The cell biometry (Ciona Data) and gene expression (Beer and Tavarzoe Data) are used to test our results. However, as mentioned in (Iverson *et al.*, 1971; Nelson, 1985; Kreps, 1990; Jimbo, 2004; Jimbo *et al.*, 2017) the number of observations remain relatively low, which contradicts further generalization.

## Gaussian Process for the MDR Game

Let  $f_k(t)$  be a function that a side  $k$  of a die appears after  $t$  throws. We define such a function as a Gaussian process, i.e., the probability distribution over the function  $f_k$  that takes values  $f_k(t)$  at time  $t$  (probability mass function, *p.m.f.*, of a given side  $k$  of a die). Analogous to the Gaussian distribution, which is fully characterized by a mean and covariance, a Gaussian process is characterized by a mean function  $E[f_k(t)]$  and a covariance function  $cov(k, k') = E[f_k(t) f_{k'}(t')] - E[f_k(t)] E[f_{k'}(t')]$  where the expectation is over function evaluations at time  $t$  with side  $k$  and time  $t'$  with side  $k'$ . We notice that  $t$  may be equal to  $t'$ . Furthermore, we assume that the functions  $f_k(t)$  are stationary, differentiable and bounded. The outcome of the game is highly stochastic. In the next section, we will show how we consider a variant of the Chapman-Kolmogorov equations for the dynamic of  $f_k(n)$  over time.

## Test Dataset

The main motivation for creating a *Test Dataset* was to have a benchmark dataset for which we may presume to know certain regulatory rules. This dataset will be necessary to compare outcomes of games with various other datasets and the accuracy can be easily checked. This is because there is no experimental data of multiple dice rolling games for which the true underlying interaction and propensity allocation are exactly known. We, therefore, create a sample data set by permuting in a random manner the numbers 1 to 6 as many times as we need. We anticipate that the availability of the test dataset will allow researchers to evaluate their own methods and compare their methods against commonly used algorithms. With the test data, what we provide will be useful for researchers who want to get started right away testing their algorithms; we emphasize that the real power of the proposed algorithm is the capability it provides to quickly produce necessary outcomes for a game when a good allocation of propensity is made. Researchers can now use their own test datasets to compare the dependency of any method on any particular parameter (number of side of dice, type of correlation, type of data and type of propensity).

## Experiments

We throw 1000 dice, six sides each, 1000 times. Each side will randomly appear over time with some probability  $p.m.f = f_k(x, t) \sim f_k(x, n)$  with  $t \rightarrow n$ .

In the tables below, we will present some results of the simulations.

Table 1 shows the minimum and maximum average probability values for each side of a dice to occur over time were obtained by our algorithm. As can be seen, side five has the smallest variance, but side two has the smallest mean. Sides four and six have the highest mean. From Table 1, we observe that the best Average Probability Value (APV) of the occurrence of a given side after  $n$  throws is  $p_{av} = 0.265$ .

Table 2 represents variance - covariance matrix of a die, where each non-diagonal entry  $(i,j)$  gives the covariance

between sides  $i$  and  $j$ . The diagonal entries are the variance of each side over all iterations. We observe evidence of weak correlation among all sides, which is reasonable and acceptable as expected for such game.

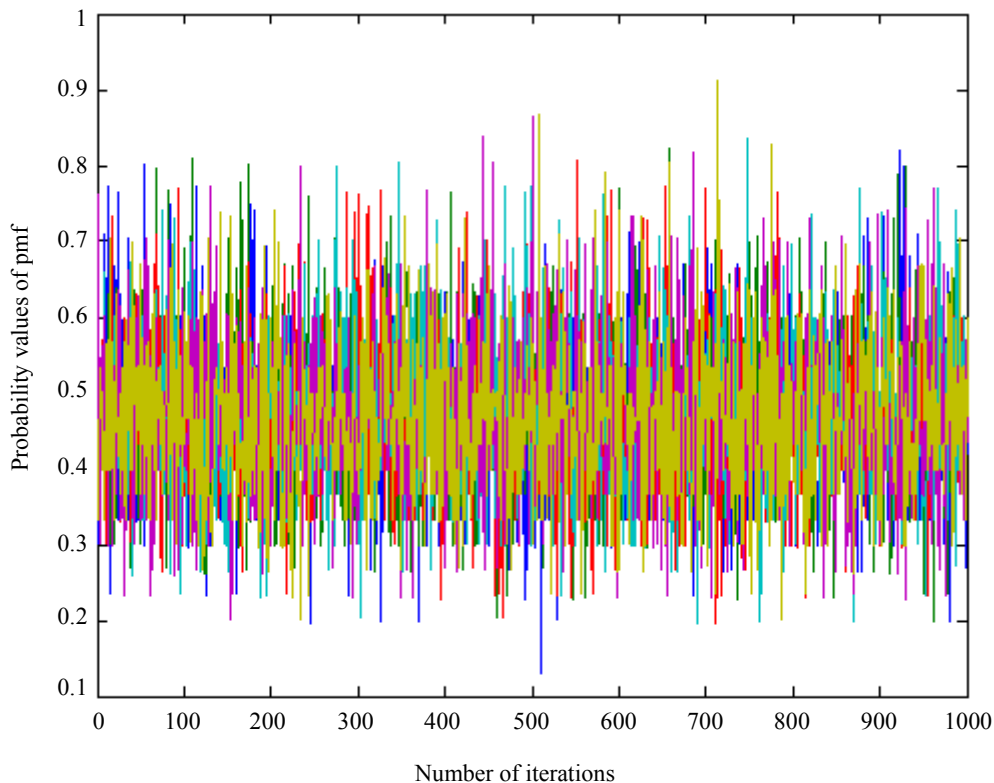
Figure 1 represents sample paths of the probability values for all dice sides with the Permutation Dataset. We assign a color to each side; the blue color represents side 1, green side 2, red side 3, grey side 4, violet side 5 and finally, yellow side 6.

