

Investigation of Theoretical Models of the Behavior of Animal Populations

Or Zuk

Investigation of Theoretical Models of the Behavior of Animal Populations

Research Thesis

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

Or Zuk

Submitted to the Senate of
the Technion - Israel Institute of Technology

Iyyar, 5761

Haifa

May, 2001

The research thesis was done under the supervision of Prof. Shay Gueron (Department of Mathematics of the University of Haifa) in the interdepartmental program of Applied Mathematics.

My special thanks are given to Prof. Shay Gueron for his supervision, cooperating work during six years, help, encouragement, support, and optimism.

The generous financial help of the Technion Graduate School is gratefully acknowledged.

Contents

Abstract	1
List of symbols and abbreviations	2
1 Introduction and literature survey	3
1.1 Literature Survey	4
1.1.1 Background on Lions	4
1.1.2 Background on birds	5
1.1.3 The secretary problem	5
2 On Smoluchowski Equations for Coagulation Processes with Multiple Absorbing States	6
2.1 introduction	7
2.2 The Coagulation Process	9
2.2.1 Preliminaries and definitions	9
2.2.2 Coagulation processes with multiple absorbing states .	10
2.2.3 Direct computation of the group size distribution . . .	10
2.2.4 The Smoluchowski equations	12
2.3 Example 1: CP with a small population and one absorbing state	13
2.4 Example 2: D-truncated CP ($D \approx N$) CP with a small population and multiple absorbing states	15
2.5 Example 3: D-truncated CP ($D \ll N$) with a large population and multiple absorbing states and	17

2.6	Concluding remarks	18
3	Mixing through coagulation processes	22
3.1	introduction	24
3.2	Defining Coagulation processes:	26
3.2.1	Computing the group size probability directly	28
3.2.2	Computing the group size probability via Monte Carlo simulations	30
3.2.3	The Smoluchowski type equations	31
3.3	DCP as a grouping strategy	33
3.3.1	Kernel Quality	33
3.3.2	Coagulation kernels for maximizing quality	35
3.3.3	Efficiency	38
3.3.4	Example 3	39
3.4	Conclusion	41
4	Comparing strategies in a multi-staged decision game	43
4.1	Introduction	45
4.2	Definition of the process	47
4.3	Strategies for solution	48
4.3.1	The Stupid Bird	48
4.3.2	The Smart Bird	49
4.3.3	Example : The Uniform Distribution	53
4.3.4	Example : The Exponential Distribution	55
4.3.5	Example : The Normal Distribution	56
4.4	Generalizing the basic model	60
4.4.1	Bounded G	60
4.4.2	Time dependent utility	61
4.4.3	Searching Cost	62
	Appendix: Description of the numerical procedure	66

List of Figures

2.1	The numerical trajectory (solid line) $f_2(t)$ (panel a), and $f_4(t)$ (panel b), of the Smoluchowski ODE system (2.11), compared with the exact expectation (dashed line). The horizontal axis is time.	20
2.2	The numerical trajectory (solid line) $f_2(t)$ (panel a), and $f_4(t)$ (panel b), of the Smoluchowski ODE system (2.15), compared with the exact expectation (dashed line). The horizontal axis is time.	20
3.1	The Resulting Group Size Probabilities For Different coagulation kernels : constant, additive and multiplicative.	36
3.2	The Quality (solid line) And Efficiency(zig-zag line) of the coagulation kernel. The horizontal axis is β	40

Abstract

In this work we consider some problems of modeling the behavior of animal populations. In the first part, we introduce and discuss a coagulation process, in which a finite population of particles coagulate to form larger groups. We discuss the quality of approximation by Smoluchowski equations in this context. In the second part, we interpret the coagulation process as a model for the social behavior of large mammals (e.g. lions), and discuss the question of when is this model indeed adequate to the population 'demand' of being in some desired group size. In the third part, we discuss a completely different topic. We consider a multi staged decision process under uncertainty conditions. Our conclusion is that developing an optimal strategy is not of great importance, and satisfactory results might be accomplished by much simpler strategies requiring significantly less memory resources.

List of symbols and abbreviations

N	– population size
D	– maximal group size
t	– time
Ω_N	– state space
$\Omega_{N,D}$	– state space of truncated process
$A_{N,D}$	– set of absorbing states
η_0	– solitary state
$n_i(\eta)$	– number of groups of size i in η
$\#\eta$	– lexicographic order of η
$\psi(i, j)$	– coagulation rate
$\Psi(i, j; \eta)$	– total merging intensity
$\Pi(t)$	– state space probability distribution
Π	– state space probability distribution at infinity
$f(i; N)$	– group size distribution
B	– transition rate matrix
C	– transition probability matrix
F	– quality probability distribution
μ	– quality expectation
M_F	– maximal value of F
T	– number of search steps
S^T	– a strategy for an T steps process
X_T	– expected score of the T steps process
x_i	– expected score with i steps remaining
$A(\alpha)$	– conditional expectation constant
$B(\alpha)$	– probability constant
C	– cost per turn

Chapter 1

Introduction and literature survey

1.1 Literature Survey

1.1.1 Background on Lions

Lions (*Panthera leo*) are often viewed as an example of a specie possessing a high level of group cooperation. Unlike all other big cats, they live in groups where all members of the group are helping each other even at the expense of their own well being. While cooperative behavior may offer some significant benefits to group member, living in a group have many costs for the individual as well (such as inner-group competition on food, higher diseases carriage etc.) However, since animals are rarely truly altruistic, when combining the payments and benefits of living in a group we expect the individuals to belong to groups of sizes which give the best overall influence. Thus, the common situation of relatively small groups sizes occurring in nature is interpreted as the result of a balanced equilibrium , between forces tending to live alone and forces tending to belong to larger groups.

Data about the serengeti lions behavior, including their group sizes can be found in ???. This data has been analyzed by many researchers, trying to propose models explaining the data collected.

In ???, it is stated that the optimal group size of a group of hunting lions is correlated with the pray's size. Moreover, the optimal size for hunting as concerning kg. food/ lion is smaller that the typical group size in nature. The authors suggest several explanations, such as maximizing the reproductive success or to protecting kills from scavengers. Similar results and conclusions are found also in ???.

In ???, it is shown that hyeanas presence may influence optimal group size among lions, due to their attempts to 'steal' pray from lions. Larger group size yields greater success probability when defending the kills against hyeanas.

In contrary to the great effort which was dedicated to the question of determining the optimal group size for lions, there were not much consideration for modeling an individual's behavior which leads to grouping of optimal or

close to optimal sizes. This article tries to fill this gap by proposing a model for lions grouping, provided that the optimal or desired group size is already given.

1.1.2 Background on birds

1.1.3 The secretary problem

The problem known as the secretary problem has a known and reach history of research. A good reference could be found in ??

Chapter 2

On Smoluchowski Equations for Coagulation Processes with Multiple Absorbing States

Abstract

Smoluchowski coagulation equations propose a model for the stochastic time evolution of a particles population in which particle clusters merge to form larger clusters, at some given rates. These equations represent the dynamics of the expected cluster size distribution. Since Smoluchowski equations were not derived as a rigorous description of the underlying stochastic process, their quality in this context is not obvious.

Here, we consider the case of a finite particles population and raise the following question: to what limit do the solutions of Smoluchowski equations converge as $t \rightarrow \infty$? In particular, we are concerned with the case where the population size is N and the coagulation rates restrict the maximal group sizes to D . For $D = N$, the stochastic process has only one absorbing state, but if $D < N$ it may have many absorbing states. We demonstrate here that when the $D \approx N$, the solutions of Smoluchowski equations do not converge, as $t \rightarrow \infty$, to the expected cluster size distribution, but when $D \ll N$, the convergence is to a limit which is close to the exact solution.

2.1 introduction

A Coagulation process (CP) describes the stochastic time evolution of a population of particles distributed into groups (clusters) which coagulate at rates that depend only on the sizes of the interacting groups. Such processes and their applications have been studied extensively (see [3] for a thorough review). The process has been studied by means of the system of ODE's

$$\dot{c}_j = \frac{1}{2} \sum_{k=1}^{j-1} (R_{j-k,k} c_{j-k} c_k) - \sum_{k=1}^{\infty} (R_{j,k} c_j c_k) \quad j = 1, 2, \dots \quad (2.1)$$

In (2.1), $c_j = c_j(t)$ denote the expected number of j -particle, and $R_{j,k}$ represents the coagulation rate of clusters of size j and k into clusters of size

$j + k$. These equations are named after Smoluchowski who formulated them in 1916 and 1917. They involve various mathematical problems which have been studied in many papers (see [3] for a review). When the coagulation rates are positive (i.e., R_{jk} for $i, j > 0$) and the particles population is assumed to be infinite, infinitely large clusters may appear at finite time (this phenomenon is called gelation). Thus, although the system (2.1) is formally a conservation law, the total mass is not always conserved. Further, under such conditions, the global existence, uniqueness and positiveness, of the solutions is not guaranteed.

The derivation of Smoluchowski equations is heuristic. Gueron [2] showed how equations similar to 2.1 can be viewed as a deterministic approximation for $c_j(t)$ if one ignores the effects of correlations. The consequences of neglecting these correlations may be significant, and therefore the value of Smoluchowski equations as an approximation for the expectation of the stochastic process is not clear. To emphasize the difficulty of ignoring correlations, we mention the case of coagulation-fragmentation processes where the model integral equation (the continuous analog of Smoluchowski equations) was shown to deviate from the exact solution [1].

In this paper we are concerned with the relation between the stochastic CP's and Smoluchowski equations for a finite particles population. For positive coagulation rates, which are those studied in the literature, the stochastic process has only one absorbing state: it always terminates with a single cluster containing the whole population. However, when the coagulation rates bound the largest cluster that can be generated, the CP may have many absorbing states. In this context, our questions are the following: to what limit do the solutions of Smoluchowski equations converge as $t \rightarrow \infty$? Does this limit approximate the expected steady state cluster size distribution? How are the multiple absorbing states being accounted for?

2.2 The Coagulation Process

2.2.1 Preliminaries and definitions

Consider a population of N particles found at time $t = 0$, at the “solitary state” where each particle forms a group of size 1. The population undergoes stochastic evolution of coagulation events where any two groups can merge into a larger one. This CP is a time-homogeneous interacting particle system whose state space is (included in) the set $\Omega = \Omega_N = \{\eta\}$ of all partitions of N . We denote a partition of N into n_i summands of size i , $i = 1, 2, \dots, N$, by $\eta = (n_1, n_2, \dots, n_N)$, where $n_i \geq 0$ and $\sum_{i=1}^N in_i = N$. In these notations, the initial state at $t = 0$ is $\eta_0 = (N, 0, 0, \dots, 0)$. Assuming mass action kinetics, we model the total merging intensity, $\Psi(i, j; \eta)$, at the state $\eta = (n_1, n_2, \dots, n_N) \in \Omega_N$ by

$$\begin{aligned}\Psi(i, j; \eta) &= \Psi(i, j; n_i, n_j) = \psi(i, j) (n_i n_j), \quad i \neq j, \quad 2 \leq i + j \leq N, \\ \Psi(i, i; \eta) &= \Psi(i, i; n_i, n_i) = \psi(i, i) \frac{n_i(n_i - 1)}{2}, \quad 2 \leq 2i \leq N,\end{aligned}\tag{2.2}$$

In 2.2, the rates of the infinitesimal (in time) transitions depend only on the sizes of the interacting groups. For i and j such that $1 \leq i, j \leq N - 1$ and $2 \leq i + j \leq N$, the merging rate of two groups of sizes i and j into one group of size $i + j$ is denoted by $\psi(i, j)$, where $\psi(i, j)$ is some function satisfying $\psi(i, j) = \psi(j, i) \geq 0$. The CP is a non-ergodic Markov chain on Ω_N : from the initial state η_0 , it eventually reaches an absorbing state.

We denote the set of all absorbing states of the process by A . For each state η , we denote the probability of being in η at time t by $\pi_\eta(t)$, and the probability of being in η when $t \rightarrow \infty$ by π_η . The CP is non-ergodic Markov chain, and therefore the existence of such a limit is guaranteed. Clearly $\pi_\eta > 0$ if and only if $\eta \in A$.

Our study concerns the resulting expected groups size distribution $f = f(i; N)$. Here, $f(i; N)$ denotes the expected number of groups of size i , when $t \rightarrow \infty$. The expected groups size distribution is given by

$$f(i; N) = \sum_{\eta \in \Omega_N} \pi_\eta n_i(\eta) = \sum_{\eta \in A} \pi_\eta n_i(\eta) \quad i = 1, 2, \dots, N \quad (2.3)$$

The time dependent groups size distribution is defined analogously. Note that mass conservation implies

$$\sum_{i=1}^N i f(i, N) = N \quad (2.4)$$

To identify the states in Ω_N we sort them lexicographically and use $\#\eta$ to denote the ordinal number of the state η .