**Table 1:** Minimum and maximum average probability and variance of data

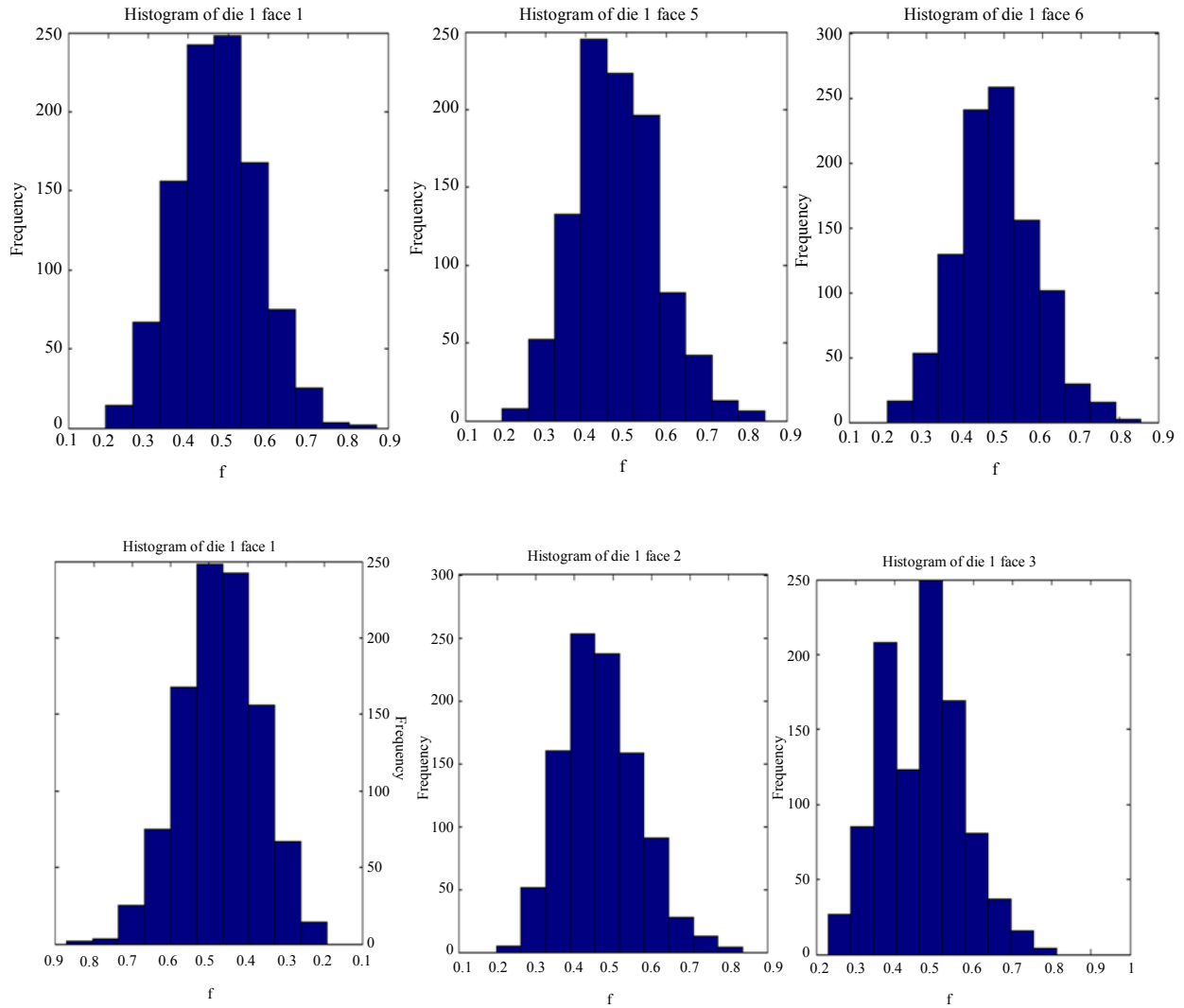
| Side | Min    | Max    | Variance   | APV    |
|------|--------|--------|------------|--------|
| 1    | 0.2428 | 0.2838 | 0.00004642 | 0.2651 |
| 2    | 0.2446 | 0.2948 | 0.00004867 | 0.2649 |
| 3    | 0.2447 | 0.2882 | 0.00004598 | 0.2650 |
| 4    | 0.2442 | 0.2918 | 0.00004428 | 0.2653 |
| 5    | 0.2456 | 0.2917 | 0.00004322 | 0.2650 |
| 6    | 0.2410 | 0.2893 | 0.00004391 | 0.2653 |

**Table 2:** Variance - covariance matrix of a die

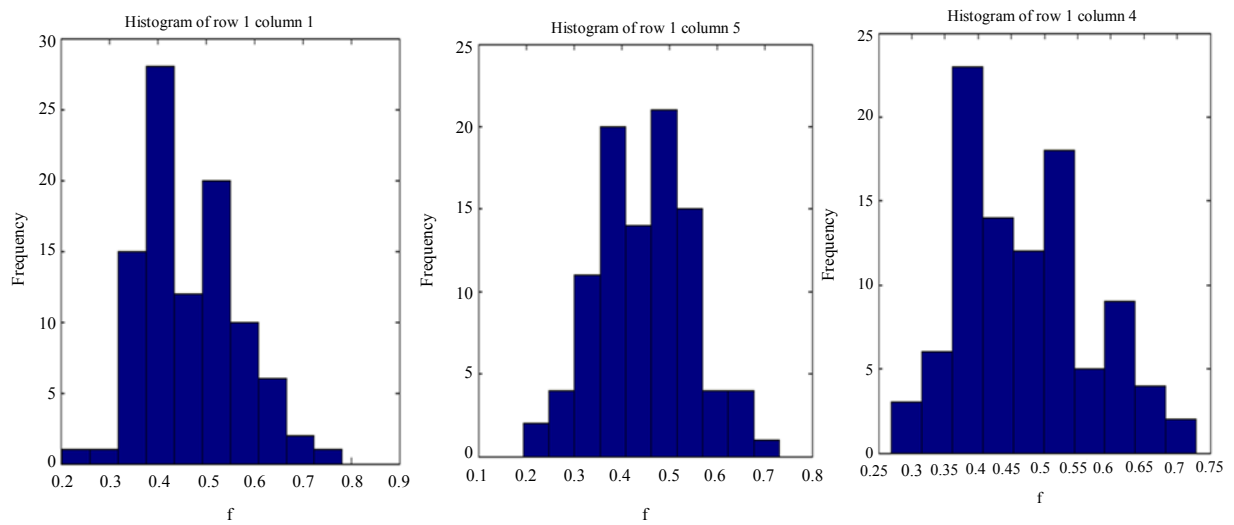
|          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|
| 0.007248 | 0.006986 | 0.006919 | 0.007226 | 0.007438 | 0.007014 |
| 0.006986 | 0.006682 | 0.006861 | 0.007246 | 0.007062 | 0.006616 |
| 0.006919 | 0.006861 | 0.006733 | 0.006895 | 0.006930 | 0.006694 |
| 0.007226 | 0.007246 | 0.006895 | 0.007161 | 0.007128 | 0.006842 |
| 0.007437 | 0.007062 | 0.006930 | 0.007218 | 0.007274 | 0.007105 |
| 0.007013 | 0.006616 | 0.006694 | 0.006642 | 0.007105 | 0.006723 |

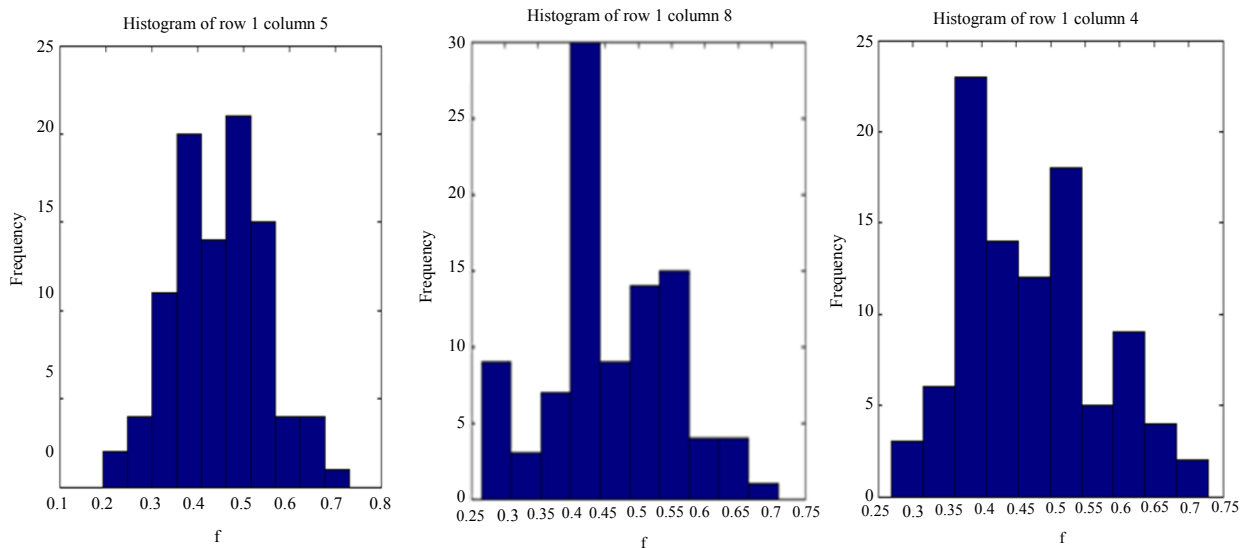


**Fig. 1:** Sample paths and probability values of all dice sides



**Fig. 2:** Histogram of each side of a dice over generation





**Fig. 3:** Histogram of six biometrics of Ciona cells

Figure 2 shows histogram of each side of a dice. It also highlights the change in the *pmf* over throws. The overall sample paths of the probability values and the *pmf* of each side at each throw are available in additional file that can be downloaded on to test our model, we decide to apply it to two real biological datasets in the next section.

The sample paths of probability values, the *pmf* and all histograms are available upon request (Fig. 3).

### Application to Real Biological Data

We applied our stochastic based-modelling approach to the MDR game to predict the likelihood of the outcome of each dice side over time. At the beginning, we use six-sided regular dice to generate synthetic initial data and create a benchmark set of probable values of the outcomes. We, further, test our results using the Ciona and Beer-Tavazoie datasets as initial data. We observed that our model may fit very well to the Ciona data, but there was a slight overfit when applied to the Beer-Tavazoie data suggesting some specificity in such data. The overfitting in the case of the Beer-Tavazoie data was explained by the presence of excessive noise in such data. Furthermore, we examined the dynamic correlation and stability of various sides of a chosen die in the game and found that in the Ciona data were more stable than Beer and Tavazoie data over iterations. Another advantage of our approach was speed, especially as we tested it by running many simulations, for real biological data, which takes more time to generate than synthetic data; we sacrificed some speed for higher accuracy and reliability. We concluded that our approach works well on our synthetic data generated according to the simulation of a real dice game. We also obtained an indication of its ability to find highly statistically

significant results in real biological datasets. The main motivation for using real biological data is to test the efficiency of our approach in the application. We applied our algorithm respectively to the Ciona embryo biometry and Beer-Tavazoie gene expression datasets; the results are presented below.