2.2.2 Coagulation processes with multiple absorbing states

Suppose the coagulation kernel is strictly positive, that is, $\psi(i, j) > 0$ for $i, j > 0$ such that $i + j \leq N$. Then, independently of the initial condition, the process terminates only when all the particles are found in one cluster of size N . In other words, $\eta = (0, 0, \dots, 0, 1)$ is the only absorbing state.

D-truncated CP's

We define here another class of CP's, which we call D-truncated CP's, where $\psi(i, j) > 0 \iff i + j \leq D$ for some $D \leq N$. In D-truncated CP's (starting from η_0) the size of a group never exceeds D . Accordingly, the process may have many absorbing states as shown the following examples.

2.2.3 Direct computation of the group size distribution

The group size distribution $f(i; N)$ is a functional of the probability distribution π . One way to compute $f(i; N)$ is to compute π first, and to obtain $f(i; N)$ from (2.3).

To compute the probability distribution π we construct the transition rate matrix B , whose entries are determined as follows. A transition accounting for the coagulation of groups of size $i \neq j$ moves the population from a state $\eta = (n_1, n_2, \dots, n_N)$ to the state $\xi = (n_1, n_2, \dots, n_i - 1, \dots, n_j - 1, \dots, n_{i+j} + 1, \dots, n_N)$. In this case, $B(\#\eta, \#\xi) = \psi(i, j)n_i(\eta)n_j(\eta)$. A transition representing the coagulation of two groups of size i moves the population from a state $\eta = (n_1, n_2, \dots, n_N)$ to the state $\xi = (n_1, n_2, \dots, n_i - 2, \dots, n_{2i} + 1, \dots, n_N)$. In this case, $B(\#\eta, \#\xi) = \frac{1}{2}\psi(i, i)n_i(\eta)(n_i(\eta) - 1)$. For all other pairs η, ξ such that $\eta \neq \xi$, the transition from η to ξ cannot occur, and therefore $B(\#\eta, \#\xi) = 0$. Since the rate of leaving the state η is the sum of all the rates of transitions emanating from η , the diagonal entries of B are determined by

$$B(\#\eta, \#\eta) = - \sum_{\eta \neq \xi \in \Omega_N} B(\#\eta, \#\xi) \quad (2.5)$$

With these definitions, the probability distribution $\pi(t)$ is

$$\pi(t) = \eta_0 \times e^{Bt} \quad (2.6)$$

To compute π (i.e., the limit as $t \rightarrow \infty$) we generate a new matrix, C , in a way that for any $\eta \neq \xi$, $C(\#\eta, \#\xi)$ is the transition probability from η to ξ . This is done in the following way:

1. If $B(\#\eta, \#\eta) \neq 0$, then row number $\#\eta$ in C is obtained by dividing row number $\#\eta$ in B by $-B(\#\eta, \#\eta)$.
2. If $B(\#\eta, \#\eta) = 0$, η is an absorbing state, and the entire row of B is 0. In this case we set $C(\#\eta, \#\eta) = -1$ and the other entries of this row remain 0.

For each absorbing state $\eta \in A$ we define 1_η as a vector of length $|\Omega_N|$, whose entries are 0 except for entry number $\#\eta$ which is set to be -1 . We now consider the linear system

$$S(\eta) : \quad Cx = 1_\eta \quad (2.7)$$

and note that π_η is the value of $x_{\#\eta_0}$ in the solution of $S(\eta)$. Thus, in order to find π , we have to solve $Cx = 1_\eta$, for every $\eta \in A$ (this requires one Gauss elimination with multiple free columns). The solution when the initial condition is not η_0 can be obtained from this system as well.

2.2.4 The Smoluchowski equations

Smoluchowski equations are an attempt to approximate the group size distribution by forming a self contained system of ODE's with the unknowns $f_i = f_i(t)$, where f_i "represents" $f(i; N)$. Following [2], the exact evolution equations for $f_i(t)$ read

$$\frac{df_i}{dt} = \sum_{k=1, k \neq \frac{i}{2}}^{i-1} f_k f_{i-k} \psi(k, i-k) + \frac{1}{2} f_{\frac{i}{2}} (f_{\frac{i}{2}} - 1) \psi\left(\frac{i}{2}, \frac{i}{2}\right)_{(i \text{ even})} - \sum_{k=1, k \neq i}^{D-1} f_i f_k \psi(i, k) - f_i (f_i - 1) \psi(i, i)_{(2i \leq D)} + CORR_i, \quad (i = 1, 2, \dots, D) \quad (2.8)$$

with initial conditions $f_1 = N$ and $f_i = 0$ otherwise. Here, $CORR_i$ are the correlation terms, and so far, f_i and $f(i; N)$ are the same (see [2]). To obtain a self contained system of equations (Smoluchowski type equations) for the unknowns f_i , we ignore the $CORR_i$ terms that depend explicitly on the distribution π_i .

For computing the equilibrium, we take $\frac{df_i}{dt} = 0$, $i = 1, 2, \dots, N$, in (2.8) and obtain a system of quadratic equations. Any steady state of the process is a solution of this system.

The question we address here concerns the comparison between limit, when $t \rightarrow \infty$, of f_i , the solution of the Smoluchowski system, and $f(i; N)$, the expected groups size distribution (which we can compute directly for small populations).

2.3 Example 1: CP with a small population and one absorbing state

In our first example we take $N = 5$, and $\psi(i, j) \equiv 1$. The state space Ω_5 has 7 states, listed in lexicographic order

$$\Omega_5 = \{(0, 0, 0, 0, 1), (0, 1, 1, 0, 0), (1, 0, 0, 1, 0), (1, 2, 0, 0, 0), \\ (2, 0, 1, 0, 0), (3, 1, 0, 0, 0), (5, 0, 0, 0, 0)\}$$

There is only one absorbing state, namely $\eta = (0, 0, 0, 0, 1)$. The transition rates matrix B is

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & -3 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 3 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 & -10 \end{pmatrix} \quad (2.9)$$

and the “normalized” matrix C is

$$C = \begin{pmatrix} -10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & -1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (2.10)$$

For finite time, the group size distribution is calculated by using (2.9) and (2.6). The steady state group size distribution (i.e., at $t \rightarrow \infty$), computed by (2.7), is $f_{exact} = (0, 0, 0, 0, 1)$ (it corresponds with having only one absorbing state).

The Smoluchowski equations for this case read

$$\begin{aligned}
\dot{f}_1(t) &= -f_1(t)(f_1(t) - 1) - f_1(t)f_2(t) - f_1(t)f_3(t) - f_1(t)f_4(t) \\
\dot{f}_2(t) &= -f_1(t)f_2(t) - f_2(t)(f_2(t) - 1) + \frac{1}{2}f_1(t)(f_1(t) - 1) - f_2(t)f_3(t) \\
\dot{f}_3(t) &= -f_1(t)f_3(t) + f_1(t)f_2(t) - f_2(t)f_3(t) \\
\dot{f}_4(t) &= \frac{1}{2}f_2(t)(f_2(t) - 1) + f_3(t)f_1(t) - f_1(t)f_4(t) \\
\dot{f}_5(t) &= f_1(t)f_4(t) + f_2(t)f_3(t)
\end{aligned} \tag{2.11}$$

with the initial conditions $f_1(0) = 5$, $f_2(0) = f_3(0) = f_4(0) = f_5(0) = 0$.

The equations of (2.11) are, by definition, functionally dependent: adding up i times equation number i , for $i = 1, 2, \dots, 5$, gives an identity $\dot{f}_1(t) + 2\dot{f}_2(t) + 3\dot{f}_3(t) + 4\dot{f}_4(t) + 5\dot{f}_5(t) = 0$, which reflects the fact that the system (2.11) is a conservation law.

To compare the solution of Smoluchowski system (2.11) with the exact expected group size distribution, we solved (2.11) numerically. 2.1 shows the trajectory of $f_2(t)$ (panel a), and $f_4(t)$ (panel b) for $0 \leq t \leq 30$, compared with the exact solution (which was obtained from (2.6) and the definition of the expectation). As one can see, the wiggling solution of (2.11) is not a good approximation to the exact solution. Further, note that the numerical trajectory of (2.11) dips below 0, which renders it completely irrelevant.

Propagating the numerical trajectory of (2.11) to large t , enables us to approximate numerically the limit to which the system converges, and the result is $f_{Smoluchowski} = (1, 0, 0, 0, 0.8)$. Recalling that the exact expected equilibrium group size distribution is $f_{exact} = (0, 0, 0, 0, 1)$, we conclude that $f_{Smoluchowski}$ is not a good approximation.

2.4 Example 2: D-truncated CP ($D \approx N$) CP with a small population and multiple absorbing states

We now study a D-truncated CP. We take $N = 5$, $D = 4$, $\psi(2, 3) = \psi(3, 2) = \psi(1, 4) = \psi(4, 1) = 0$, and $\psi(i, j) = 1$ for all other i, j . The state space is Ω_5 , which was listed above. Unlike the case with Example 1, we have here two absorbing states. These are

$$A = \{\eta = (0, 1, 1, 0, 0), \xi = (1, 0, 0, 1, 0)\}$$

The transition rates matrix B is

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & -3 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 3 & -6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 & -10 \end{pmatrix} \quad (2.12)$$

and the normalized matrix C is

$$C = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & -1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (2.13)$$

To compute the probability of being absorbed in $\eta = (0, 1, 1, 0, 0)$ (the first absorbing state of A), we set $1_\eta = (0, -1, 0, 0, 0, 0, 0)$ and solve the system $S(\eta) : Cx = 1_\eta$. The solution is

$$x = \left(0, 1, 0, \frac{2}{3}, \frac{1}{3}, \frac{1}{2}, \frac{1}{2}\right) \quad (2.14)$$

From (2.14) it follows that $\pi_\eta = x_7 = \frac{1}{2}$. Recalling that there are only two absorbing states, it follows that $\pi_\xi = 1 - \pi_\eta = \frac{1}{2}$, i.e., the probability of being absorbed in either one of the absorbing states equals $\frac{1}{2}$. The resulting steady state group size distribution, computed by (2.3), is $f_{exact} = (0.5, 0.5, 0.5, 0.5, 0)$.

The Smoluchowski equations in this case read

$$\begin{aligned}
 \dot{f}_1(t) &= -f_1(t)(f_1(t) - 1) - f_1(t)f_2(t) - f_1(t)f_3(t) \\
 \dot{f}_2(t) &= -f_1(t)f_2(t) - f_2(t)(f_2(t) - 1) + \frac{1}{2}f_1(t)(f_1(t) - 1) \\
 \dot{f}_3(t) &= -f_1(t)f_3(t) + f_1(t)f_2(t) \\
 \dot{f}_4(t) &= \frac{1}{2}f_2(t)(f_2(t) - 1) + f_3(t)f_1(t) \\
 \dot{f}_5(t) &= 0
 \end{aligned} \tag{2.15}$$

with the initial conditions $f_1(0) = 5$, $f_2(0) = f_3(0) = f_4(0) = f_5(0) = 0$.

To compare the solution of (2.15) with the exact expected group size distribution, the system (2.15) was solved numerically. Fig. 2 shows the numerical trajectory of $f_2(t)$ (panel a), and $f_4(t)$ (panel b) for $0 \leq t \leq 30$, compared with the exact solution. As with the previous example, the wiggling solution of (2.15) is not a good approximation to the exact solution, and the numerical trajectory of (2.15) dips below 0.

By propagating the numerical solution of (2.15) to large t , we find (numerically) that the solution of the Smoluchowski system converges to $f_{Smoluchowski} = (1, 0, 0, 1, 0)$. It fails to approximate the exact solution $f_{exact} = (0.5, 0.5, 0.5, 0.5, 0)$. Moreover, it is easy to check (by substitution) that the exact solution $f_{exact} = (0.5, 0.5, 0.5, 0.5, 0)$ does not even satisfy the Smoluchowski equilibrium equations.

Note that in the given example, the numerical solution of Smoluchowski system converges to one of the two absorbing states $(1, 0, 0, 1, 0)$, whereas the expected equilibrium group size distribution is a nontrivial linear combination of them.

However, in general, the Smoluchowski system does not necessarily converge to an absorbing state. For example, with the initial condition $f_{Smoluchowski} = (3, 1, 0, 0, 0)$, the solution of (2.15) converges to $\approx (0, 1, 0.5787, 0.3159, 0)$. This solution is not an absorbing state and not even a linear combination of the two absorbing states.