#### A. Cell Biometry - Ciona Embryo Biometry Data

We used the Ciona embryo data from the link below:

<http://crfb.univ-mrs.fr/aniseed/embryo-collection.php>

#### Remark 1

We can clearly observe that  $\text{var}_{av(cio)}$  and  $\text{var}_{av}$  are in the same range, indicating a perfect match in the measure of variability for both datasets. Where  $\text{var}_{av(cio)}$  and  $\text{var}_{av}$  represent respectively the average variance of gene expression dataset and the average variance of the simulated data set.

In the Table 3 it can be seen that, the best average probability value “avp” of the occurrence of a given side after  $n$  throws is  $p_{av} = 0.2708$ . This result indicates that if we were to bet on the outcome of a given dice after  $n$  throws, we must choose a probability value around  $p_{av}$ . Comparing  $p_{av}$  to  $p_{av}$ , we then obtain a marginal error of 0.72%, suggesting that our model can be effectively used to explain and capture the randomness in the Ciona biometry data.

In the Table 4 the non-diagonal entries are: the covariance between side  $i$  and biometry  $j$ . The diagonal entries are the variance of the  $i$ -th biometry of cell. Comparing Min and Max values of *p.m.f.*, we found that Min (Min) and Max (Max) are always in the same ranges. These findings suggest that the order can be formed from disorder, especially in biodynamic systems. Also, capturing randomness in the Ciona biometry data is essential.

**Table 3:** Dynamic of “avp” values over time for Ciona dataset

| Side | Min    | Max    | Variance   | APV    |
|------|--------|--------|------------|--------|
| 1    | 0.2460 | 0.7510 | 0.00250800 | 0.2697 |
| 2    | 0.2491 | 0.2899 | 0.00005115 | 0.2656 |
| 3    | 0.2511 | 0.4211 | 0.00030000 | 0.2669 |
| 4    | 0.2459 | 0.4995 | 0.00063040 | 0.2667 |
| 5    | 0.2473 | 0.8280 | 0.00335700 | 0.2700 |
| 6    | 0.2489 | 0.3906 | 0.00020870 | 0.2656 |
| 7    | 0.2487 | 0.7667 | 0.00265500 | 0.2708 |
| 8    | 0.2517 | 0.2832 | 0.00005105 | 0.2656 |

**Table 4:** Variance - covariance and matrix of cell biometry prediction after 100 iterations

|          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|
| 0.006330 | 0.006759 | 0.007390 | 0.005933 | 0.005761 | 0.006912 | 0.006635 | 0.005501 |
| 0.006759 | 0.006149 | 0.007286 | 0.006379 | 0.006572 | 0.005942 | 0.006489 | 0.005702 |
| 0.007390 | 0.007286 | 0.007560 | 0.006869 | 0.006706 | 0.007245 | 0.006931 | 0.006578 |
| 0.005933 | 0.006379 | 0.006869 | 0.006362 | 0.006636 | 0.007858 | 0.006776 | 0.006436 |
| 0.005761 | 0.006572 | 0.006706 | 0.006636 | 0.006414 | 0.006844 | 0.006052 | 0.005669 |
| 0.006912 | 0.005943 | 0.007145 | 0.007858 | 0.006844 | 0.007160 | 0.006491 | 0.006729 |
| 0.006635 | 0.006489 | 0.006931 | 0.006776 | 0.006052 | 0.006491 | 0.006751 | 0.007334 |
| 0.005501 | 0.005702 | 0.006578 | 0.006436 | 0.005669 | 0.006729 | 0.007334 | 0.006029 |

**Table 5:** Minimum and maximum average probability values

| Side | Min    | Max    | Variance   | APV    |
|------|--------|--------|------------|--------|
| 1    | 0.2528 | 0.2746 | 0.00003715 | 0.2638 |
| 2    | 0.2551 | 0.2827 | 0.00005207 | 0.2653 |
| 3    | 0.2547 | 0.2825 | 0.00006549 | 0.2654 |
| 4    | 0.2503 | 0.2742 | 0.00003441 | 0.2643 |
| 5    | 0.2510 | 0.2766 | 0.00005161 | 0.2641 |
| 6    | 0.2504 | 0.2807 | 0.00006210 | 0.2663 |
| 7    | 0.2534 | 0.2789 | 0.00004886 | 0.2658 |
| 8    | 0.2544 | 0.2808 | 0.00005535 | 0.2660 |

**Table 6:** Variance - covariance matrix of genes expression levels/conditions

|          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|
| 0.006451 | 0.005405 | 0.007751 | 0.007778 | 0.006855 | 0.009141 | 0.007652 | 0.007923 |
| 0.005405 | 0.004138 | 0.005522 | 0.004195 | 0.005579 | 0.006986 | 0.004782 | 0.008410 |
| 0.007751 | 0.005522 | 0.008101 | 0.006632 | 0.008874 | 0.007422 | 0.005767 | 0.007410 |
| 0.007778 | 0.004195 | 0.006632 | 0.005630 | 0.007684 | 0.006929 | 0.006274 | 0.006479 |
| 0.006855 | 0.005579 | 0.008874 | 0.007684 | 0.008483 | 0.007586 | 0.006447 | 0.007643 |
| 0.009141 | 0.006986 | 0.007422 | 0.006929 | 0.007586 | 0.007812 | 0.006327 | 0.001051 |
| 0.007653 | 0.004782 | 0.005767 | 0.006274 | 0.006447 | 0.006327 | 0.005167 | 0.006448 |
| 0.007923 | 0.008410 | 0.007410 | 0.006479 | 0.007643 | 0.001051 | 0.006448 | 0.008233 |

Table 5 shows that minimum and maximum average probability values obtained from our algorithm are respectively 0.260 and 0.2638 for eight gene expression conditions over generation time. As can be seen, all average probabilities are around 0.2650 with a margin error of 0.22%. This result shows that margin errors in gene expressions are far smaller than in cell biometry.