2.5 Example 3: D-truncated CP ($D \ll N$) with a large population and multiple absorbing states and

Here we study a D-truncated CP with a large population and $D \ll N$. We take $N = 100$, $D = 3$, $\psi(1, 1) = \psi(1, 2) = \psi(2, 1) = 1$ with $\psi(i, j) = 0$ otherwise. Here, we cannot use Ω_{100} directly because it too large. However, the relevant state space (containing only states with group of size not exceeding 3) has only 884 states (we do not list these states here). Also, there are 18 absorbing states, listed in lexicographic order:

$$A = \{(0, 2, 32), (0, 5, 30), (0, 8, 28), (0, 11, 26), (0, 14, 24), (0, 17, 22), \\ (0, 20, 20), (0, 23, 18), (0, 26, 16), (0, 29, 14), (0, 32, 12), (0, 35, 10), \\ (0, 38, 8), (0, 41, 6), (0, 44, 4), (0, 47, 2), (0, 50, 0), (1, 0, 33)\}$$

To compute π , we solve the 884×884 linear system (2.7) with 18 free columns (one for each absorbing state). The resulting group size distribution is $f_{exact} = (0.0091, 7.9031, 28.0616)$.

The Smoluchowski equations this case reads

$$\begin{aligned} \dot{f}_1(t) &= -f_1(t)(f_1(t) - 1) - f_1(t)f_2(t) \\ \dot{f}_2(t) &= -f_1(t)f_2(t) + \frac{1}{2}f_1(t)(f_1(t) - 1) \\ \dot{f}_3(t) &= f_1(t)f_2(t) \end{aligned} \tag{2.16}$$

with the initial conditions $f_1(0) = 100$, $f_2(0) = f_3(0) = 0$. This system was solved numerically, and the solution was propagated to large t in order

to approximate the equilibrium group size distribution to which the Smoluchowski system converges as $t \rightarrow \infty$. The resulting limit is $f_{Smoluchowski} = (0.0000, 6.4703, 29.0198)$. This distribution is a fairly reasonable approximation to the exact solution (note also that if we substitute the exact solution into the Smoluchowski equilibrium system, we get relatively small deviations from zero.)

Larger populations

Our conjecture is that as the population size N grows, the Smoluchowski equations (at least at equilibrium) provide a better approximation to the exact expected group size distribution. Verifying this conjecture is difficult because computing the exact solution directly is infeasible for large values of N , due to the large state space and the corresponding dimensions of the transition rates matrix. To replace the infeasible direct computation, we used a Monte Carlo method to simulate the actual stochastic process (see [2] for details). Simulation of the D-truncated CP with $N = 900$, $D = 3$, $\psi(1, 1) = \psi(1, 2) = \psi(2, 1) = 1$ with $\psi(i, j) = 0$ otherwise, gave $f_{exact} = (0.0000, 69.3329, 253.7781)$. The Smoluchowski system with these parameters yields $f_{sm} = (0.0000, 67.9305, 254.7130)$, which is a good approximation of the exact solution.

2.6 Concluding remarks

Smoluchowski equations are not a good model for D-truncated CP's when $D \approx N$ (or $D = N$). This is due to the correlation between relatively large interacting groups (see [2] for details).

For small populations, we always have $D \approx N$ (or $D = N$), and we demonstrated that Smoluchowski equations produce a poor approximation to the stochastic CP. With nonpositive coagulation kernels, there are several absorb-

ing states. The Smoluchowski system does not necessarily converge to the exact solution, neither to one of the absorbing states, or to a linear combination of absorbing states, and the solutions are not positive. The same occurs for strictly positive coagulation kernels.

For a large population and $D \ll N$, we gave an example where the Smoluchowski system converges, as $t \rightarrow \infty$, to a good approximation of the correct solution. We conjecture that this is the case for finite populations with $D \ll N$. The classical study of CP's deals with strictly positive coagulation kernels and $N = \infty$, and is modeled by the infinite system of OD's (2.1). Many of the mathematical difficulties and phenomena associated with this model (e.g., violation of mass conservation) do not correspond to analogous phenomena in a finite particles population. Note that the classical study allows for only one absorbing state: one cluster containing the whole population. Therefore, in our terminology, the classical Smoluchowski equations can be viewed as the limit as $N \rightarrow \infty$ of a D-truncated CP where $D = N$. Recalling that for finite populations, the Smoluchowski equations for D-truncated CP's with $D = N$ produce a poor approximation to the exact group size distribution, illustrates an intrinsic difficulty.

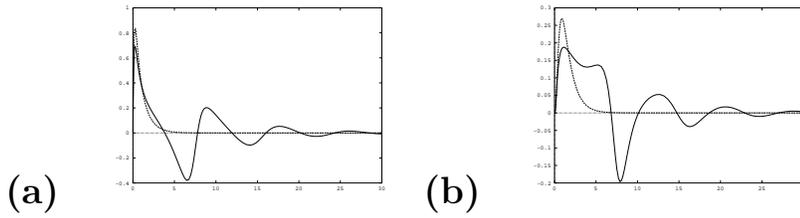


Figure 2.1: The numerical trajectory (solid line) $f_2(t)$ (panel a), and $f_4(t)$ (panel b), of the Smoluchowski ODE system (2.11), compared with the exact expectation (dashed line). The horizontal axis is time.

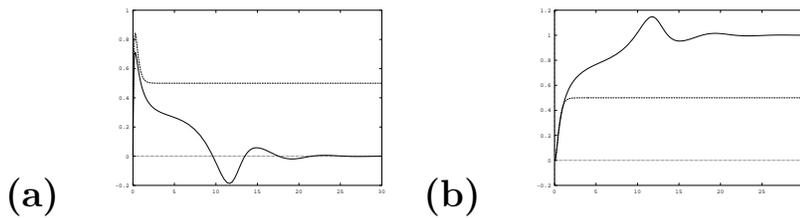


Figure 2.2: The numerical trajectory (solid line) $f_2(t)$ (panel a), and $f_4(t)$ (panel b), of the Smoluchowski ODE system (2.15), compared with the exact expectation (dashed line). The horizontal axis is time.

Bibliography

- [1] R. Durrett, B. L. Granovsky, S. Gueron. Equilibrium behavior of the reversible coagulation-fragmentation processes. *Journal of Theoretical Probability*. 12:447-474 (1999).
- [2] S. Gueron. The Steady-state distributions of coagulation-fragmentation processes. *Journal of Mathematical Biology* 37:1-27 (1998).
- [3] D. Aldous. Deterministic and stochastic models for coalescence (aggregation, coagulation): a review of the mean-field theory for probabilists. *Bernoulli* 5:3-48 (1999).

Chapter 3

Mixing through coagulation processes

Abstract

This paper deals with an application of a stochastic model for coagulation processes. A coagulation process describes a population of N individuals which are divided into groups that coagulate according to some given rates. The fact that this process allows groups only to coagulate, assures that after some finite time the population reaches one of the absorbing states of the state space. We are interested in the expected final group size distribution.

Our study considers a situation where each individual “wishes“ to end up as a member of a group of a desired size D . To this end, we define the special class of bounded coagulation processes, where coagulation rates vanish if the number of individuals in the two interacting groups exceeds D .

In cases where $D \nmid N$, the population cannot be partitioned into groups of size D , and therefore the desire of some individuals cannot be accommodated. In this context, we study the quality of coagulation processes with different coagulation rates as a good mixing and partitioning strategy.

3.1 introduction

Animal grouping is a very common phenomenon in nature. Although it has been studied extensively, the factors that determine the typical group sizes for various animal species are often not completely understood. This question may have ecological and evolutionary significance, since it can be assumed that characteristic group size distributions evolved to maximize some desired fitness.

We consider here a situation where each individual “wishes” to be a member of a (small) group of a desired size that is denoted by D . This might be interpreted for example, as the optimal group size of a pack of lions. A reasonable assumption is that the typical group size evolved as an optimizing balance between the ability to defend the territory, and required food supply. Therefore, at the individual’s level, an individual seeks a strategy that increases its chances of becoming a member in a group of size D . At the population level, at the steady state, we expect to see an approximately perfect partitioning of the population into groups of size D .

Achieving perfect partitioning at the population level seems to require some of cooperation or leadership at the individual level, but this is not always known to be the case. We therefore seek a simple model that enables autonomous partitioning of the population, based only on the local behavior of the individuals, and without requiring knowledge on the population scale. To this end, we study bounded coagulation processes. These are coagulation processes in which coagulation rates vanish if the number of individuals in the two interacting groups exceeds D . Our main question is whether this individual based behavior can lead to perfect or almost perfect distributions of the population into groups. We also note that if $D \nmid N$, where N is the population size, the population cannot be completely partitioned into groups of size D , and the desire of some individuals cannot be accommodated. In such cases we study the quality of coagulation processes with different coagulation rates as

a good mixing and partitioning strategy.

The paper is organized as follows. Section 2 describes the basic coagulation model. In section 3 we present three approaches for computing the stationary distribution. The first two are a direct Markov Chain approach which is accurate but not practical for large populations. The second approach is a Monte Carlo method which provides a good approximation for the distribution, by means of collecting the statistics of simulations. The third approach uses the Smoluchowski type equations, which are based on a heuristic argument [2], [1]. We show here that the quality of the resulting approximation varies, and in some cases can be poor.

In Section 4, we deal with bounded coagulation processes, where we calculate number of individuals that actually find themselves in a group of size D . We prove that bounded coagulation processes with an appropriate choice of (local decision) kernels can lead to a stationary almost perfect distribution, and conclude that such processes can become a grouping strategy. The proof of existence of such kernels involves an upperbound which is sufficient but not necessarily tight. We show by an example for small populations, that in practice, much smaller bounds may be used in order to achieve satisfactory results.

3.2 Defining Coagulation processes:

The general coagulation process

Consider a population of N particles that is found, at time $t = 0$, at the initial state where each particle forms a group of size 1. (This is called also “solitary state”) The population undergoes stochastic evolution (in time) of coagulation events where two groups can merge into a larger one. This stochastic coagulation process (CP) is a time-homogeneous interacting particle system whose state space is (included in) the set $\Omega = \Omega_N = \{\eta\}$ of all partitions of N . For a given N , we denote a partition of N into n_i summands of size i , $i = 1, 2, \dots, N$, by $\eta = [n_1, n_2, \dots, n_N]$. Here, $n_i \geq 0$ and $\sum_{i=1}^N i n_i = N$. The initial state at $t = 0$ is, by our definition, denoted $\eta_0 = [N, 0, 0, \dots, 0]$. Assuming mass action kinetics, we model the total merging intensity, $\Psi(i, j; \eta)$, at the state $\eta = [n_1, \dots, n_N] \in \Omega_N$ by

$$\begin{aligned} \Psi(i, j; \eta) &= \Psi(i, j; n_i, n_j) = \psi(i, j) (n_i n_j), \quad i \neq j, \quad 2 \leq i + j \leq N, \\ \Psi(i, i; \eta) &= \Psi(i, i; n_i, n_i) = \psi(i, i) \frac{n_i(n_i - 1)}{2}, \quad 2 \leq 2i \leq N, \end{aligned} \tag{3.1}$$

In 3.1, the rates of the infinitesimal (in time) transitions depend only on the sizes of the interacting groups. For i and j such that $1 \leq i, j \leq N - 1$ and $2 \leq i + j \leq N$, the merging rate of two groups of sizes i and j into one group of size $i + j$ is denoted by $\psi(i, j)$, for some function satisfying $\psi(i, j) = \psi(j, i) \geq 0$. The CP is a (non ergodic) Markov chain on Ω_N that starts from η_0 and reaches eventually an absorbing state.

The D -truncated coagulation processes

In this paper, we consider only D -truncated coagulation processes (DCP) which are the special case of a CP, where the rates $\psi(i, j)$ satisfy

$$\psi(i, j) = \begin{cases} a(i, j), & \text{if } i + j \leq D \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

for some $D > 0$, and some symmetric function $a(i, j) > 0$. Obviously, since we start from the solitary state, groups of size larger than D are never generated in a DCP. Therefore, we use hereafter an abbreviated notation for the states of a DCP, namely $\eta = (n_1, n_2, \dots, n_D)$, where n_i , $i = 1, 2, \dots, D$ count the number of i -groups in the state. For example, the initial solitary state is represented as a vector of length D , namely $\eta_0 = (N, 0, 0, \dots, 0)$. We also denote the state space of a DCP by $\Omega_{N,D}$. $\Omega_{N,D}$ is the set of partitions of N whose summands are bounded by D . Clearly, $\Omega_{N,D} \subseteq \Omega_N$. The set of absorbing states is denoted by $A_{N,D}$ where $A_{N,D} \subseteq \Omega_{N,D}$. A state $\eta = (n_1, n_2, \dots, n_D) \in \Omega_{N,D}$ is an absorbing state if and only if it does not contain two groups of sizes i, j with $i + j \leq D$, that is, $n_i \times n_j = 0$ for all i, j such that $i + j \leq D$.

The group size probability

For each state η , we denote the probability of being in η at time t by $\pi_\eta(t)$, and the probability of being in η when $t \rightarrow \infty$ by π_η . Clearly $\pi_\eta > 0$ if and only if $\eta \in A_{N,D}$.

We define the group size probability $f(i; N)$ as the probability of ending up in a group of size i as $t \rightarrow \infty$. If the DCP is absorbed in the state $\eta = (n_1, n_2, \dots, n_D)$, the probability of being in a group of size i is $\frac{in_i}{N}$. Therefore $f(i; N)$ is defined by

$$f(i; N) = \frac{i}{N} \sum_{\eta \in \Omega_N} \pi_\eta n_i(\eta) = \frac{i}{N} \sum_{\eta \in A_{N,D}} \pi_\eta n_i(\eta) \quad i = 1, 2, \dots, D \quad (3.3)$$

Relation 3.3, together with conservation of mass, imply that

$$\sum_{i=1}^D f(i, N) = 1 \quad (3.4)$$

Note that the last two relations still hold if we replace f and π by $f(t)$ and $\pi(t)$ respectively.

3.2.1 Computing the group size probability directly

The group size probability $f(i; N)$ is a functional of the probability distribution π . Thus, one way to compute $f(i; N)$ is to compute π first, and to obtain the functional $f(i; N)$ by using Eq. 3.3.

To compute the probability distribution π we construct the transition rate matrix B , whose entries are determined as follows. A transition generated by the coagulation of groups of size $i \neq j$ moves the population from a state $\eta = (n_1, n_2, \dots, n_D)$ to the state $\xi = (n_1, n_2, \dots, n_i - 1, \dots, n_j - 1, \dots, n_{i+j} + 1, \dots, n_D)$. In this case, $B(\#\eta, \#\xi) = \psi(i, j)n_i(\eta)n_j(\eta)$, where we use $\#\eta$ to denote the ordinal number of the state η when the states in $\Omega_{N,D}$ are sorted lexicographically. A transition generated by the coagulation of two groups of size i moves the population from a state $\eta = (n_1, n_2, \dots, n_D)$ to the state $\xi = (n_1, n_2, \dots, n_i - 2, \dots, n_{2i} + 1, \dots, n_D)$. In this case, $B(\#\eta, \#\xi) = \frac{1}{2}\psi(i, i)n_i(\eta)(n_i(\eta) - 1)$. For all other pairs η, ξ such that $\eta \neq \xi$, the transition from η to ξ is not allowed, and we therefore set $B(\#\eta, \#\xi) = 0$. Since the rate of leaving the state η is the sum of all the rates of transitions emanating from η , the diagonal of B is determined by

$$B(\#\eta, \#\eta) = - \sum_{\eta \neq \xi \in \Omega_{N,D}} B(\#\eta, \#\xi) \quad (3.5)$$

With these definitions, the probability distribution $\pi(t)$ is

$$\pi(t) = \eta_0 \times e^{Bt} \quad (3.6)$$

To compute π we generate a new matrix, C , in a way that for any $\eta \neq \xi$, $C(\#\eta, \#\xi)$ is the transition probability from η to ξ . This is done in the following way:

1. If $B(\#\eta, \#\eta) \neq 0$, then row number $\#\eta$ in C is obtained by dividing row number $\#\eta$ in B by $B(\#\eta, \#\eta)$.
2. If $B(\#\eta, \#\eta) = 0$, η is an absorbing state, and the entire row of B is 0. In this case we set $C(\#\eta, \#\eta) = -1$ and the other entries of this row to be 0.

For each absorbing state $\eta \in A_{N,D}$ we define 1_η as a vector of length $|\Omega_{N,D}|$, whose entries are 0 except for entry number $\#\eta$ which is set to be -1 . We now consider the linear system

$$S(\eta) : \quad Cx = 1_\eta \tag{3.7}$$

and note that π_η is the value of $x_{\#\eta_0}$ in the solution of $S(\eta)$.

To illustrate, we provide the following example.

Example 1

We take $N = 5$, $D = 4$, and $a(i, j) \equiv 1$. The state space $\Omega_{5,4}$ and the two absorbing states of $A_{5,4}$, listed in lexicographic order, are

$$\Omega_{5,4} = \{(0, 1, 1, 0), (1, 0, 0, 1), (1, 2, 0, 0), (2, 0, 1, 0), (3, 1, 0, 0), (5, 0, 0, 0)\}$$

$$A_{5,4} = \{\eta = (0, 1, 1, 0), \xi = (1, 0, 0, 1)\}$$

The transition rates matrix B is

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1-3 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0-3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 3-6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10-10 & 0 \end{pmatrix} \tag{3.8}$$

The “normalized” matrix C is

$$C = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & -1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & -1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (3.9)$$

We now compute the probability of being absorbed in $\eta = (0, 1, 1, 0)$, i.e., the first absorbing state in $A_{5,4}$. We set $1_\eta = (-1, 0, 0, 0, 0, 0)$ and solve the system $S(\eta) : Cx = 1_\eta$. The solution is

$$x = \left(1, 0, \frac{2}{3}, \frac{1}{3}, \frac{1}{2}, \frac{1}{2}\right) \quad (3.10)$$

From 3.10 it follows that $\pi_\eta = x_6 = \frac{1}{2}$, and thus $\pi_\xi = 1 - \pi_\eta = \frac{1}{2}$. Therefore, the absorption probability in either of the absorbing states is equal.

The resulting group size probability, computed by 3.3, is $f = (0.1, 0.2, 0.3, 0.4)$.

3.2.2 Computing the group size probability via Monte Carlo simulations

The relevant state space $\Omega_{N,D}$, and therefore the matrix B , may be too large for feasible implementation of the direct approach. Clearly, this occurs when N is large, even if D remains relatively small. In such cases we can estimate $f(i; N)$ by means of Monte Carlo simulations where we do not need to store or manipulate the matrix B . We briefly describe this procedure here.

We start the simulation from the initial state η_0 . Suppose that at some stage we are at the state η . This state is connected, by a single transition, to at most $\frac{D(D-1)}{2}$ other states, and we can compute the related transition probabilities to these states. We simulate a legal transition by choosing a target state at random, while accounting for the relative transition probabilities to the connected states. We continue the procedure until hitting an absorbing state,

and then restart the simulation from η_0 . The statistics of a large number of such simulations gives us an estimate for f .

We note here that this simulation can also be used for estimating the expected time to absorption, in the following manner. Suppose that we are currently in the non-absorbing state η . The expected time to leaving η is the inverse of the rate of leaving η , that is $\frac{-1}{B(\#\eta, \#\eta)}$. Thus, if we sum $\frac{-1}{B(\#\eta, \#\eta)}$ over all the visited states from the initial state until the simulation is stopped at an absorbing state, we obtain the expected time to absorption of the specific route chosen by the simulated realization of the CP. The statistics of a large number of such simulations gives an estimate for the expected time until absorption.

3.2.3 The Smoluchowski type equations

Another method for computing the group size probability is implemented by a system of ODE's. As we illustrate here it gives unsatisfactory results. More details can be found in [2]. Following [2], the Smoluchowski type equations for the group size probability read

$$\begin{aligned} \frac{df_i}{dt} = & Ni \sum_{k=1, k \neq \frac{i}{2}}^{i-1} \frac{f_k f_{i-k}}{k(i-k)} \psi(k, i-k) + f_{\frac{i}{2}} \left(\frac{2N}{i} f_{\frac{i}{2}} - 1 \right) \psi\left(\frac{i}{2}, \frac{i}{2}\right)_{(i \text{ even})} - \\ N \sum_{k=1, k \neq i}^{D-1} \frac{f_i f_k}{k} \psi(i, k) - & f_i \left(\frac{N}{i} f_i - 1 \right) \psi(i, i)_{(2i \leq D)} + CORR_i, \quad (i = 1, 2, \dots, D) \end{aligned} \quad (3.11)$$

where $CORR_i$ are the correlation terms. To obtain a self contained system of equations for the unknowns f_i we ignore the $CORR_i$ terms that depend explicitly on the distribution π_i . The initial conditions are $f_1 = 1$ and $f_i = 0$ otherwise.

Solving 3.11, and computing the limit of the solution as $t \rightarrow \infty$, we may expect to get a reasonable approximation for the stationary group size probability if $D \ll N$. This, however, is not the case if D and N are comparable.

To illustrate, we use the data of the previous example, namely $N = 5$, $D = 4$, and $a(i, j) = 1$. The resulting system of ODE's reads

$$\begin{aligned}
 \dot{f}_1(t) &= -f_1(t)(5f_1(t) - 1) - \frac{5}{2}f_1(t)f_2(t) - \frac{5}{3}f_1(t)f_3(t) \\
 \dot{f}_2(t) &= -5f_1(t)f_2(t) - f_2(t)\left(\frac{5}{2}f_2(t) - 1\right) + f_1(t)(5f_1(t) - 1) \\
 \dot{f}_3(t) &= -5f_1(t)f_3(t) + \frac{15}{2}f_1(t)f_2(t) \\
 \dot{f}_4(t) &= f_2(t)\left(\frac{5}{2}f_2(t) - 1\right) + \frac{20}{3}f_3(t)f_1(t)
 \end{aligned} \tag{3.12}$$

with the initial conditions $f_1(0) = 1$, $f_2(0) = f_3(0) = f_4(0) = 0$.

Numerical solution gives $f = (0.2, 0, 0, 0.8)$, which does not approximate the exact solution, $f = (0.1, 0.2, 0.3, 0.4)$, detailed above.

3.3 DCP as a grouping strategy

3.3.1 Kernel Quality

We assume here that all individuals in the population wish to be found, eventually, in a group of size D . With this intention in mind, we ask if the DCP is a good mixing strategy, that is, what coagulation kernels lead to a “good” group size probability?

From the definition, it is clear that the “ideal” group size probability is $f_{ideal} = (0, 0, 0, \dots, 1)$, corresponding to the absorbing state $\eta_{ideal} = (0, 0, 0, \dots, \frac{N}{D})$. However, note that η_{ideal} is a legal absorbing state if and only if $D \mid N$, and is not accessible if $D \nmid N$. In such cases, the best possible group size probability, denoted by f_{best} , is

$$f_{best} : f(i) = \begin{cases} \frac{N \bmod D}{N}, & \text{if } i = N \bmod D, \\ 1 - \frac{N \bmod D}{N}, & \text{if } i = D, \\ 0 & \text{otherwise.} \end{cases} \quad (3.13)$$

Which corresponds to the absorbing state $\eta_{best} = (0, 0, \dots, 0, 1, 0, \dots, [\frac{N}{D}])$, where the ‘1’ is in the place $N \bmod D$.

As one can see, To assess the individual’s fitness resulting from a given coagulation kernel $a = a(i, j)$ we need to define a quality function which is maximized at f_{best} . We use here linear quality function.

Definition: Let α_i , $i = 1, 2, \dots, D$ be some nonnegative constants. Suppose that a is the coagulation kernel leading to the steady state group size probability f . The quality of the kernel a is defined by

$$S = S(a, \alpha, f) = \sum_{i=1}^D \alpha_i f_i \quad (3.14)$$

Since individuals want to form large groups, up to the desired size D , it is reasonable to choose α_i that are monotonically increasing in i . Such choice, however, does not guarantee that the quality $S(a, \alpha, f)$ is maximized at f_{best} .

To illustrate, we take $N = 5$, $D = 4$, and $a(i, j) = 1$ as in 3.2.1, and choose $\alpha = (0, 9, 10, 11)$. Here, $f_{best} = (0.2, 0, 0, 0.8)$. Setting $f^1 = (0, 0.4, 0.6, 0)$ we can compute $S(a, \alpha, f_{best}) = 8.8 < 9.6 = S(a, \alpha, f^1)$, so f_{best} does not attain maximal quality. Therefore, we further restrict the weight α 's and prove the following assertion.

Claim 1: Suppose that the weights α_i , $i = 1, 2, \dots, D$ are the values $\alpha_i = A(i)$ of a nonnegative convex function A defined on the interval $[0, D]$ with $A(0) = 0$. Then, f_{best} maximizes the quality function defined in 3.14 over all possible group size distributions.