Table 6 shows variance, covariance and matrix of genes expression levels/conditions, where the non-diagonal entries are the covariance between the *i*-th and *j*-th expression conditions. The diagonal entries are the variances of expression conditions over time. Also values are in the same range of 3 digits.

Figure 4 presents the histogram of each expression condition of six genes. Additionally, the sample paths of probability values and the *pmf* of each expression

condition over many throws are available in the supplementary materials, available upon request.

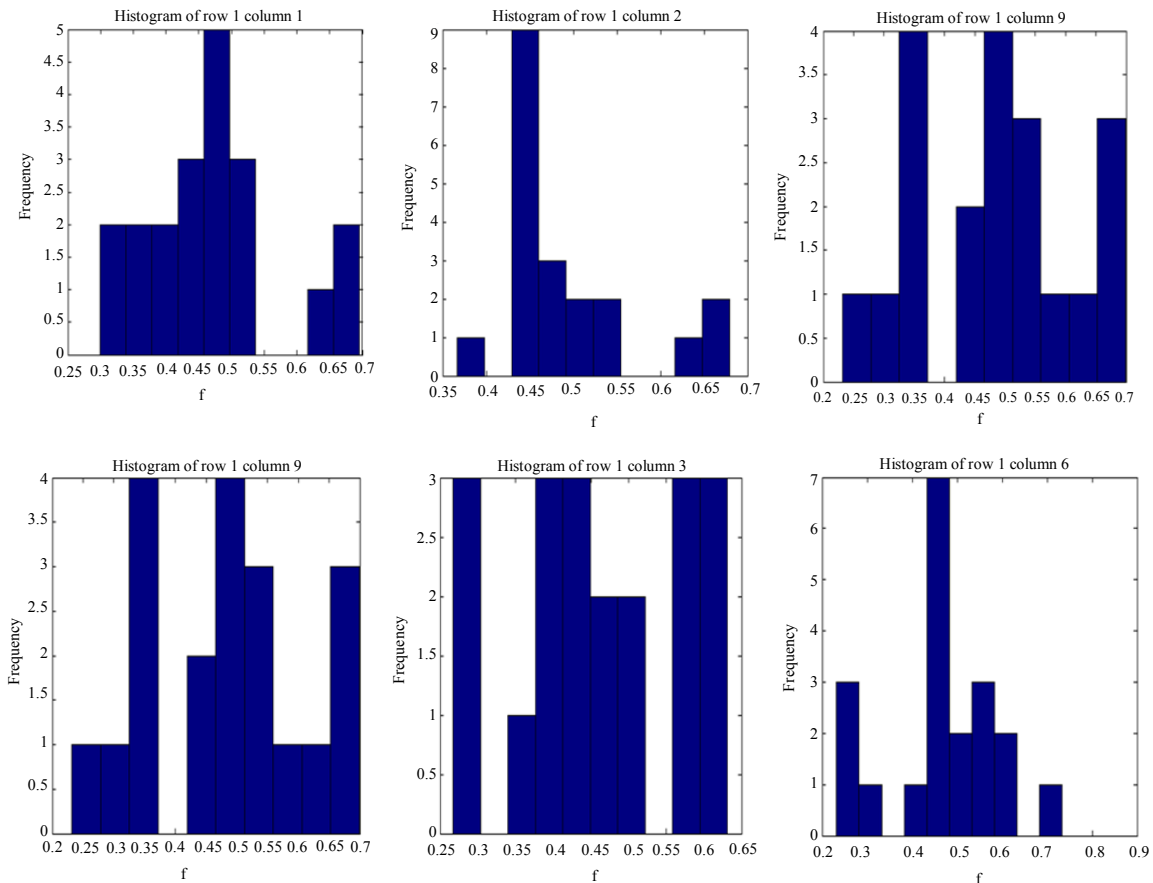
### B. Gene Expression - Beer-Tavazoie Data

We used the gene expression data from the following website:

[http://genomics.princeton.edu/tavazoie/web/Beer\\_Tavazoie\\_Supp\\_Mat/expression\\_data.TXT](http://genomics.princeton.edu/tavazoie/web/Beer_Tavazoie_Supp_Mat/expression_data.TXT).

We rearranged the data in a matrix form, rows representing genes (dice) and columns the expression conditions (sides of dice). We obtained a data matrix with 6189 rows and 255 columns, meaning 1,578,195 total entries. We next applied our mathematical model and computational algorithm to the data, obtaining the following results:





**Fig. 4:** Histograms of each expression condition of six genes

*Remark 2*

We can easily observe that  $var_{av(ge)}$  and  $var_{av}$  are in the same range indicating a perfect match in the measure of variability for both datasets. Where  $var_{av(ge)}$  and  $var_{av}$  represent respectively the average variance of gene expression dataset and the average variance of the simulated data set.

Now, comparing  $p_{avBT}$  with  $p_{av}$ , we obtain a marginal error of 0.03%, suggesting that our model can be used to explain and capture the randomness in the Beer-Tavazoie gene expression data with a small error (as we preferentially focus on the most frequent outcomes).