Proof: Let $\eta \in A_{N,D}$ be an absorbing state and let f^η be the corresponding group size distribution (that is $f^\eta(i) = \frac{in_i(\eta)}{N}$). Clearly, the set of all possible group size distributions is the convex hull of the set $H = \{f^\eta | \eta \in A_{N,D}\}$. Due to the given linearity of the quality function, the quality of any $f \in Conv(H)$ is the weighted average of the qualities of some elements in H . Therefore, it is sufficient to show that for every $\eta \in A_{N,D}$, we have $S(a, \alpha, f^\eta) \leq S(a, \alpha, f_{best})$. Let $\eta \in A_{N,D}$ be an absorbing state. If all but maybe one of the groups in η are of size D , then $f_{best} = f^\eta$ and we are done. Therefore, suppose that η contains two groups of sizes $k_1, k_2 < D$. We consider two cases: if $k_1 + k_2 > D$ we replace these two groups with two other groups of sizes $D, k_1 + k_2 - D$. Since A is convex and non-negative it can be shown that the resulting quality cannot decrease. If on the other hand, $k_1 + k_2 \leq D$, we replace the two groups with one group of size $k_1 + k_2$. Now, let us extend the definition of α_k by defining $\alpha_0 = 0$. We can now reduce this case to the first case, by considering the group of size $k_1 + k_2$, and an empty group of size 0. As before, the resulting quality cannot decrease. After a finite number of such steps, we finally reach f_{best} , and therefore f_{best} has the maximal quality. ■

Two possible selections for α_i are:

$$Q_1 : \quad \alpha_D = 1, \alpha_i = 0, \forall i \neq D, \quad i = 1, 2, \dots, D \quad (3.15)$$

With this quality function being in a group of size D is the only case that gets credit, and the quality is the number of such groups. Being in any other group is considered as equally bad. This quality function is weakly convex, but f_{best} is still a strong maximum since there is only one absorbing state with the maximal number of groups of size D .

$$Q_2 : \quad \alpha_i = i^\gamma, \quad i = 1, 2, \dots, \quad D, \gamma \geq 1 \quad (3.16)$$

With this quality function being in a larger group gets a larger credit. We assume that $\gamma \geq 1$, to ensure the convexity of S . When $\gamma = 1$, the quality function is simply the expected group size.

Example 2

We set $N = 20$ and $D = 6$. The initial state is $\eta_0 = (20, 0, 0, 0, 0, 0)$. The set of absorbing states, $A_{20,6}$, listed in lexicographic order is

$$A_{20,6} = \{(0, 0, 0, 0, 4, 0), (0, 0, 0, 1, 2, 1), (0, 0, 0, 2, 0, 2), (0, 0, 0, 5, 0, 0), \\ (0, 0, 1, 0, 1, 2), (0, 0, 1, 3, 1, 0), (0, 1, 0, 0, 0, 3)\}$$

Figure 3.1 shows the results obtained when using the three coagulation kernels $a(i, j) = 1$, $a(i, j) = i + j$, $a(i, j) = ij$.

We see that the coagulation kernels affect significantly the resulting group size probability, and the resulting quality. Higher quality is attained when the coagulation rates of large groups is larger than those of the smaller groups.

3.3.2 Coagulation kernels for maximizing quality

If D is fixed and $N \rightarrow \infty$, we see from the definition of f_{best} that f_D approaches 1, or, equivalently, that $f_{best} \rightarrow f_{ideal}$ in the infinity norm. We are therefore

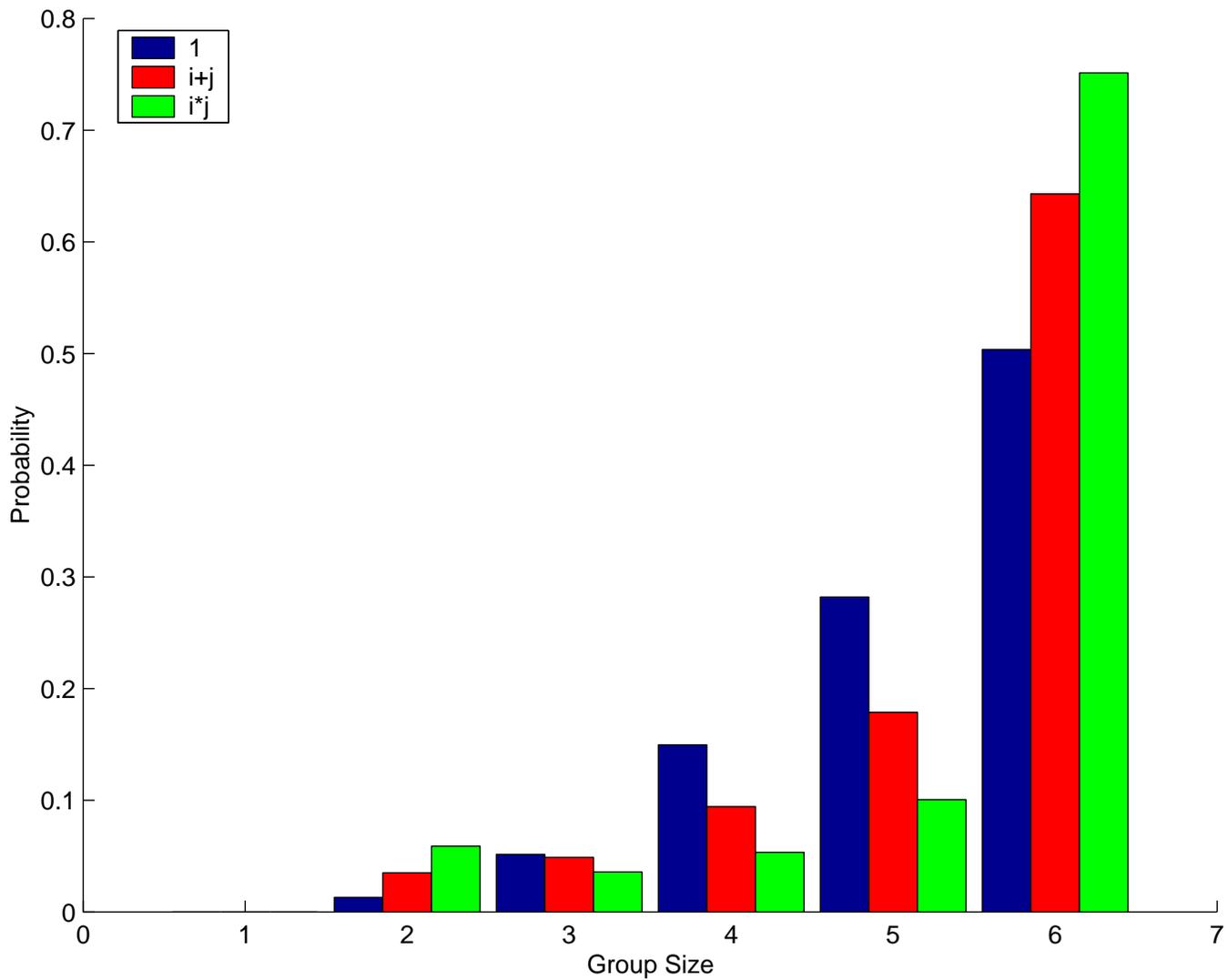


Figure 3.1: The Resulting Group Size Probabilities For Different coagulation kernels : constant, additive and multiplicative.

interested in finding kernels that lead to a stationary group size probability where the value of f_D is arbitrarily close to 1. In the following claim we prove the existence of such kernels.

Claim 2: There exist a series of kernels $\{a_N = a_N(i, j)\}$ such that when using the kernel a_N on the population size N , the resulting stationary group size probability $f(i; N)$ satisfies

$$\lim_{N \rightarrow \infty} f_D = 1 \quad (3.17)$$

Proof : Let N be arbitrary. We take some positive large $U = U(N)$ whose value is determined later, and define the kernel $a_N(i, j)$ by

$$a_N(i, j) = \begin{cases} U^i, & \text{if } j = 1, \\ U^j, & \text{if } i = 1, \\ \frac{1}{U} & \text{otherwise.} \end{cases} \quad (3.18)$$

For $i = 0, 1, \dots, \lfloor \frac{N}{D} \rfloor$, we use η_i to denote the state $\eta_i = (N - Di, 0, 0, \dots, 0, i)$. Let ξ_0 be the state that corresponds to the best group size probability, that is, $\xi_0 = (0, 0, \dots, 0, 1, 0, \dots, 0, \lfloor \frac{N}{D} \rfloor)$. For every two states $\eta, \xi \in \Omega_{N,D}$ let $P(\xi | \eta)$ be the conditional probability that the DCP reaches the state ξ , given that it is now in η . Then,

$$P(\xi_0 | \eta_0) \geq P(\xi_0 | \eta_{\lfloor \frac{N}{D} \rfloor}) \prod_{i=1}^{\lfloor \frac{N}{D} \rfloor} P(\eta_i | \eta_{i-1}) = \prod_{i=1}^{\lfloor \frac{N}{D} \rfloor} P(\eta_i | \eta_{i-1}) \geq (P(\eta_1 | \eta_0))^{\lfloor \frac{N}{D} \rfloor} \quad (3.19)$$

The last inequality holds because $P(\eta_i | \eta_{i-1})$ is monotonically increasing in i . Since $a(i, j) > 0$ it is possible to get from η_0 (the initial state) to η_1 . One possible way is by the following chain of coagulation events: $(1, 1), (1, 2), \dots, (1, D-1)$. Therefore,

$$P(\eta_1 | \eta_0) \geq 1 \times \frac{U}{U + \frac{1}{2}(N-3)} \times \frac{U^2}{U^2 + \frac{1}{2}(N-4)} \times \dots \times \frac{U^{D-2}}{U^{D-2} + \frac{1}{2}(N-D)} \geq$$

$$\left(\frac{U}{U + \frac{1}{2}(N - 3)}\right)^D \quad (3.20)$$

We conclude that

$$P(\xi_0 | \eta_0) \geq \left(\frac{U}{U + \frac{1}{2}(N - 3)}\right)^N \quad (3.21)$$

Clearly, for every $\epsilon = \epsilon(N) > 0$, we can find a value $U = U(N)$ such that $P(\xi_0 | \eta_0) > 1 - \epsilon$. Therefore, we have $f_D \geq (1 - \epsilon)^{\frac{N - N \bmod D}{N}}$. Determining, for example $\epsilon(N) = \frac{1}{N}$, we get

$$\lim_{N \rightarrow \infty} f_D = 1 \quad (3.22)$$

■

3.3.3 Efficiency

We showed that the maximal quality can be approached with an appropriate choice of coagulation kernels. However, it turns out that with these kernels, the expected time to a coagulation event ranges extremely from state to state. In other words, if we normalize the kernel and make the sum of the $a(i, j)$ values 1, the expected time to absorption would be very large. This happens, since to achieve 'good' kernels, we need to have more control on the coagulations of groups, such that groups will be formed in the sizes, we require. This causes a relatively large variation in the rates of 'good' kernels, and therefore a long time until absorption, while kernels which are more homogeneous result shorter time until absorption but achieve poorer quality.

This discussion leads to the following definition:

Efficiency: The efficiency, ν , of the kernel $a = a(i, j)$, normalized such that $\sum_{i \leq j}^D a(i, j) = 1$, is

$$\nu = \frac{1}{E} \quad (3.23)$$

where E is the expected time to absorption.

Note that with no loss of generality, we can assume that the kernel is normalized, because if all the $a(i, j)$ values are multiplied by a constant $C > 0$, we end up with the same CP but with rescaled time.

As was mentioned before, it seems that raising the efficiency of a kernel reduces the quality in the solution. Therefore one may ask if there is a point in trying to find kernels with both good quality and efficiency. The next example shows that although efficiency and quality are indeed anti correlated, raising one of them does not necessarily reduces the other, therefore there are kernels that are 'better' then others with respect to both of this quantities, which gives a motivation for trying to find such 'good' kernels.

3.3.4 Example 3

Consider a DCP with $D = 3$, arbitrary N , and the quality function Q_1 .

We assume that $a(1, 1) = \beta = 1 - a(1, 2) = 1 - a(2, 1)$. Note that the quality is a monotonically decreasing function of β since increasing $a(1, 2)$ increases f_2 . However, the efficiency is not monotonically increasing in β . We computed the efficiency ν , for the case of $N = 100$, using the monte carlo method ,

and found that $\nu = \nu(\beta)$ has a local maximum in the interval $\beta \in [0, 1)$

Moreover, ν is not continuous from the left at $\beta = 1$. We propose a possible interpretation to the latter phenomenon. Assume, for example, that N is even, and let μ be the expected time to absorption when $\beta = 1$. Than for $\beta = 1 - \epsilon$ where $\epsilon \rightarrow 0$, we get that the expected time until absorbing will be $(1 - O(\epsilon))\beta$ when a coagulation of 2 and 1 didn't occur, while if such a coagulation did occur (with probability $O(\epsilon)$) we have to wait for another occurrence of such a coagulation until absorbing (because N is even), and therefore in such a case our expected time until absorbing will be $O(\frac{1}{\epsilon})$, so this case contributes $O(1)$ to the total expected time, and so when $\epsilon \rightarrow 0$ we get $\beta + O(1)$ which is larger than β .