**Measures of Dynamic Correlation and Stability**

To understand the level of stability in each biological process (cell biometry and Gene expression), we introduce two new measures:

i. *Dynamic Correlation:*

$$d_t = |\text{corr}(k, k'; nt) - \text{corr}(k, k'; n(t-1))| \tag{5}$$

where,  $\text{corr}(k, k'; nt)$  is the correlation between even  $k$  and  $k'$  at time  $t$  over  $n$  throws.

ii. *Dynamic Stability*

$$d_s = \frac{M(d_s)}{Var(d_s)} \tag{6}$$

$$s = 1, \dots, N$$

**Comparison of the Stability in Ciona and Beer-Tavazoie Data**

All cases presented in this section are representative; full details are presented in the supplemental material on stability. We calculate  $d_t$  and  $d_s$  for both the Ciona and Beer-Tavazoie data. The above figures confirm that cell biometry is more stable than gene expression. It can be seen that the Beer-Tavazoie data is a mixture of stable and unstable genes, even if the whole system is unstable. Now, we wish to understand the level of fluctuations in our MDR game and for that end we need a measure to evaluate noise strength.

## Noise Strength Analysis

We define *noise strength* as  $NS(k)$  of the dynamic system (or game) as follows:

$$NS(k) = \frac{1(1-p_k)}{n p_k} \quad (7)$$

with:

$$k = 1, \dots, N$$

where,  $n$  is the number of throws,  $k$  is a die side and  $f = p_k$  is the probability that side  $k$  appears on the  $n$ -th throw. The derivation of the above formula is presented in the supplemental material. We use the formulation of  $NS(k)$  to understand how the number of iterations and the probability values affect the noise strength. The obtained figures with  $n=1000$ ,  $n=500$ ,  $n=100$  and  $p=0.95$ ,  $p=0.50$ ,  $p=0.25$  show that as  $n$  increases, the noise strengths overall decreases.

## Characterization of MDR Game

### Robustness

The proposed model will not be robust against a game that predicts or influences a dice. If we assume that the

roller can influence the throw of the dice or use any other trick to influence the outcome, such games will not be robust. But, if we consider that the rolling conditions remain the same and the dice are identical, then the game becomes robust.

### Simplicity

Our model for MDR game is extremely straightforward. Additional constraints on some properties of the game may decrease simplicity, but the MDR game is very clear at the beginning.

### Efficiency

Our model efficiency will depend on the initial condition of the game, the propensity, the number of iterations and the number of dice rolled.

As the number of throws or iteration of dice decreases, the noise strength increases for presumably fixed probability. As the probability of appearance of a side of dice decreases, the noise strength increases for a fixed number of iterations (Fig. 5).

Figure 6 Ciona entropy is generally within the interval (2.4, 2.9) and Beer-Tavazoie entropy varies within the interval (86, 89). Clearly Ciona has lower entropy over all iterations and, therefore, higher probability values or less randomness in contrast to the Beer-Tavazoie data which has higher randomness.

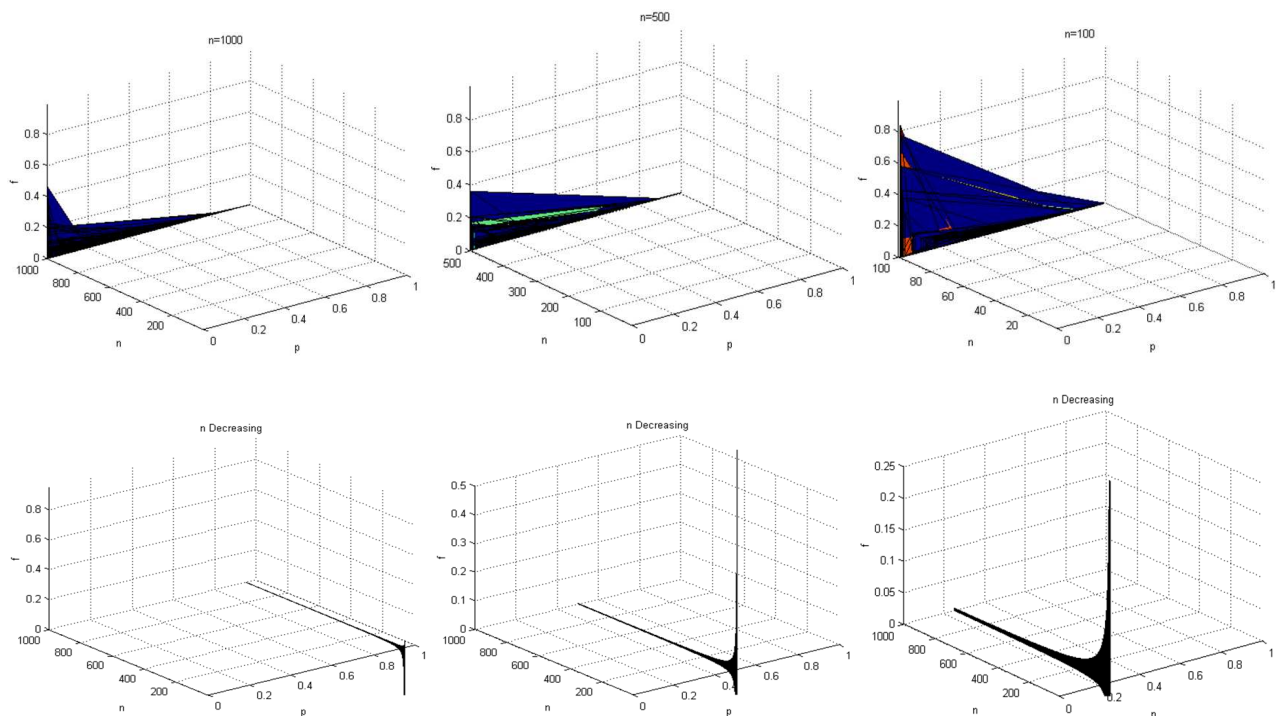
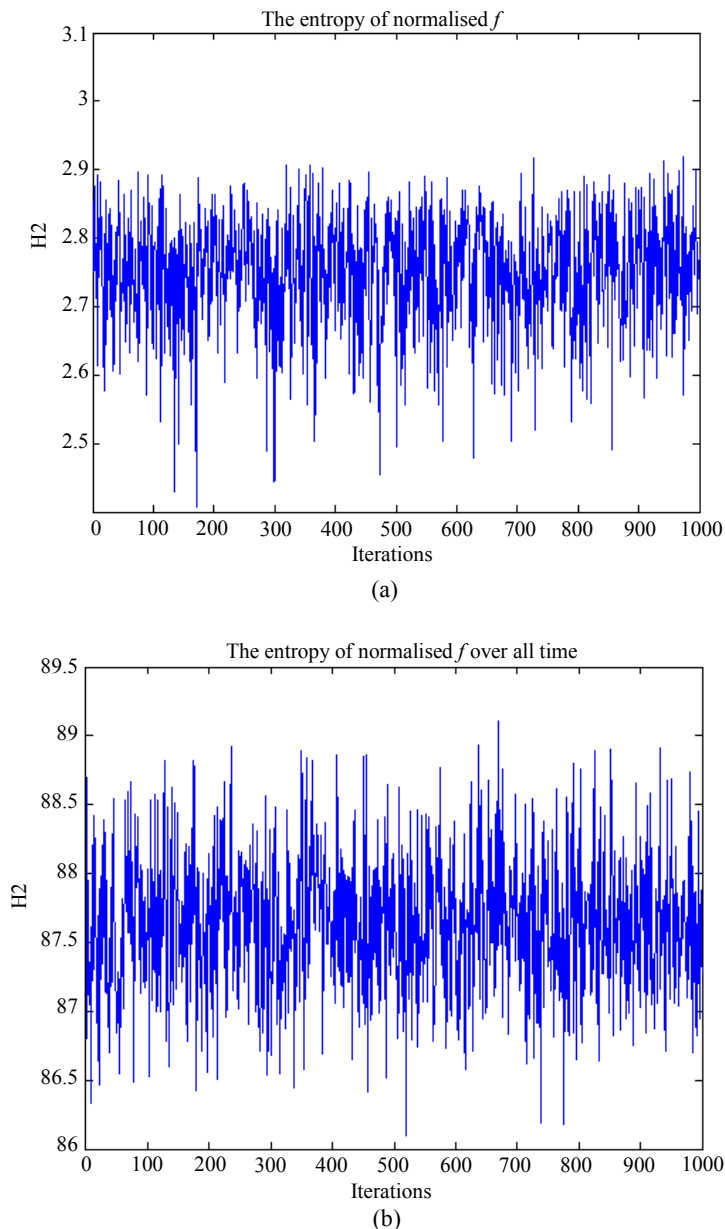