The following graphs show the quality and efficiency for $N = 100$.

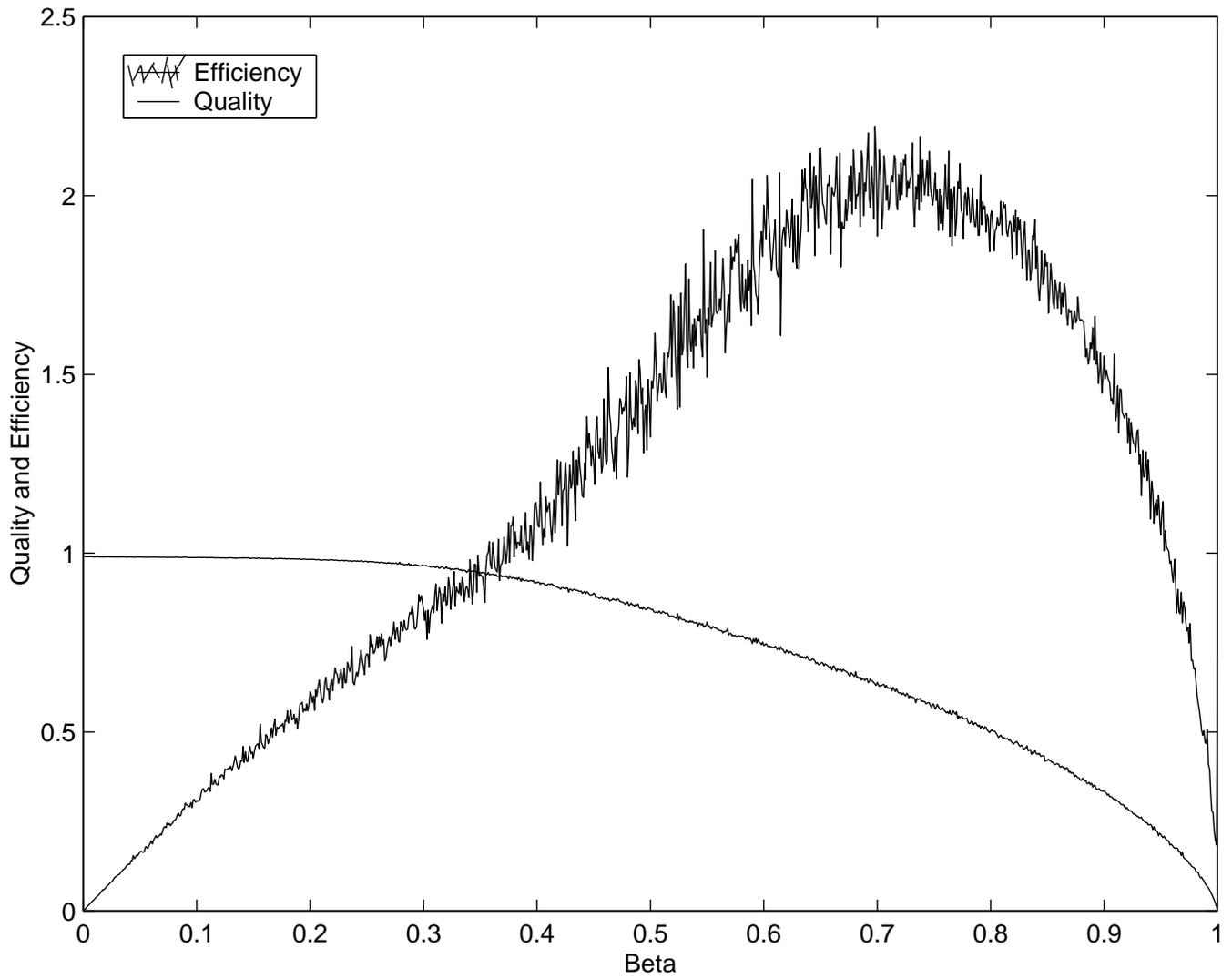


Figure 3.2: The Quality (solid line) And Efficiency(zig-zag line) of the coagulation kernel. The horizontal axis is β .

So, our conclusion is that there are kernels which are better than others with respect to both quality and efficiency, thus there is a motivation for finding such good kernels.

newpage

3.4 Conclusion

Our primary question in this paper was : Can one achieve an optimal group size distribution via a DCP. We have shown that for the simplest model, i.e. a constant kernel, this is not the case and the DCP yields unsatisfying results. However, when choosing the kernel coefficient carefully, it is possible to converge to the optimal group size distribution. Next, we introduced the term efficiency of a kernel, which represent the time until the process converges, and showed that "optimal" kernels have poor efficiency. Thus, we raise the question of the possibility of finding kernels which are both 'optimal' and efficient.

Bibliography

- [1] Caraco, T., and L. L. Wolf. 1975. Ecological determinants of group sizes of foraging lions. *American Naturalist* 109:343-352.
- [2] Giraldeau, L. A. and D. Gillis. 1988. Do lions hunt in group sizes that maximize hunters' daily food returns? *Anim. Behav.* 36:611-613.
- [3] Cooper, S. M. 1991. Optimal hunting group size: the need for lions to defend their kills against loss to spotted hyaenas. *Afr. J. Ecol.* 29:130-136.
- [4] Zemel Amos and Lubin Yael. 1994. Inter-group competition and stable group sizes
- [5] Schaller, G.B. 1972. *The Serengeti Lion*. Univ. Chicago Press, Chicago, IL, USA.

Chapter 4

Comparing strategies in a multi-staged decision game

Abstract

This paper deals with a decision game that can apply to various biological situations. Through simple models, our goal is to investigate the connection between the objective living conditions, and the need to develop complex and sophisticated strategies in order to survive. Our main result is, that both in comfortable and rough conditions, developing an optimal strategy is not of great importance, and a relatively simple memoryless strategy, achieves results almost as good as optimal, and therefore might be enough for surviving.

4.1 Introduction

In various biological situations, an organism faces a stage on which he should make a decision that would have great effect on its surviving chances, due to influence on food amount, number of off springs, chances of escaping from a predator etc.

As evolution and natural selection operate, we might expect the organism to be able to decide the 'correct' decision, that is, select the optimal strategy, based on the knowledge it has when facing a certain situation. Here 'optimal' means a strategy which maximizes, for example the expected amount of food gained, or the expected number of off springs. However, the optimal strategy is usually complex and developing a brain complex enough to perform such a strategy has a cost. Therefore, if the difference between the gain from applying an optimal strategy and from applying a simpler strategy is not large, it may be not cost-effective to develop a sophisticated brain, and therefore relatively primitive species might survive. On the contrary, if the difference is large, then developing a sophisticated brain is crucial for surviving, and in this case only more advanced species may survive.

In order to demonstrate the above, we construct a model of a multi-staged decision game a certain organism (bird in our case) plays. At each stage the bird finds a place on which it is able to nest, and decides if it does indeed take this place, or keeps searching, as the bird's goal is to find the best nesting place possible, which means, for example, the highest expected number of off springs. The difference between different situations or species will be achieved by changing a single parameter of the game, which is the search cost per turn. This parameter can also be treated as the rate on which the bird loses its chances to reproduce. We compare the results obtained for two strategies, the optimal ('smart') strategy, and a more simple ('stupid') strategy.

When examining the influence of search cost, our main conclusion is that if there is a little or no cost at all involved in searching (living, moving) then

there is no significant difference between the smart and the stupid strategies. If there is a relatively large cost involved in searching then the smarter bird is better.

Discrete multi staged decision games have been widely investigated in literature, as various models were proposed, many of which have interesting economical, biological and social science interpretations. As an example, we bring the well known problem of taking the max of T proposals, which are assumed to have a linear order relation. At each stage, we may stop the game or proceed, and our goal is to maximize the probability of stopping at the maximal proposal. It can be shown that the optimal strategy is not to stop at all in the first $a(T)$ proposals, and then stop right after getting a proposal which is better than the first $a(T)$, while $a(T)$ satisfy the asymptotic relation $a(T) \sim \frac{T}{e}$ as $T \rightarrow \infty$. In our problem, however, we want to maximize the expected score, and are not particularly interested in the best proposal.

In [1], a multi-staged decision game is discussed, on which at each step the proposal is assumed to be normally distributed, with parameters μ and σ , and the player can control those parameters. In this example the goal is maximizing the surviving probability, which is the probability of the score to be above a certain bound B , which is the amount of food crucial for surviving. Our model differs since the bird cannot control the proposal's distribution, since we are interested in maximizing the expected score, and since in our model the bird can choose only one proposal while in [1] it eats at each step.

In [3], the connection between body-size and foraging strategy is discussed.

In [4], it is assumed that different foraging strategies yields different searching costs, and it is showed that for seed-harvester ants, the searching costs are so small compared to the expected gain, that they can be neglected. However, in this paper different strategies gives different searching times, so it is stated that time is the main resource to be considered when comparing between different strategies. In our game, however, the searching cost contains both the energy spent and the price of time, and therefore it is important.

4.2 Definition of the process

Consider the following decision game: a (virtual) bird is given a sequence of T proposals. All proposals are drawn from a quality distribution function denoted by F . We assume that the expectation μ of F , namely $\mu = E(F)$ is finite.

We denote $M_F = \min \{x \mid F(x) = 1\}$. If F is the distribution function of a bounded random variable, then $M_F < \infty$. The density function of F is denoted by f .

A strategy for accepting/rejecting a proposal at each step, when the total number of steps is T is denoted by S^T . A strategy can be represented as a sequence of nonnegative numbers $(\alpha_0, \alpha_1, \dots, \alpha_T)$ where in the k -th step, $1 \leq k \leq T$, the proposal is accepted if it is at least α_k .

When the maximal number of steps, T , is given we denote the expected score attained by applying the strategy S^T by $X_T = X_T(S^T)$ (this is the utility function).

$x_i^T = x_i = x_i(S^T)$ is the conditional expected score attained by applying the strategy S^T when only i steps remain. When T is known we will use x_i .

We are interested in maximizing the expected score, that is finding strategies S^T that maximizes $X_T(S^T)$

4.3 Strategies for solution

We compare here two strategies: the “stupid” bird uses a constant bound strategy and the “smart” bird uses an optimal strategy, which we define below.

4.3.1 The Stupid Bird

The stupid bird does not know how many steps are left in the game. Therefore, it uses a single bound strategy, that is accepting any proposal whose score is no less than some constant α . This implies $S^T = (\alpha, \alpha, \dots, \alpha)$. For convenience, we might call the strategy $S^T = (\alpha, \alpha, \dots, \alpha)$, just α . We further assume that (through evolutionary selection) the stupid bird it uses the value of α that maximizes the total expectation from all possible single bound strategies. Thus, for a given number of steps T , we seek to compute the optimal α which is denoted $\alpha(T)$. We define

$$A = A(\alpha) = \int_{\alpha}^{\infty} tf(t)dt \quad (4.1)$$

and

$$B = B(\alpha) = F(\alpha) \quad (4.2)$$

Let x_k be the expected score, when there are exactly k steps left. At this stage, the bird can choose to accept the proposal or to reject it. If the proposal is accepted, with probability $1 - B$, the expected score is $\frac{A}{1-B}$. If it is reject, with probability B , the expected score is x_{k-1} . We therefore obtain the following recurrence relation:

$$x_k(\alpha) = A(\alpha) + B(\alpha)x_{k-1}(\alpha) \quad (4.3)$$

with the initial condition $x_0(\alpha) = \mu$. (We assume that in the last stage the bird always accept the proposal). This yields

$$x_k(\alpha) = A(\alpha) \frac{1 - B(\alpha)^k}{1 - B(\alpha)} + \mu B(\alpha)^k, \quad k = 1, 2, \dots, T. \quad (4.4)$$

The optimal value of $\alpha = \alpha(T)$ which the stupid bird should use, is the value that maximizes $X_T = X_T(\alpha) = x_T(\alpha)$.