Fig. 5: Noise Vs number of throws and evaluation results of noise strength



**Fig. 6:** Ciona and Beer-Tavazoie entropy of data (a) Ciona Entropy (b) Beer-Tavazoie Entropy

### Entropy over Iterations

By definition, dice entropy  $H(x_k)$  is the measure of uncertainty or energy associated with the appearance of the side  $x_k$  after  $n$  throws; it is given by the formula:

$$H(x_k) = \sum_{j=1}^n f_j(x_k) \log f_j(x_k) \quad (8)$$

where:

$$k = 1, \dots, N$$

and:

$$n = 1, \dots, \max(\text{iterations})$$

In this study, we calculate the entropy for Ciona and gene expression datasets and present the results in the figures that follow:

#### Remark 1

We observe in general that  $H_{Cio}(x_k) < 30$  and  $H_{B\&T}(x_k) < 90$  indicating that there are more fluctuations in Beer - Tavazoie gene expression data comparing to Ciona cell biometry data.

## Discussion

We have shown using a mathematical model that the stochasticity in the system game can be captured by MDR game. The extended version of the *Chapmann-Kolmogorov Equations* (CKEs) we used is still a simplification and could even be unrealistic for certain designs of the game, but it can capture the randomness of the game if we use good initial conditions and reasonably stable parameters. The main novelty of our approach is the use of the extended CKEs and the propensity of the game applying our algorithm to predict the average probability of appearance of each side of a die after a fixed number of throws. However in practice, when using MDR game to solve real world problems, it is not enough to find the average probability value and fix it. Again, all outcomes in our model depend on the initial inputs and the parameters. We, then, have to continuously update new probability values as dice are thrown. We face the challenge of iteratively improving the solution to a series of problems that change over time and we must do this in light of unanticipated events and incomplete knowledge. The fact that *uncertainty* and *inaccuracy* enter into our model, prediction must be accepted and even used to the great possible benefit.

In summary, we conclude that our method can efficiently predict the future evolution of biological systems, which are not chaotic. We have limited ourselves to evaluating the average probability of sides of a die with preset parameters and fixed initial conditions. The Gaussian process approach can also be used to compute the likelihood for the model (Kac and Logan, 1997; Knizer, 1999), but since our model is inherently nonlinear, the likelihood is no more tractable. Therefore, Bayesian inference and even more sophisticated statistical methods could be required. However, it seems doubtful that a more complex mathematical model or techniques may be needed to calculate the probability or likelihood that a face of dice appears after a fixed number of throws. We are pursuing a number of extensions of the current results because it is unlikely that these probabilities will remain the same over time (iteration) due to possible external (initial conditions, additional parameters) and internal (dynamic correlation, noise strength and entropy) variabilities at each throw.

## Conclusion

The proposed algorithm has made it possible to detect interactions among all dice sides as dynamic correlations and evaluates the *p.m.f* of a side to appear over many throws. We use the variance-covariance matrix to extract such information in the game over time. The input of our algorithm is associated to a matrix of columns, which represent the dice thrown and rows representing the faces of dice. The outcome of games in

terms of the prediction or *p.m.f* values greatly depends on the propensity and the parameters of our model. We found that the time complexity of the proposed algorithm was polynomial of order  $O(n^2)$  where  $n$  is the number of data points. We showed the performance of the method by applying it to two typical real datasets. We found similarities in *p.m.f* values obtained from synthetic and real datasets. This suggests that it is possible to predict the dynamic of the real dataset using our method. Finally, we calculate the entropy and evaluate stability in the Ciona and Beer-Tavazoie datasets and observed that the Ciona data was more stable (less variable) with low entropy, which means higher probability values and, therefore, lower uncertainty. In contrast, the Beer-Tavazoie data was noisier with high entropy and, therefore, low probability values and higher uncertainty.