More turns yields more opportunities for the bird therefore we expect X_T to be monotonically increasing with respect to T . Moreover, when the number of turns approaches infinity, we expect the gain to be as close as we like to the maximum possible value, M_F . The proof for these observations is as following : Assume, with no loss of generality, that $E(F) = 0$. Then :

$$\begin{aligned} X_T &= A(\alpha(T)) \frac{1 - B(\alpha(T))^T}{1 - B(\alpha(T))} \geq \\ &A(\alpha(T-1)) \frac{1 - B(\alpha(T-1))^T}{1 - B(\alpha(T-1))} \geq \\ &A(\alpha(T-1)) \frac{1 - B(\alpha(T-1))^{T-1}}{1 - B(\alpha(T-1))} = X_{T-1} \end{aligned} \quad (4.5)$$

Now, we know that there exists some X (possibly infinite) with $X_T \nearrow X$. Clearly $X \leq M_F$. To show that equality holds, we suppose that $X < M_F$. In this case, $F(M_F) - F(\frac{M_F+X}{2}) = p > 0$. Therefore, if we set $\alpha(T) = \frac{M_F+X}{2}$ for every T , we get

$$\lim_{T \rightarrow \infty} X_T \geq \lim_{T \rightarrow \infty} (1 - (1-p)^T) \frac{M_F + X}{2} = \frac{M_F + X}{2} > X \quad (4.6)$$

which is a contradiction. Therefore $X_T \nearrow \infty$

4.3.2 The Smart Bird

The smart bird has complete information on the game, and therefore establishes an optimal strategy (that is a strategy that maximizes the expected score X_T for the T steps game). It can be easily shown (by induction), that there exists a non-decreasing sequence of numbers : $\alpha_0, \alpha_1, \dots, \alpha_k, \dots$ such

that if there are exactly k steps remaining, the smart bird will accept a proposal if it's score is not smaller than α_k . Note that this property implies $X_k(S_k) = X_k = x_k = x_k(S_T)$ for every $T \geq k$. Therefore, in this section we may use only the lower case x_k .

For every natural T we wish to find the optimal strategy S_T and the resulting expected score x_T . With the above notations, we use the dynamic programming approach in order to get a recurrence formula for x_T . Suppose that there are exactly k steps left in the game. The bird may accept or reject the current proposal. If the proposal's score is smaller than x_{k-1} , the proposal must be rejected, because the expected score when rejecting the proposal is greater than the score of the proposal itself. For the same reason, the proposal must be accepted if it's score is greater than x_{k-1} . Therefore, we get the relation $\alpha_k = x_{k-1}$. If the proposal is rejected, with probability $B(x_{k-1})$, the expected score is x_{k-1} , while if it is accepted, with probability $1 - B(x_{k-1})$, the expected score is $\frac{A(x_{k-1})}{1 - B(x_{k-1})}$. We therefore obtain the following recurrence formula :

$$x_k = A(x_{k-1}) + B(x_{k-1})x_{k-1}$$

With the initial condition $x_0 = \mu$. (We assume that in the last stage the proposal is always accepted)

The optimality of the smart bird's strategy, can be easily shown by induction, directly from the recurrence formula.

This recurrence formula, however, cannot be brought into a closed form in the general case, and even in specific simple F distributions. Instead, it is often possible to find at least an asymptotic expression for x_T when $T \rightarrow \infty$, as will be shown later.

In similar to the stupid bird, here also the expected gain is monotonically increasing to M_F . Looking at the recurrence formula we get :

$$\begin{aligned} x_k &= A(x_{k-1}) + B(x_{k-1})x_{k-1} = \\ &(1 - F(x_{k-1}))E(x/x > x_{k-1}) + F(x_{k-1})x_{k-1} \geq \end{aligned}$$

$$(1 - F(x_{k-1}))x_{k-1} + F(x_{k-1})x_{k-1} = x_{k-1} \quad (4.7)$$

Therefore x_k is non-decreasing. (It also follows that x_k is strongly increasing unless $F(x_{k-1}) = 1$, that is, we have already reached the maximal value M_F). Since the smart bird's strategy is optimal, it is obvious that $x_k \nearrow M_F$ because this was the case for the sub-optimal strategy of the stupid bird.

Though a closed expression for x_k cannot be obtained usually, it is possible to set a bound on x_k , which is valid for every F . If F is bounded, then $M_F < \infty$, and x_k is of course bounded since $x_k \leq M_F$. If F is not bounded, then $x_k \rightarrow \infty$. In order to establish an asymptotic bound, we first observe that the gain of adding an extra turn, is monotonically decreasing with respect to the number of steps left. This is proven in the following lemma

lemma : Define d_k as the difference between two consecutive elements of the series x_k , that is $d_k = \Delta(x_k) = x_k - x_{k-1}$. Then $d_k \searrow d$ for some real d . If $\mu = E(F) < \infty$ then $d_k \searrow 0$.

proof : First we prove that d_k is a decreasing sequence.

$$\begin{aligned} d_{k+1} = x_{k+1} - x_k &= \int_{x_k}^{\infty} (t - x_k) \cdot f(t) dt < \\ \int_{x_k}^{\infty} (t - x_{k-1}) \cdot f(t) dt &< \int_{x_{k-1}}^{\infty} (t - x_{k-1}) \cdot f(t) dt = \\ x_k - x_{k-1} &= d_k \end{aligned} \quad (4.8)$$

And therefore $d_k \searrow d$ for some real d . Now suppose that $E(F) < \infty$, then :

$$d_k = x_k - x_{k-1} = \int_{x_{k-1}}^{\infty} (t - x_{k-1}) \cdot f(t) dt < \int_{x_{k-1}}^{\infty} t \cdot f(t) dt \quad (4.9)$$

But $x_k \nearrow M_F$, therefore :

$$\lim_{k \rightarrow \infty} \int_{x_k}^{\infty} t \cdot f(t) dt = 0 \quad (4.10)$$

and therefore $d_k \searrow 0$

■

From the last lemma an asymptotic bound on x_k can be easily derived, by the following claim :

claim : If $\mu < \infty$ then $\frac{x_k}{k} \rightarrow 0$ as $k \rightarrow \infty$.

proof : $d_k \rightarrow 0$ as $k \rightarrow \infty$, therefore :

$$\frac{x_k}{k} = \frac{x_0 + \sum_{i=1}^k d_i}{k} \rightarrow 0 \quad (4.11)$$

■

This result implies that $x_k = o(k)$, meaning that the growing rate is sub-linear. Examples of specific distributions which of course satisfy this property are viewed in the next section.

4.3.3 Example : The Uniform Distribution

In this example F is uniformly distributed on the interval $[-1, 1]$, that is $f \sim U[-1, 1]$. Here we get :

$$A = A(\alpha) = \frac{1}{4}(1 - \alpha^2) \quad (4.12)$$

And

$$B = B(\alpha) = \frac{\alpha + 1}{2} \quad (4.13)$$

For the stupid bird we get :

$$x_k(\alpha) = 2\alpha(1 - (\frac{\alpha^2 - 1}{\alpha^2})^k) \quad (4.14)$$

Derivation and comparing to zero gives after simplifying :

$$2 - 2(\frac{\alpha^2 - 1}{\alpha^2})^k(1 + \frac{4k}{\alpha^2 - 1}) = 0 \quad (4.15)$$

And we get a polynomial equation of degree k which we are unable to solve analytically. However, simulations shows that when $T \rightarrow \infty$, then also $T(1 - x_T) \rightarrow \infty$.

For the smart bird we get the following recursion formula :

$$x_k = \frac{(x_{k-1} + 1)^2}{2} \quad (4.16)$$

Next, we we show that $1 - x_k \sim \frac{1}{k}$.

claim : as $k \rightarrow \infty$, we have : $1 - x_k \sim \frac{1}{k}$

proof : write $y_k = 1 - x_k$, then the recursion formula becomes : $y_k = y_{k-1} - \frac{y_{k-1}^2}{2}$.

Now assume that $y_k > \frac{1}{k}$. Then :

$$y_{k+1} = y_k - \frac{y_k^2}{2} > \frac{1}{k} - \frac{1}{4k^2} > \frac{1}{k} - \frac{1}{k(k+1)} = \frac{1}{k+1} \quad (4.17)$$

Assume now that $y_k < \frac{4}{k}$. Then :

$$y_{k+1} = y_k - \frac{y_k^2}{2} < \frac{4}{k} - \frac{4}{k^2} < \frac{4}{k} - \frac{4}{k(k+1)} = \frac{4}{k+1} \quad (4.18)$$

And from these observations the claim is easily proved by induction. Comparing the two birds shows that while the difference in expected score is relatively small (it has one peak whose value is .025) the asymptotic behavior is different, and in fact, we get :

$$\lim_{T \rightarrow \infty} \frac{1 - x_T(\text{smart})}{1 - x_T(\text{stupid})} = 0 \quad (4.19)$$

4.3.4 Example : The Exponential Distribution

In this example F is exponentially distributed with parameter λ , that is $f(t) = \lambda e^{-\lambda t}$. Here we get :

$$A = A(\alpha) = \left(\alpha - \frac{1}{\lambda}\right)e^{-\lambda\alpha} \quad (4.20)$$

And

$$B = B(\alpha) = 1 - e^{-\lambda\alpha} \quad (4.21)$$

For the smart bird we get the following recursion formula :

$$x_k = x_{k-1} + \frac{1}{\lambda}e^{-\lambda x_{k-1}} \quad (4.22)$$

Now, we show that $x_k \sim \ln(k)$.

claim : as $k \rightarrow \infty$, we have : $x_k \sim \ln(k)$

proof : Assume that $x_k > \frac{1}{\lambda} \ln(k)$. Then :

$$x_{k+1} \geq \frac{1}{\lambda} \ln(k) + \frac{1}{\lambda k} > \frac{1}{\lambda} \ln(k+1) \quad (4.23)$$

Assume now that $x_k < \frac{2}{\lambda} \ln(k)$ Then :

$$x_{k+1} \leq \frac{2}{\lambda} \ln(k) + \frac{1}{\lambda k^2} < \frac{1}{\lambda} \ln(k+1) \quad (4.24)$$

And from these observations the claim is easily proved by induction.

For the stupid bird the formula we get is too complicated to be solved analytically. However, numerical solution shows that for a large number of steps the gain of the stupid bird is at the same order of that of the smart bird. Thus, a reasonable assumption is that both possess the same asymptotic behavior.

4.3.5 Example : The Normal Distribution

In this example F is a standard normal distribution, that is $f(t) \sim T(0, 1)$.

Here we get :

$$A = A(\alpha) = \phi(\alpha) \quad (4.25)$$

And

$$B = B(\alpha) = \Phi(\alpha) \quad (4.26)$$

Where ϕ and Φ are, respectively the normal density and probability functions.

For the smart bird we get the following recursion formula :

$$x_k = x_{k-1} \cdot \Phi(x_{k-1}) + \phi(x_{k-1}) \quad (4.27)$$

We cannot solve this recurrence formula, but can give an upper bound to it's asymptotic behavior. when $k \rightarrow \infty$, then x_k is $O(\ln(k))$. Let us prove it :

claim : as $k \rightarrow \infty$, we have : $x_k = O(\ln(k))$

proof : Use the following asymptotic expansion :

$$1 - \Phi(x) \sim \left(\frac{1}{x} - \frac{1}{x^3}\right) \cdot \phi(x), \text{ as } x \rightarrow \infty \quad (4.28)$$

Then we get :

$$x_{k+1} \sim x_k \cdot \left(1 - \frac{1}{x_k} + \frac{1}{x_k^3}\right) \cdot \phi(x_k) + \phi(x_k) = x_k + \frac{1}{\sqrt{2\pi} \cdot x_k^2} e^{-\frac{x_k^2}{2}} \quad (4.29)$$

From here it is obvious that x_k grows more slowly than in the exponential case, and therefore, $x_k = O(\ln(k))$.

Like in the exponential case, we cannot solve analytically for the stupid bird, but numerical solution shows the same phenomena of close results to the smart bird for a large number of steps.

Those examples (and others) yields the following to observations : 1. The asymptotic behavior of the smart and stupid birds benefits are the same 2.

When F is bounded, the asymptotic behavior of the 2 differences $M_F - x_k$ (that is, the second term in the asymptotic expansion) is different for the two birds, (where the term of the smart bird is of course smaller).

The first assertion is proved under some conditions in the next section.