In conclusion, we have developed a stochastic theory for the MDR game. This novel theory was for the first time applied to the Ciona and Beer-Tavazoie real datasets. We found that the Ciona data (cell biometry) was relatively stable and the Beer-Tavazoie data (gene expression) was noisier. These facts confirm the well-known biologists' intuitions that cells are stable, but genes which are components of cells are unstable. Furthermore, we demonstrated quantitatively the cellular dogma. Future direction to our work will be to develop a unified theory between game theory and stochastic theory.

## Acknowledgement

This paper was stimulated by discussion during meetings at NACA 2016, Niigata, Japan. The research presented in this study was supported by the Sakura Research, Waseda University and AUAF. We thank Dr. Sakumura for fruitful discussions and providing us with the ANISEED webpage where we could find the Ciona embryo dataset. We equally thank Hamidullah Hamidy for reviewing and editing the final version of this manuscript. The initial test data was used for simulation and analysis of the MDR game in a real life situation. This publication reflects only the authors' views and any remaining mistakes are ours.

## Availability

Complete Matlab code for repeating the experiment is available at <http://www.naist.bs.sakumLab>

## Author Contributions

**Jimbo Henri Claver:** Designed and wrote the model.

**Takeru Suzuki:** Implemented part of the algorithm and the Matlab code.

**Achille Mbassi, Jesus Pascal, Charles Awono, Gabriel Andjiga and Pascal Foumane:** Contributed with their respective expertise in this research work.

**Jimbo Henri Claver:** Supervised the work as a whole and drafted the manuscript. All authors read and approved the final manuscript.

## Ethics

There are no ethical issues with this article.

## References

- Alon, N. and M. Krilievich, 2006. Extremal and probabilistic combinatory. *Canad. J. Math.*, 11: 34-38.
- Aumann, R.J., 1989. Lecture note on game theory. Westview Press Inc., Boulder, Colorado.
- Backwell, D. and M.A. Girshick, 1954. *Theory of Games and Statistical Decisions*. 1st Edn., John Wiley and Sons, New York, pp: 355.
- Conway, J.H., 1976. *On Numbers and Games*. 1st Edn., Academic Press, New York, ISBN-10: 0121863506, pp: 238.
- Erdős, P., 1961. Graph theory and probability. *Canad. J. Math.*, 13: 346-352. DOI: 10.4153/CJM-1961-029-9
- Feller, W., 1957. *An Introduction to Probability Theory and its Applications*. 1st Edn., John Wiley and Son, New York.
- Feistel, R. and W. Ebeling, 1989. *Evolution of complex systems*. Kluwer Academic Publ., Dordrecht.
- Ferguson, T.S., 1969. *Mathematical Statistics: A Decision Theoretic Approach*. 1st Edn., Academic Press, New York.
- Iverson, G.R., W.H. Longcour, F. Mosteller, J.P. Gilbert and C. Youtz, 1971. Bias and runs in dice throwing and recording: A few million throws. *Psychometrika*, 36: 1-19. DOI: 10.1007/BF02291418
- Jimbo, H.C., J. Azimi and T. Suzuki, 2017. New prospective on multiple dice game and its statistical implications. *Am. J. Applied Math. Stat.*, 5: 169-174. DOI: 10.12691/ajams-5-5-3
- Jimbo, H.C., 2004. Distribution characterization in a practical moment problem. *Acta Math.*, 73: 107-114.
- Kac, K.J. and L. Logan, 1976. In *fluctuating phenomena*. Montroll, E.S.W. and J.L. Lebowitz (Eds.), North Holland, Amsterdam.
- Knizer, R., 1999. *Dice Games Properly Explained*. 1st Edn., Right Wy, Tadworth, ISBN-10: 0716021129, pp: 224.
- Kolmogorov, A.N., 1950. *Foundation of Probability Theory*. 2nd Edn., Chelsea Publishing Company, New York, pp: 71.
- Krasovskij, N.N., 1970. *Game theoretic problem of capture*. Nauka, Moscow.
- Kreps, D.M., 1990. *Game Theory and Economic Modelling*. 1st Edn., Oxford University Press, ISBN-10: 0198283814, pp: 195.
- Maitra, A.P. and W.D. Sudderth, 1996. *Discrete Gambling and Stochastic Games*. 1st Edn., Springer, ISBN-10: 0387946284, pp: 244.
- Mayerson, R.B., 1991. *Game Theory: Analysis of Conflict*. 1st Edn., Harvard University Press, Cambridge, ISBN-10: 0674341155, pp: 568.
- Morris, P., 1994. *Introduction to Game Theory*. 1st Edn., Springer, New York, ISBN-10: 038794284X, pp: 252.
- Murray, H.J.R., 2002. *A History of Chess*. 1st Edn., Oxford University Press, London.
- Nelson, E., 1985. *Quantum Fluctuations*. 1st Edn., Princeton University Press, Prince Town, ISBN-10: 0691083797, pp: 146.
- Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1st Edn., Morgan Kaufmann, San Monteo, CA., USA., ISBN-10: 1558604790, pp: 552.
- Roth, A.E. and M.A.O. Sotomayor, 1990. *Two-Sided Matching: A Study in Game-theoretic Modeling and Analysis*. 1st Edn., Cambridge University Press, New York, ISBN-10: 052139015X, pp: 265.