In order to prove similarity in the asymptotic expansion of the two birds, we need to demand certain conditions on the probability distribution function F .

claim : Suppose x_k is the expected benefit of the smart bird when there are k steps left. If :

$$\lim_{k \rightarrow \infty} B(x_k)^k = q < 1 \quad (4.30)$$

then the asymptotic behavior of the smart and stupid birds is the same. Moreover, if $q = 0$ in the last equation, then the ratio between the benefits of the smart and stupid birds approaches 1 as $k \rightarrow \infty$.

proof : Suppose, that for the stupid bird, when there are k steps left, our strategy is to take α to be x_k of the smart bird. (This is clearly a sub-optimal strategy). Denote the expected profit from such a strategy by z_k . Then :

$$z_k = A(x_k) \frac{1 - B(x_k)^k}{1 - B(x_k)} + \mu B(x_k)^k \sim A(x_k) \frac{1 - q}{1 - B(x_k)} + \mu q > (1 - q)x_k \quad (4.31)$$

And, since $1 - q > 0$, we have $z_k = O(x_k)$ and therefore $x_k(\text{stupid}) = O(x_k(\text{smart}))$. If $q = 0$, we get, of course : $z_k \sim x_k$ and therefore $x_k(\text{stupid}) \sim x_k(\text{smart})$.

■

claim :

$$\lim_{T \rightarrow \infty} \frac{X_T(\text{smart})}{X_T(\text{stupid})} = 1 \quad (4.32)$$

proof : Let X_T be the expected score of the smart bird when there are exactly T steps left. Suppose that for the stupid bird, in the T steps game, we take a modified sub-optimal α which is $h \cdot X_T$ where $0 < h < 1$ is an arbitrary constant. (This strategy will, of course, bring benefit which is not more than the stupid bird's optimal strategy α_T). Now denote the expected benefit from such a strategy by Z_T . Also, define :

$$i_T = \min\{i/X_i > h \cdot X_T\} \quad (4.33)$$

(Clearly, for T sufficiently large : $0 < i_T < T$). Then :

$$Z_T = (1 - (B(X_{i_T}))^{T-i_T})E(x/x \geq i_T) \approx (1 - (B(h \cdot X_T))^{T-i_T})E(x/x \geq h \cdot X_T) > \\ (1 - \prod_{i=i_T}^T B(X_i))E(x/x \geq h \cdot X_T) > (1 - \prod_{i=i_T}^T B(X_i))h \cdot X_T. \quad (4.34)$$

Now, we shall prove :

$$\lim_{T \rightarrow \infty} \prod_{i=i_T}^T B(X_i) = 0 \quad (4.35)$$

Define Y_T to be the maximum of T i.i.d R.V. with probability distribution F . Then , clearly, $X_T < EY_T$ (The bird can't achieve the maximum since it cannot go back to previous proposals). Therefore $B(X_T) < B(EY_T)$. But :

$$1 - B(EY_T) = P(x \geq EY_T) \geq P(x \geq Y_T, Y_T \geq EY_T) = \frac{1}{T}P(Y_T \geq EY_T) > \frac{\lambda}{T} \quad (4.36)$$

For some $\lambda > 0$. Therefore, we get :

$$B(X_T) < B(EY_T) < 1 - \frac{\lambda}{T} \quad (4.37)$$

and therefore :

$$\lim_{T \rightarrow \infty} \prod_{i=i_T}^T B(X_i) < \lim_{T \rightarrow \infty} \prod_{i=i_T}^T (1 - \frac{\lambda}{T}) = 0 \quad (4.38)$$

The last limit is indeed 0 since $\frac{i_T}{T} \rightarrow 0$ as $T \rightarrow \infty$. Now, for T sufficiently large, Z_T can be as close as we like to $h \cdot X_T$, but h was an arbitrary constant. Therefore, for larger and larger values of T we may select h to be closer and closer to 1, and therefore, for T sufficiently large, Z_T can be as close as we like to X_T , and therefore we conclude that the stupid birds expected gain can be as close as we like to the smart's one and the two birds are asymptotically equivalent. ■

It can be easily verified that all the examples shown so far satisfy the claim's condition, and therefore, in all those examples the smart and stupid birds are asymptotically equivalent.

4.4 Generalizing the basic model

In this section we would like to extend the results obtained for the basic model to some more complicated cases.

4.4.1 Bounded G

Suppose that G is not constant but bounded by $T < \infty$. Mark by $r(n)$ the probability that we have another step, when there is a total of n steps possible, that is $r(n) = P(x \geq n \mid x \geq n - 1)$ where $x \sim G$. Let us start with the smart bird this time. In the recurrence formula of the basic model we have to replace x_{k-1} by $x_{k-1}r(n - k + 1)$ and the rest stays the same. The recurrence formula now looks like this :

$$x_k = A(x_{k-1} \cdot r(n - k + 1)) + B(x_{k-1} \cdot r(n - k + 1))x_{k-1} \cdot r(n - k + 1) \quad (4.39)$$

With $x_0 = 0$.

For the stupid bird, in the same manner, we get :

$$x_k = A(\alpha \cdot r(n - k + 1)) + B(\alpha \cdot r(n - k + 1))x_{k-1} \cdot r(n - k + 1) \quad (4.40)$$

With $x_0 = 0$. Here we cannot obtain a closed formula for X_k as a function of α , and therefore usually this function should be computed recursively when knowing F , and only then we can proceed like in the basic model, that is find an α which maximizes $X_k(\alpha)$

A natural question concerning our game, is the influence of uncertainty on the bird's strategy and utility. More precisely, given that the expectation of G remains constant, how does changing the variance affects the output of the game. In all simulations done with uniform and binomial distributions, it turned out that the larger the variance the smaller the bird gains. This can be explained by the following intuitive argument : Increasing the variance increase the probability of getting values which are far from the expectation. However, since the differences series d_k is monotonically decreasing, getting

a very low number of steps has a more drastic affect on the payment then a very high number of steps. Therefore, in the mean, the bird looses. For the uniform distribution, it can be actually proved, at least for the smart bird, that the best G is the constant G .

claim : For each $k = 0, 1, \dots, T - 1$ Let $G \sim U[T - k, T + k]$ and mark by z_k the resulting X_{T+k} . Then :

$$z_k < z_0, \forall k > 0 \quad (4.41)$$

proof : Denote by X_i^* the X_i we would have get from a game with constant G , with i steps. Clearly, if in our game exactly i steps have occurred, the bird's payment is not larger than X_i^* , since the last value is the output of the optimal strategy for the game with exactly i steps. Therefore, we get :

$$z_k \leq \frac{1}{2k + 1} \cdot \sum_{i=T-k}^{T+k} X_i^* = \frac{1}{2k + 1} \cdot (X_T^* + \sum_{i=1}^k (X_{T+k}^* + X_{T+k}^*)) \quad (4.42)$$

But because d_k is a decreasing sequence , we get :

$$X_{T+k}^* + X_{T+k}^* < 2X_T^* \quad (4.43)$$

And the claim immediately follows.

4.4.2 Time dependent utility

Suppose that at the k -th turn, the utility distribution is F_k . If we set A_k and B_k to be the A and B corresponding to F_k , then the smart's bird's formula will be :

$$x_{k+1} = A_{k+1}(x_k) + B_{k+1} \quad (4.44)$$

While for the stupid bird, we get :

$$x_{k+1} = A_{k+1}(\alpha) + B_{k+1}(\alpha)x_k \quad (4.45)$$

And we cannot express $x_{k+1}(\alpha)$ in closed form as we did in the previous cases.

4.4.3 Searching Cost

In this section we try to figure out how the results obtained for the smart and stupid birds in the previous chapters, are changed when searching has a cost. Suppose that for every turn that passes, the bird pays a constant amount of utility (because waiting demands some of the bird's resources) - C . Then, for the smart bird, our recursion formula becomes :

$$x_{k+1} = A(x_k - C) + B(x_k - C)(x_k - C) \quad (4.46)$$

Note that here x_k is still monotonically increasing, but the limit is no longer M_F but smaller. In fact, it can be shown that the limit is always final even if F is unbounded, as long as $E(F) < \infty$.

claim : If $\mu < \infty$ and $C > 0$ then :

$$\lim_{k \rightarrow \infty} x_k < \infty \quad (4.47)$$

proof : Clearly, x_k is a monotonically increasing sequence, with $x_0 = \mu$. In order to show that the limit is final it is enough to show that there exists a final solution $x > \mu$ to the equation :

$$x = A(x - C) + B(x - C)(x - C) \quad (4.48)$$

Or :

$$C = \int_{x-C}^{\infty} (t - x + C) f(t) dt \equiv j(x) \quad (4.49)$$

But, it is easy to see that $j(x)$ is monotonically decreasing, and that

$$\lim_{x \rightarrow \infty} j(x) = 0$$

so for C small enough there is a finite solution to the equation, which is larger than μ , and therefore the limit is finite. For larger values of C the limit is surely finite because it is monotonically decreasing with respect to C . ■

So, to conclude, we have shown that for the smart bird we have $x_k \nearrow x < \infty$, where x is the solution of the previous equation.

For the stupid bird, we get :

$$x_{k+1}(\alpha) = A(\alpha) + B(\alpha) \cdot (x_k - C), \quad k = 1, 2, \dots, T. \quad (4.50)$$

And the solution is :

$$X_T(\alpha) = x_T(\alpha) = \frac{(A(\alpha) - C \cdot B(\alpha))(1 - B(\alpha)^T)}{1 - B(\alpha)} + B(\alpha)^T \cdot (\mu - C) \quad (4.51)$$

Here, as in the case of no cost per turn, we have to find the optimal α and then calculate X_T . The limit of X_T when $T \rightarrow \infty$ will of course be also finite, since it is smaller than the limit for the smart bird. When letting T approach ∞ , and keeping a constant α , we get :

$$X_T(\alpha) \rightarrow X(\alpha) = \frac{A(\alpha) - C \cdot B(\alpha)}{1 - B(\alpha)} \quad (4.52)$$

Define :

$$\alpha_\infty = \lim_{T \rightarrow \infty} \alpha(T) \quad (4.53)$$

Then, from continuity :,

$$\lim_{T \rightarrow \infty} X_T = X(\alpha_\infty), \quad (4.54)$$

And α_∞ can be found from deviating $X(\alpha)$ and comparing to zero. Doing so, we get the following equation :

$$0 = X'(\alpha) = \frac{(-\alpha f(\alpha) - C f(\alpha))(1 - B(\alpha)) + f(\alpha)(A(\alpha) - C B(\alpha))}{(1 - B(\alpha))^2} = \frac{f(\alpha)(A(\alpha) + \alpha(B(\alpha) - 1) - C)}{(1 - B(\alpha))^2} \quad (4.55)$$

We therefore get :

$$f(\alpha) = 0 \quad \text{or} \quad C = \int_\alpha^\infty (t - \alpha) f(t) dt \quad (4.56)$$

It can be shown by deviating $X(\alpha)$ further, that the left equation will give us a saddle point, while the right one will give us the desired maximum. As we

see the equation on the right is very similar to the one obtained for the smart bird. Comparing the two equations yields :

$$\alpha_{\infty}(stupid) = X(smart) - C \quad (4.57)$$

Assigning in $X(\alpha)$ for the stupid bird, we get :

$$\begin{aligned} X(stupid) &= \frac{A(X(smart) - C) - CB(X(smart) - C)}{1 - B(X(smart) - C)} = \\ &= \frac{X(smart) - (X(smart) - C)B(X(smart) - C) - CB(x(smart) - C)}{1 - B(X(smart) - C)} = \\ &= \frac{X(smart)(1 - B(X(smart) - C))}{1 - B(X(smart) - C)} = X(smart) \end{aligned} \quad (4.58)$$

And therefore, just like in the no-cost model, the smart and stupid strategies yields similar results when $T \rightarrow \infty$, meaning that the limit is the same for both strategies, even though for every finite T , the smart strategy is of course slightly better. This could suggest an explanation for success of relatively simple species which don't have the concept of time, and therefore have a constant bound when deciding whether to accept a suggestion or not. It seems, that after all, those species doesn't lose much, compared to ideal species which use the optimal strategy.

Bibliography

- [1] Stephens D.K. and KREBS J.R. (1986). Foraging Theory. Princeton University Press, Princeton
- [2] Bellman R.E. (1957). Dynamic Programming. Princeton University Press, Princeton.
- [3] Yamauchi A. and Iwasa Y. (1995) .Coupling of fermentation and foraging strategies of herbivorous mammals, Journal of Theoretical Biology, v 172, n 1, January 7, 1995, p1-11 (ID jt950001)
- [4] Weier J.A. and Feener D.H. (1995). Foraging in the seed-harvester ant genus *Pogonomyrmex*: are energy costs important? Behavioral Ecology and Sociobiology Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA
- [5] K. S. Erickson and D. H. Morse (1996). Predator size and the suitability of a common prey Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, RI 02912, USA
- [6] Loeve M. (1977). Probability Theory 1. Springer-Verlag.
- [7] Ben Gum (2000) The secretary problem with uniform distribution Princeton , USA
- [8] T. S. Ferguson (1989) Who solved the secretary problem. Statistical Science 4, 282-296

Appendix: Description of the numerical procedure

The numerical procedure is:

