A Note on Varying Mutation Rates in 2×2 Coordination Games*

Alexander F. Tieman† and Harold Houba‡

June 9, 1998

Abstract

The model of 2×2 coordination games in Kandori, Mailath, and Rob (1993) is extended to allow for a mutation rate that is stochastic over time. The expected time the system spends in the risk dominated equilibrium is systematically underestimated by the standard model in Kandori, Mailath, and Rob (1993) when the latter model's (fixed) mutation rate is equal to the expected mutation rate. A small population result corrects a minor omission in Kandori, Mailath, and Rob (1993).

*This paper has benefitted significantly from comments by and discussions with Gerard van der Laan, Nam-Kyoo Boots, Oddvar Kaarboe and various participants at the 1998 Stony Brook Summer Festival on Game Theory. The authors would like to thank Larry Samuelson for encouraging this research project and participants at the 1997 Summer school in Oberwesel, Germany, for comments.

†Tinbergen Institute and Free University, Department of Econometrics, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands. E-mail: xtieman@econ.vu.nl, Phone: +31-20-4446022, Fax: +31-20-4446020, URL: http://www.econ.vu.nl/~xtieman/ .

‡Free University.

1. Introduction.

The central message in both Kandori, Mailath, and Rob (1993) (henceforth KMR) and Young (1993) (and in related model like Ellison (1993), Bergin and Lipman (1996) or Rhode and Stegeman (1996)) is that noise in the form of mutations matters for the selection of one Nash equilibrium among strict Nash equilibria. In this note the evolutionary process for 2×2 coordination games in KMR[1] is extended to replace the constant mutation rate by a per-period mutation rate that is small but stochastic over time. To be precise, in each time period the mutation rate is a draw from a non-degenerate i.i.d. random variable with support on a subset of a small bounded interval of non-negative numbers. A non-constant mutation rate is closer to (economic) reality than a fixed mutation rate, since in the three interpretations of mutations most frequently mentioned in the literature, namely experimentation, (computational) errors in the implementation of an action and genetic mutation, are all three difficult to reconcile with the assumption that the mutation rate is constant, as already noted by Bergin and Lipman (1996). Events outside of the model will cause economic subjects to experiment more, make more mistakes or experience a higher rate of genetic mutation at certain times than at others. We model such outside influence indirectly through the stochastic mutation rate.

Our main result states that in 2×2 coordination games the expected time the system spends in the risk dominated equilibrium (the equilibrium with the smaller basin of attraction) is systematically underestimated by the results of KMR when the constant mutation rate in the latter model is set equal to the expected per-period mutation rate in our extended model. A numerical example shows that the magnitude of the underestimation can be quite large. These results are of importance for (economic) phenomena that can be modelled as coordination games, e.g. for the macroeconomic literature on coordination failure (see e.g. Cooper and John (1988)).

Our main result is derived for a sufficiently large population. However, we also investigate small populations and find different results. As a topic of related interest our analysis reveals a minor mistake in the results reported in KMR and we provide the correct specification.

In most of the literature, the mutation rate is taken to zero in the limit in order to state a

[1] It is not hard to make the same extension in the closely related model of Young (1993).

clear result on the invariant distribution on the state space. Performing a similar exercise in our extended model by letting the support of the stochastic mutation rate pile up at 0 in the limit would yield the same invariant distribution as found in KMR. For that reason we omit this routine exercise. Furthermore, in all of the proofs attention is restricted to the case in which the Darwinian dynamics are the best reply dynamics. Subsequently, we argue that the results holds for all Darwinian dynamics.

2. The Model.

In this note we use the framework of KMR, section 7. Specifically, we focus on the (fixed) coordination game

$1\backslash 2$	s_1	s_2
s_1	a, a	$0, 0$
s_2	$0, 0$	$1, 1$

with $a > 1$, played repeatedly in a population consisting of N players. By $z_t \in Z$, $Z = \{0, 1, \ldots, N\}$, we denote the number of players adopting action s_1 at time t. Thus $z_t = N$, denoted by $E_1 = (s_1, s_1)$, yields the risk dominant equilibrium at any play of the stage game between two players from the population, while $z_t = 0$, denoted by $E_2 = (s_2, s_2)$, yields the risk dominated equilibrium. The population state $z^* = \frac{N+a-1}{a+1}$ reflects the (unstable) mixed equilibrium. By the entier function $[z^*]$, we denote the largest integer smaller or equal to z^*. We focus on best reply deterministic dynamics given by

$$z_{t+1} = B(z) = \begin{cases} 0, & \text{if } z_t < z^*, \\ z^*, & \text{if } z_t = z^*, \\ N, & \text{if } z_t > z^*. \end{cases}$$

As in KMR we add a stochastic component representing random mutation. The best reply dynamics with random mutation enable us to model the process of switching between states E_1 and E_2 as a two state Markov chain. We define $P(\varepsilon)$ to be the transition matrix of this Markov chain, where $\varepsilon > 0$ is the constant mutation rate. Now, the expected time spent in the equilibrium E_1 respectively E_2 is given by $\frac{1}{p(\varepsilon)}$ and $\frac{1}{p'(\varepsilon)}$, where

$$p(\varepsilon) = \sum_{j=N-[z^*]}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j} \text{ and } p'(\varepsilon) = \sum_{j=[z^*]+1}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}.$$

The transition matrix $P(\varepsilon)$ of this Markov chain is given by

$$P(\varepsilon) = \begin{pmatrix} 1 - p(\varepsilon) & p(\varepsilon) \\ p'(\varepsilon) & 1 - p'(\varepsilon) \end{pmatrix}$$

and the stationary distribution, denoted by $\mu(P(\varepsilon)) = (\mu_1(P(\varepsilon)), \mu_2(P(\varepsilon)))$, only puts weight on the two states E_1 and E_2. The weight on E_1 respectively E_2 is given by

$$\mu_1(P(\varepsilon)) = \frac{p'(\varepsilon)}{p(\varepsilon) + p'(\varepsilon)} \text{ and } \mu_2(P(\varepsilon)) = 1 - \mu_1(P(\varepsilon)) = \frac{p(\varepsilon)}{p(\varepsilon) + p'(\varepsilon)}.$$

The innovative feature of our model lies in the assumption that the mutation rate ε_t at time t is stochastic. At every time t, the mutation rate ε_t is a realization of a random variable θ with an arbitrary non-degenerate, discrete probability distribution Θ, which has its support on a subset of $[0, \bar{\varepsilon}]$, for some $\bar{\varepsilon} > 0$. Since every continuous probability distribution can be approximated arbitrarily close by a discrete probability distribution there is no loss in generality in using a discrete Θ.

The case of a stochastic mutation rate yields an inhomogeneous Markov chain, which can be handled as follows. In every period, the dynamics consist of a compounded lottery. At each time t, first a draw ε_t from $\theta \sim \Theta$ determines the mutation rate for period t, and next, the mutation rate ε_t determines the transition probabilities $p(\varepsilon_t)$ and $p'(\varepsilon_t)$ in the transition matrix of the Markov chain. These two stochastic events can be compounded into a single transition matrix $\mathbb{E}P(\theta)$, $\theta \sim \Theta$, where

$$\mathbb{E}P(\theta) = \begin{pmatrix} 1 - \mathbb{E}p(\theta) & \mathbb{E}p(\theta) \\ \mathbb{E}p'(\theta) & 1 - \mathbb{E}p'(\theta) \end{pmatrix},$$

and $\mathbb{E}(.)$ is the expectation operator. Therefore, the invariant distribution of our extended model is the invariant probability distribution of the transition matrix $\mathbb{E}P(\theta)$.

3. Results.

Our main result states an order on the following three numbers, namely $\mu_2(P(\mathbb{E}\theta))$ is the smallest number, $\mu_2(\mathbb{E}P(\theta))$ is the largest number and $\mathbb{E}\mu_2(P(\theta))$ lies in between these two numbers. Since $\mu_2(\mathbb{E}P(\theta))$ corresponds to the true invariant distribution the other two can be regarded as (inexact) approximations of $\mu_2(\mathbb{E}P(\theta))$. The number $\mu_2(P(\mathbb{E}\theta))$ corresponds to the standard model in KMR, where the constant mutation rate is taken equal to the per-period expected mutation rate $\mathbb{E}\theta$ of the stochastic variable $\theta \sim \Theta$. The main result shows that the latter interpretation systematically underestimates the expected time the system spends in the dominated equilibrium E_2. The number $\mathbb{E}\mu_2(P(\theta))$ refers to the standard KMR model with the constant mutation rate ε_0 drawn randomly at time $t = 0$ and, once drawn, remaining fixed at ε_0 over time. Then $\mathbb{E}\mu_2(P(\theta))$ is the ex-ante (before drawing ε_0) expected probability of being in E_2 at an arbitrary time t far enough away from $t = 0$. This number is a better approximation for $\mu_2(\mathbb{E}P(\theta))$ than $\mu_2(P(\mathbb{E}\theta))$. The order of these three numbers is the result stated in the following theorem, which applies to relatively large populations.

Theorem 3.1. *Let $a > 1 + \frac{2}{N-3}$ for N odd and $a > 1 + \frac{4}{N-4}$ for N even. Then there exists an $\overline{\varepsilon} > 0$ such that $\mu_2(P(\mathbb{E}\theta)) < \mathbb{E}\mu_2(P(\theta)) < \mu_2(\mathbb{E}P(\theta))$, where θ is a random variable with distribution Θ, with Θ an arbitrary non-degenerate discrete probability distribution with support on a subset of $[0, \overline{\varepsilon}]$.*

Proof of the first inequality in Theorem 3.1.

In order to proof the first inequality, it suffices to show that $\mu_2(P(\varepsilon))$ is convex in ε and to apply Jensen's inequality. Substituting $p(\varepsilon)$ and $p'(\varepsilon)$ in $\mu_2(P(\varepsilon))$ yields

$$\mu_2(P(\varepsilon)) = \frac{\sum_{j=N-[z^*]}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}}{\sum_{j=N-[z^*]}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j} + \sum_{j=[z^*]+1}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}}.$$

By dividing both the numerator and the denominator by $\varepsilon^{[z^*]+1}(1-\varepsilon)^{N-[z^*]-1}$ we get

$$\mu_2(P(\varepsilon)) = \frac{\sum_{j=N-2[z^*]-1}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j}{\sum_{j=N-2[z^*]-1}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j + \sum_{j=0}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j},$$

which is equivalent to

$$\mu_2(P(\varepsilon)) = \frac{\sum_{j=N-2[z^*]-1}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j}{2 \cdot \sum_{j=N-2[z^*]-1}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j + \sum_{j=0}^{N-2[z^*]-2} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j}.$$

We define the terms $C(\varepsilon)$ and $D(\varepsilon)$ as

$$C(\varepsilon) = \sum_{j=0}^{N-2[z^*]-2} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j \tag{3.1}$$

and

$$D(\varepsilon) = \sum_{j=N-2[z^*]-1}^{N-[z^*]-1} \binom{N}{j+[z^*]+1} \left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j \tag{3.2}$$

leading to $\mu_2(P(\varepsilon)) = \frac{D(\varepsilon)}{C(\varepsilon)+2D(\varepsilon)}$.

The second derivative of $\mu_2(P(\varepsilon))$ with respect to ε, $\mu_2''(P(\varepsilon))$, is given by

$$\frac{\{C(\varepsilon)D''(\varepsilon) - C''(\varepsilon)D(\varepsilon)\}\{C(\varepsilon)+2D(\varepsilon)\}}{\{C(\varepsilon)+2D(\varepsilon)\}^3} + \frac{2\{C'(\varepsilon)D(\varepsilon) - C(\varepsilon)D'(\varepsilon)\}\{C'(\varepsilon)+2D'(\varepsilon)\}}{\{C(\varepsilon)+2D(\varepsilon)\}^3}.$$

Explicit expressions for the derivatives $C'(\varepsilon)$, $C''(\varepsilon)$, $D'(\varepsilon)$ and $D''(\varepsilon)$ can be found in the appendix. From these expressions we see that the denominator of $\mu_2''(P(\varepsilon))$ is strictly positive for all $\varepsilon \in (0,1)$ and that the numerator of $\mu_2''(P(\varepsilon))$ is a higher order polynomial function of ε. Label the smallest positive root of the numerator of $\mu_2''(P(\varepsilon))$ as $\bar{\varepsilon}_1$. If the polynomial has no root larger than 0, we set $\bar{\varepsilon}_1$ to 1. The function $\mu_2''(P(\varepsilon))$ is either positive or negative on the whole interval $(0,\bar{\varepsilon}_1)$. To determine the sign of $\mu_2''(P(\varepsilon))$ on $(0,\bar{\varepsilon}_1)$, we distinguish the case N even and the case N odd.

For N even, note that since $[z^*] < \frac{1}{2}N$ is always integer, $N - 2[z^*] - 1$ can only take the values $1, 3, 5, \ldots$. The condition $a > 1 + \frac{4}{N-4}$ guarantees that $N - 2[z^*] - 1 > 1$, and thus that $N - 2[z^*] - 1 \geq 3$, ensuring that $D(0) = D'(0) = D''(0) = 0$, since the lowest power of ε in these polynomials is at least 1. Thus $\mu_2'(P(\varepsilon))$ and $\mu_2''(P(\varepsilon))$ are both 0 in $\varepsilon = 0$. Furthermore, we see that $\mu_2(P(0)) = 0$ and $\mu_2(P(\varepsilon)) > 0$ for $\varepsilon > 0$, leading to the conclusion $\mu_2'(P(\varepsilon)) > 0$ for small $\varepsilon > 0$. Repeating this line of reasoning we conclude from $\mu_2'(P(0)) = 0$ and $\mu_2'(P(\varepsilon)) > 0$, for small $\varepsilon > 0$, that $\mu_2''(P(\varepsilon)) > 0$ for small $\varepsilon > 0$. Thus $\mu_2''(P(\varepsilon))$ is positive on $(0,\bar{\varepsilon}_1)$.

Now consider the case N odd. Note that since $[z^*] < \frac{1}{2}N$ is always integer, $N - 2[z^*] - 1$ can only take the values $0, 2, 4, \ldots$. Furthermore, the condition $a > 1 + \frac{2}{N-3}$ guarantees that $N - 2[z^*] - 1 > 0$, and thus that $N - 2[z^*] - 1 \geq 2$, ensuring that $D(0) = D'(0) = 0$. We now consider two cases, $D''(0) = 0$, when $N - 2[z^*] - 1 > 2$ and $D''(0) > 0$, when $N - 2[z^*] - 1 = 2$. In the former case ($D''(0) = 0$), the same argument used above for N even ensures that $\mu_2''(P(\varepsilon)) > 0$ on $(0,\bar{\varepsilon}_2)$, for some $\bar{\varepsilon}_2 > 0$.[2] In the latter case ($D''(0) > 0$), the numerator of $\mu_2''(P(\varepsilon))$ is equal to $C^2(\varepsilon)D''(\varepsilon) > 0$ in $\varepsilon = 0$. By a continuity argument this proves that $\mu_2''(P(\varepsilon)) > 0$ for all $\varepsilon \in (0,\bar{\varepsilon}_3)$, for some $\bar{\varepsilon}_3 > 0$.

Thus $\mu_2(P(\varepsilon))$ is a convex function on $(0,\bar{\varepsilon})$, with $\bar{\varepsilon} = \min\{\bar{\varepsilon}_1, \bar{\varepsilon}_2, \bar{\varepsilon}_3\} > 0$. Applying Jensen's inequality results in $\mu_2(P(\mathbb{E}\theta)) < \mathbb{E}\mu_2(P(\theta))$. □

[2] The maximal $\bar{\varepsilon}_2$ is the smallest positive root of $\mu_2''(P(\varepsilon))$, which will generically be different from $\bar{\varepsilon}_1$.

Proof of the second inequality in Theorem 3.1.

In this proof we will first focus attention on a Bernoulli distribution Θ, i.e. $\Pr(\varepsilon_t = \varepsilon_{high}) = \lambda$, $\Pr(\varepsilon_t = \varepsilon_{low}) = 1 - \lambda$, $\varepsilon_{high}, \varepsilon_{low} > 0$. Subsequently we will argue that the obtained results hold for all non-degenerate discrete probability distributions Θ.

We assume w.l.o.g. that $\varepsilon_{low} = \varepsilon > 0$ and $\varepsilon_{high} = (1+\alpha)\varepsilon$, $\alpha > 0$. In table 3.1, we have calculated $\mathbb{E}\mu_i(P(\theta)) = \lambda\mu_i(P(\{1+\alpha\}\varepsilon)) + (1-\lambda)\mu_i(P(\varepsilon))$, $i = 1,2$, and $\mu_i(\mathbb{E}P(\theta)) = \mu_i(\lambda P(\{1+\alpha\}\varepsilon) + (1-\lambda)P(\varepsilon))$, $i = 1,2$.

i	$\mathbb{E}\mu_i(P(\theta))$	$\mu_i(\mathbb{E}P(\theta))$
1	$\lambda\frac{p'((1+\alpha)\varepsilon)}{p((1+\alpha)\varepsilon)+p'((1+\alpha)\varepsilon)} + (1-\lambda)\frac{p'(\varepsilon)}{p(\varepsilon)+p'(\varepsilon)}$	$\frac{\lambda p'(\varepsilon)-\lambda p'((1+\alpha)\varepsilon)-p'(\varepsilon)}{\lambda p(\varepsilon)-\lambda p((1+\alpha)\varepsilon)-p(\varepsilon)+\lambda p'(\varepsilon)-\lambda p'((1+\alpha)\varepsilon)-p'(\varepsilon)}$
2	$\lambda\frac{p((1+\alpha)\varepsilon)}{p((1+\alpha)\varepsilon)+p'((1+\alpha)\varepsilon)} + (1-\lambda)\frac{p(\varepsilon)}{p(\varepsilon)+p'(\varepsilon)}$	$\frac{\lambda p(\varepsilon)-\lambda p((1+\alpha)\varepsilon)-p(\varepsilon)}{\lambda p(\varepsilon)-\lambda p((1+\alpha)\varepsilon)-p(\varepsilon)+\lambda p'(\varepsilon)-\lambda p'((1+\alpha)\varepsilon)-p'(\varepsilon)}$

Table 3.1: The expressions $\mathbb{E}\mu_i(P(\theta))$ and $\mu_i(\mathbb{E}P(\theta))$ for $i = 1,2$.

To show that

$$\begin{aligned}
& \lambda\frac{p((1+\alpha)\varepsilon)}{p((1+\alpha)\varepsilon)+p'((1+\alpha)\varepsilon)} + (1-\lambda)\frac{p(\varepsilon)}{p(\varepsilon)+p'(\varepsilon)} \\
< \quad & \frac{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)}{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)+p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)}
\end{aligned}$$

we rewrite the LHS, resulting in

$$\begin{aligned}
& \frac{p(\varepsilon)p((1+\alpha)\varepsilon)+p(\varepsilon)p'((1+\alpha)\varepsilon)}{\{p((1+\alpha)\varepsilon)+p'((1+\alpha)\varepsilon)\}\{p(\varepsilon)+p'(\varepsilon)\}} \\
< \quad & \frac{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)}{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)+p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)}.
\end{aligned}$$

Since both numerators and denominators are positive, cross multiplication results in

$$\begin{aligned}
& \left\{p(\varepsilon)p((1+\alpha)\varepsilon)+p(\varepsilon)p'((1+\alpha)\varepsilon)\right\}\cdot \\
& \{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)+p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)\} \\
< \quad & \left\{p((1+\alpha)\varepsilon)+p'((1+\alpha)\varepsilon)\right\}\left\{p(\varepsilon)+p'(\varepsilon)\right\}\{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)\}.
\end{aligned}$$

We substract the common term

$$\begin{aligned}
& p^2(\varepsilon)p((1+\alpha)\varepsilon)-\lambda p(\varepsilon)^2p((1+\alpha)\varepsilon)+\lambda p(\varepsilon)p^2((1+\alpha)\varepsilon)+p(\varepsilon)p((1+\alpha)\varepsilon)p'(\varepsilon) \\
& -\lambda p(\varepsilon)p((1+\alpha)\varepsilon)p'(\varepsilon)+\lambda p(\varepsilon)p((1+\alpha)\varepsilon)p'((1+\alpha)\varepsilon)+p^2(\varepsilon)p'((1+\alpha)\varepsilon) \\
& -\lambda p^2(\varepsilon)p'((1+\alpha)\varepsilon)+p(\varepsilon)p'((1+\alpha)\varepsilon)p'(\varepsilon)-\lambda p(\varepsilon)p'((1+\alpha)\varepsilon)p'(\varepsilon)
\end{aligned}$$

from both sides and get

$$
\begin{aligned}
& \left\{p\left((1+\alpha)\varepsilon\right)+p^{'}\left((1+\alpha)\varepsilon\right)\right\}\left\{\lambda p\left(\varepsilon\right)p^{'}\left((1+\alpha)\varepsilon\right)\right\} \\
< & \left\{p\left((1+\alpha)\varepsilon\right)+p^{'}\left((1+\alpha)\varepsilon\right)\right\}\left\{\lambda p\left((1+\alpha)\varepsilon\right)p^{'}\left(\varepsilon\right)\right\}.
\end{aligned}
$$

Thus, it remains to be shown that

$$
\lambda p\left(\varepsilon\right)p^{'}\left((1+\alpha)\varepsilon\right)<\lambda p\left((1+\alpha)\varepsilon\right)p^{'}\left(\varepsilon\right). \tag{3.3}
$$

We divide by $\lambda>0$ and write down the Taylor expansion for small α around $p^{'}\left(\varepsilon\right)$ and $p\left(\varepsilon\right)$

$$
p\left(\varepsilon\right)\left\{p^{'}\left(\varepsilon\right)+\alpha\frac{d}{d\varepsilon}p^{'}\left(\varepsilon\right)+O\left(\alpha^{2}\right)\right\}<p^{'}\left(\varepsilon\right)\left\{p\left(\varepsilon\right)+\alpha\frac{d}{d\varepsilon}p\left(\varepsilon\right)+O\left(\alpha^{2}\right)\right\}.
$$

Then we rewrite the expression disregarding the higher order term, deduct the common term $p\left(\varepsilon\right)p^{'}\left(\varepsilon\right)$ and divide by α, resulting in $p\left(\varepsilon\right)\frac{d}{d\varepsilon}p^{'}\left(\varepsilon\right)<p^{'}\left(\varepsilon\right)\frac{d}{d\varepsilon}p\left(\varepsilon\right).$ We now write down these terms explicitly and isolate the lowest orders of ε on both sides. Furthermore we divide by the common term and neglect all powers of $1-\varepsilon\approx1$. This results in

$$
\left([z^{*}]+1\right)\varepsilon^{N}-\left(N-[z^{*}]-1\right)\varepsilon^{N+1}<\left(N-[z^{*}]\right)\varepsilon^{N}-[z^{*}]\,\varepsilon^{N+1},
$$

an inequality that is true for $[z^{*}]<\frac{N}{2}-\frac{1}{2}$ (i.e. $a>\frac{N-1}{N-3}$), since then $N-[z^{*}]>[z^{*}]+1$ and $[z^{*}]<N-[z^{*}]-1$. Note that we do not need the restriction $a>\frac{N}{N-4}$ for N even in this proof. Furthermore, note that we made use of the fact that $\varepsilon>0$ is small, resulting in an upperbound $\overline{\varepsilon}_{4}>0$ for the value of ε for which this proof hold.

A simple induction argument now shows that the inequality holds for all non-degenerate discrete probability distributions Θ with support on a subset of the interval $[0,\overline{\varepsilon}_{5}]$, where $\overline{\varepsilon}_{5}>0$ is an upperbound that can be from an expression similar to (3.3) for any specific distribution. We conclude the proof with the remark that the $\overline{\varepsilon}$ in the Theorem is the minimum of the upperbounds mentioned in the proof, thus $\overline{\varepsilon}=\min_{i}\overline{\varepsilon}_{i}$. □

Corollary 3.2. *Let $a>1+\frac{2}{N-3}$ for N odd and $a>1+\frac{4}{N-4}$ for N even. Then there exists an $\overline{\varepsilon}>0$ such that $\mu_{1}(P(\mathbb{E}\theta))>\mathbb{E}\mu_{1}(P(\theta))>\mu_{1}(\mathbb{E}P(\theta))$, where θ is a random variable with distribution Θ, with Θ an arbitrary non-degenerate discrete probability distribution with support on a subset of $[0,\overline{\varepsilon}]$.*

Driving the proof is the result that the probability of switches between the two equilibria is higher than in the fixed mutation rate model, i.e. $\mathbb{E}p(\theta) > p(\mathbb{E}\theta)$ and $\mathbb{E}p'(\theta) > p'(\mathbb{E}\theta)$, and that the probability of a switch from E_1 to E_2 increases faster than the probability of a switch from E_2 to E_1. Furthermore, it is clear from the proof that the expected time spent in the risk dominated equilibrium grows more than linear in the probability mass that is put on high mutation rate. Note however, that $\mu_2(\varepsilon) \leq \frac{1}{2}$ for all $\varepsilon \in [0, 1]$.

In order to obtain some insight in the magnitude of the underestimation, we consider the example $a = 2$, $N = 10$ and $\Pr(\varepsilon_t = \frac{1}{50}) = \Pr(\varepsilon_t = \frac{1}{200}) = \frac{1}{2}$. Then $\mathbb{E}\theta = \frac{1}{80}$ and

$$\mu_2(P(\mathbb{E}\theta)) = 1.15 \times 10^{-6} \text{ and } \mu_2(\mathbb{E}P(\theta)) = 4.76 \times 10^{-6}.$$

Thus, $\mu_2(\mathbb{E}P(\theta))$ is roughly 315% larger than $\mu_2(P(\mathbb{E}\theta))$, showing a rather severe underestimation of the time spent in E_2 when $\mathbb{E}\theta$ is taken as the mutation rate in the standard KMR framework.

The above results only holds for sufficiently large populations. At this point we take a further look at $1 < a \leq 1 + \frac{4}{N-4}$, i.e. $N - 2[z^*] - 1 = 1$, for N even and at $1 < a \leq 1 + \frac{2}{N-3}$, i.e. $N - 2[z^*] - 1 = 0$, for N odd. From the restriction on a and $\lim_{N\to\infty} 1 + \frac{4}{N-4} = \lim_{N\to\infty} 1 + \frac{2}{N-3} = 1$ it is clear that for a fixed a, the following results only hold for small N and thus that they are only useful in relatively small populations. The following proposition states that for small populations, $\mu_2(P(\mathbb{E}\theta))$ is a better (but still inaccurate) approximation of $\mu_2(\mathbb{E}P(\theta))$ than $\mathbb{E}\mu_2(P(\theta))$ is.

Proposition 3.3. *For N even, let $a \in \left(1, 1 + \frac{4}{N-4}\right]$, i.e. $N - 2[z^*] - 1 = 1$. Then there exists an $\overline{\varepsilon} > 0$ such that $\mathbb{E}\mu_2(P(\theta)) < \mu_2(P(\mathbb{E}\theta)) < \mu_2(\mathbb{E}P(\theta))$, where θ is a random variable with distribution Θ, with Θ an arbitrary non-degenerate discrete probability distribution with support on a subset of $[0, \overline{\varepsilon}]$.*

Proof of the first inequality of Proposition 3.3.

Restricting $a \in \left(1, 1 + \frac{4}{N-4}\right]$ is equivalent to requiring that $z^* \in \left[\frac{1}{2}N - 1, \frac{1}{2}N\right)$. Since $\frac{1}{2}N - 1$ is an integer for N even, this means that $[z^*] = \frac{1}{2}N - 1$ and thus that $N - 2[z^*] - 1 = 1$. This leads to $C(\varepsilon)=\binom{N}{[z^*]+1}=\binom{N}{\frac{1}{2}N}$, being a constant for all $\varepsilon\in[0,1]$. Thus $\mu_2(P(\varepsilon))=\frac{D(\varepsilon)}{\binom{N}{\frac{1}{2}N}+2D(\varepsilon)}$ and

$$\lim_{\varepsilon\downarrow 0}\mu_2''(P(\varepsilon))=\frac{\binom{N}{\frac{1}{2}N}D''(0)-4\left[D'(0)\right]^2}{\binom{N}{\frac{1}{2}N}^2}=\frac{\binom{N}{\frac{1}{2}N}\cdot 2\left[\binom{N}{\frac{1}{2}N+1}+\binom{N}{\frac{1}{2}N+2}\right]-4\binom{N}{\frac{1}{2}N+1}^2}{\binom{N}{\frac{1}{2}N}^2}.$$

We focus on the numerator $\binom{N}{\frac{1}{2}N}\cdot 2\left\{\binom{N}{\frac{1}{2}N+1}+\binom{N}{\frac{1}{2}N+2}\right\}-4\binom{N}{\frac{1}{2}N+1}^2$ which is equal to

$$\frac{(N!)^2}{\left(\frac{1}{2}N+1\right)!\left(\frac{1}{2}N\right)!\left(\frac{1}{2}N-1\right)!\left(\frac{1}{2}N-2\right)!}\cdot$$
$$\left\{\frac{2}{\frac{1}{2}N\left(\frac{1}{2}N-1\right)}+\frac{2}{\left(\frac{1}{2}N+2\right)\cdot\frac{1}{2}N}-\frac{4}{\left(\frac{1}{2}N+1\right)\left(\frac{1}{2}N-1\right)}\right\}.$$

To determine the sign of this expression we look at the part in brackets

$$\frac{2}{\frac{1}{2}N\left(\frac{1}{2}N-1\right)}+\frac{2}{\left(\frac{1}{2}N+2\right)\cdot\frac{1}{2}N}-\frac{4}{\left(\frac{1}{2}N+1\right)\left(\frac{1}{2}N-1\right)}=\frac{-16}{(N+2)(N+4)N}<0,\ \forall N\in\mathbb{N}.$$

Thus we have that $\mu_2''(P(0))<0$ and by a continuity argument thus $\mu_2''(P(\varepsilon))<0,\ \varepsilon\in(0,\overline{\varepsilon}_1)$, for some $\overline{\varepsilon}_1>0$.[3] Applying Jensen's inequality to the concave function $\mu_2(P(\varepsilon))$ results in the first part of the proposition. □

Proof of the second inequality of Proposition 3.3.

Restricting $a\in\left(1,1+\frac{4}{N-4}\right]$ is equivalent to requiring that $[z^*]=\frac{1}{2}N-1$. Under this condition we first prove that $\mu_2(P(\mathbb{E}\theta))<\mu_2(\mathbb{E}P(\theta))$ for θ a random variable with a Bernoulli distribution Θ, i.e. $\Pr(\varepsilon_t=(1+\alpha)\varepsilon)=\lambda$, $\Pr(\varepsilon_t=\varepsilon)=1-\lambda$, $\alpha>0$. For such θ, $\mu_2(P(\mathbb{E}\theta))<\mu_2(\mathbb{E}P(\theta))$ yields

$$\frac{p(\lambda\alpha\varepsilon+\varepsilon)}{p(\lambda\alpha\varepsilon+\varepsilon)+p'(\lambda\alpha\varepsilon+\varepsilon)}<\frac{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)}{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)+p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)}.$$

Since both the numerators and denominators are strictly positive, cross multiplying does not alter the sign and we obtain

$$\begin{aligned}&p(\lambda\alpha\varepsilon+\varepsilon)\left\{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)+p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)\right\}\\ <\ &\left\{p(\lambda\alpha\varepsilon+\varepsilon)+p'(\lambda\alpha\varepsilon+\varepsilon)\right\}\left\{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)\right\}.\end{aligned}$$

When we expand this expression and deduct the common (positive) term $p(\varepsilon)p(\lambda\alpha\varepsilon+\varepsilon)-\lambda p(\varepsilon)p(\lambda\alpha\varepsilon+\varepsilon)+\lambda p((1+\alpha)\varepsilon)p(\lambda\alpha\varepsilon+\varepsilon)$ from both sides of the inequality and then rearrange the terms, we get

$$p(\lambda\alpha\varepsilon+\varepsilon)\left\{p'(\varepsilon)-\lambda p'(\varepsilon)+\lambda p'((1+\alpha)\varepsilon)\right\}<p'(\lambda\alpha\varepsilon+\varepsilon)\left\{p(\varepsilon)-\lambda p(\varepsilon)+\lambda p((1+\alpha)\varepsilon)\right\}.$$

[3]The maximal value for $\overline{\varepsilon}_1$ is the smallest positive root of the numerator of $\mu_2''(P(\varepsilon))$ or 1 in case $\mu_2''(P(\varepsilon))$ has no root larger than 0.

Now, we use the explicit expressions for $p(\varepsilon)$ and $p'(\varepsilon)$ and, since ε is small, we focus on the lowest power of ε on both sides of the inequality sign. We replace all terms $1-\varepsilon \approx 1$ by 1 and divide by $\binom{N}{[z^*]+1}\binom{N}{N-[z^*]}$ and get

$$\begin{aligned} & \varepsilon^{[z^*]+1}(\lambda\alpha\varepsilon+\varepsilon)^{N-[z^*]} - \lambda\varepsilon^{[z^*]+1}(\lambda\alpha\varepsilon+\varepsilon)^{N-[z^*]} + \lambda(1+\alpha)^{[z^*]+1}\varepsilon^{[z^*]+1}(\lambda\alpha\varepsilon+\varepsilon)^{N-[z^*]} \\ < \; & \varepsilon^{N-[z^*]}(\lambda\alpha\varepsilon+\varepsilon)^{[z^*]+1} - \lambda\varepsilon^{N-[z^*]}(\lambda\alpha\varepsilon+\varepsilon)^{[z^*]+1} + \lambda(1+\alpha)^{N-[z^*]}\varepsilon^{N-[z^*]}(\lambda\alpha\varepsilon+\varepsilon)^{[z^*]+1}. \end{aligned}$$

We now divide both sides by the common term[4] $(\lambda\alpha+1)^{[z^*]+1}\varepsilon^{N-1}$, resulting in

$$(\lambda\alpha+1)^{N-2[z^*]-1} - \lambda(\lambda\alpha+1)^{N-2[z^*]-1} + \lambda(1+\alpha)^{[z^*]+1}(\lambda\alpha+1)^{N-2[z^*]-1} < 1-\lambda+\lambda(1+\alpha)^{N-[z^*]}.$$

since $[z^*] = \frac{1}{2}N-1$, this expression reduces to

$$(\lambda\alpha+1)\left\{1-\lambda+\lambda(1+\alpha)^{\frac{N}{2}}\right\} < 1-\lambda+\lambda(1+\alpha)^{\frac{N}{2}+1}.$$

Although $(\lambda\alpha+1)(1-\lambda) > 1-\lambda$, the second terms on both sides compensate this, as follows

$$(\lambda\alpha+1)(1-\lambda)-(1-\lambda) < \lambda(1+\alpha)^{\frac{N}{2}+1} - \lambda(1+\alpha)^{\frac{N}{2}}(\lambda\alpha+1)$$

which, after some rewriting, gives

$$\lambda\alpha(1-\lambda) < (1+\alpha)^{\frac{N}{2}}\lambda\alpha(1-\lambda) \;\Leftrightarrow\; 1 < (1+\alpha)^{\frac{N}{2}},$$

a true statement for any $\alpha > 0$ and $N \in \mathbb{N}^+$. Thus, we have proven the result for $\varepsilon \in (0, \overline{\varepsilon}_2)$, for some small $\overline{\varepsilon}_2 > 0$, and a Bernoulli distribution Θ. An induction argument suffices to show that it holds for any random variable θ with a non-degenerate discrete probability distribution Θ with support on a subset of $[0, \overline{\varepsilon}_3]$ for some small $\overline{\varepsilon}_3 > 0$. Finally, in Proposition 3.3, $\overline{\varepsilon} = \min\{\overline{\varepsilon}_1, \overline{\varepsilon}_2, \overline{\varepsilon}_3\}$. □

The following proposition states the result for small populations with an odd number of agents. Then all three numbers coincide and are equal to a half.

Proposition 3.4. *For N odd, let $a \in \left(1, 1+\frac{2}{N-3}\right]$, i.e. $N-2[z^*]-1=0$. Then it holds that $\mu_2(P(\mathbb{E}\theta)) = \mathbb{E}\mu_2(P(\theta)) = \mu_2(\mathbb{E}P(\theta)) = \frac{1}{2}$, where θ is a random variable with distribution Θ, with Θ an arbitrary non-degenerate discrete probability distribution with support on a subset of $[0,1]$.*

[4] $(\lambda\alpha+1)^{[z^*]+1}$ is the lowest power of $(\lambda\alpha+1)$, since $[z^*]+1 = \frac{1}{2}N$ and $N-[z^*] = \frac{1}{2}N+1$

Proof.

Restricting $a \in \left(1, 1+\frac{2}{N-3}\right]$ is equivalent to requiring that $z^* \in \left[\frac{1}{2}N - \frac{1}{2}, \frac{1}{2}N\right)$. Since $\frac{1}{2}N - \frac{1}{2}$ is an integer for N odd, this means that $[z^*] = \frac{1}{2}N - \frac{1}{2}$ and thus that $N - 2[z^*] - 1 = 0$. This results in $C(\varepsilon) = 0$ for all $\varepsilon \in [0,1]$, what comes down to $\lim_{\varepsilon \downarrow 0} \mu_2(P(\varepsilon)) = \lim_{\varepsilon \uparrow 1} \mu_2(P(\varepsilon)) = \mu_2(P(\varepsilon)) = \frac{D(\varepsilon)}{C(\varepsilon)+2D(\varepsilon)} = \frac{1}{2} \; \forall \varepsilon \in (0,1)$. □

Since the result of Proposition 3.4 also applies to fixed mutation rate $\varepsilon = \mathbb{E}\theta$, our result is different from KMR's Theorem 3. The explanation is that $[z^*] = \frac{1}{2}N - \frac{1}{2}$ implies that the assertion in KMR's proof of their Theorem 3, that there will always exist an integer α, $z^* < \alpha \leq \frac{1}{2}N$ is incorrect. For $[z^*] = \frac{1}{2}N - \frac{1}{2}$, it takes the same number of mutations to upset the efficient equilibrium E_1, as it takes to upset E_2. A symmetry argument leads to the conclusion that, for $[z^*] = \frac{1}{2}N - \frac{1}{2}$, the invariant distribution in KMR should be $\mu = (\mu_1, \mu_2) = \left(\frac{1}{2}, \frac{1}{2}\right)$. Thus, the above result shows that the restriction to even N in KMR does have an impact, thereby contradicting the assertion in Rhode and Stegeman (1996) that it does not.

In order to show the different regions in the (a, N) plane, we present figure 3.1 for N even. For N odd, the figure looks very similar to this figure. From the figure we see that for either N large or a large, we are in the region where Theorem 3.1 applies. Only for small N and a not too large, we are in the region where Propositions 3.3 and 3.4 are applicable.

Figure 3.1: The different regions in the (a, N) plane for N even.

To obtain some insight in the magnitude of the maximum value of the upperbound $\overline{\varepsilon}$ in Theorem 3.1, we present the value for ε_1 in the proof of the Theorem in table 3.2 (to be precise, Proposition 3.3 applies to the case $N = 10$ and $a = 1.5$). For different combinations (a, N), it presents $[z^*]$ and ε_1. From this table we see that ε_1 is far away from 0, indicating that the

$N\backslash a$	1.5	2	5
10	4; 1.31	3; 0.58	2; 0.69
20	8; 0.51	7; 0.57	4; 0.73
100	40; 0.57	33; 0.64	17; 0.81

Table 3.2: The values for $[z^*]$ (1^{st} entry) and the values for ε_1 (2^{nd} entry) as a function of N and a.

interval $[0, \varepsilon_1]$ is large. Furthermore we note that, for the cases where Theorem 3.1 applies, the value of ε_1 is increasing in a.

An upperbound for the second inequality in Theorem 3.1 (ε_4 resp. ε_5) can be derived from solving equation (3.3) if Θ is a Bernoulli distribution and from solving an expression similar to (3.3) for a general non-degenerate discrete probability distribution Θ.

4. Concluding Remarks.

Thus far we have shown that, given that the payoff from the risk dominant equilibrium is sufficiently high, or, equivalently, when the population N is sufficiently large, the expected time the evolutionary system with best reply dynamics and a stochastic mutation rate spends in the risk dominated equilibrium is systematically underestimated when one calculates this time in the conventional way by assuming a fixed mutation rate that is equal to the expected mutation rate. The crucial feature in reaching this conclusion is the minimum number of mutations needed in order to escape the basin of attraction of an equilibrium (the cost minimizing i-trees, $i = 0, N$), as is the case in KMR and Young (1993). Since this minimum number of mutations does not change if we assume Darwinian dynamics instead of best reply dynamics as the deterministic component of the dynamics, our results go through for general deterministic Darwinian dynamics.

References

Bergin, J. and B.L. Lipman (1996). Evolution with state-dependent mutations. *Econometrica 64*, 943–956.

Cooper, R.W. and A. John (1988). Coordinating coordination failures in keynesian models. *Quarterly Journal of Economics 103*, 441–463.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*, 1047–1071.

Kandori, M., G.J. Mailath, and R. Rob (1993). Learning, mutation and long run equilibria in games. *Econometrica 61*, 29–56.

Rhode, P. and M. Stegeman (1996). A comment on "learning, mutation, and long-run equilibria in games". *Econometrica 64*, 443–449.

Young, H.P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.

A. Explicit Expressions for the Derivatives of $C(\varepsilon)$ and $D(\varepsilon)$.

Differentiating (3.1) with respect to ε, using $\frac{d}{d\varepsilon}\left(\frac{j+1}{(1-\varepsilon)^2}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j\right) = \frac{2(j+1)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j + \frac{j(j+1)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j-1}$, $j \geq 1$, yields $C'(\varepsilon) = \sum_{j=1}^{N-2[z^*]-2}\binom{N}{j+[z^*]+1}\frac{j}{(1-\varepsilon)^2}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j-1} = \sum_{j=0}^{N-2[z^*]-3}\binom{N}{j+[z^*]+2}\frac{j+1}{(1-\varepsilon)^2}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j$ and $C''(\varepsilon) = \binom{N}{[z^*]+2}\frac{2}{(1-\varepsilon)^3} + \sum_{j=1}^{N-2[z^*]-3}\binom{N}{j+[z^*]+2}\left\{\frac{2(j+1)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j + \frac{j(j+1)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j-1}\right\} = \binom{N}{[z^*]+2}\frac{2}{(1-\varepsilon)^3} + \sum_{j=0}^{N-2[z^*]-4}\binom{N}{j+[z^*]+3}\left\{\frac{2(j+2)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j+1} + \frac{(j+1)(j+2)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j\right\}$.

Similarly, differentiating (3.2) with respect to ε yields $D'(\varepsilon) = \sum_{j=N-2[z^*]-2}^{N-[z^*]-2}\binom{N}{j+[z^*]+2}\frac{j+1}{(1-\varepsilon)^2}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j$ and $D''(\varepsilon) = \sum_{j=N-2[z^*]-3}^{N-[z^*]-3}\binom{N}{j+[z^*]+3}\left\{\frac{2(j+2)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j+1} + \frac{(j+1)(j+2)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j\right\}$.

From these expressions and $\lim_{j\to 0}\lim_{\varepsilon\to 0}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j = \lim_{\varepsilon\to 0}\lim_{j\to 0}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j = 1$, we see that

$C(0) = \binom{N}{[z^*]+1}$, when $N - 2[z^*] - 2 \geq 0$, i.e. when $N - 2[z^*] - 1 \geq 1$, and $C(\varepsilon) = 0$ for all $\varepsilon \in [0,1]$, when $N - 2[z^*] - 1 < 1$,

$C'(0) = \binom{N}{[z^*]+2}$, when $N - 2[z^*] - 3 \geq 0$, i.e. when $N - 2[z^*] - 2 \geq 1$, and $C'(0) = 0$, when $N - 2[z^*] - 3 < 0$,

$C''(0) = 2\binom{N}{[z^*]+2} + 2\binom{N}{[z^*]+3}$ when $N - 2[z^*] - 4 \geq 0$, i.e. when $N - 2[z^*] - 3 \geq 1$, and $C''(0) = 2\binom{N}{[z^*]+2}$, when $C'(0) > 0$, i.e. when $N - 2[z^*] - 2 \geq 1$, and $C''(0) = 0$, when $N - 2[z^*] - 2 < 1$.

Similarly, we see that $D(0) = 0$, when $N - 2[z^*] - 1 \geq 1$, $D'(0) = 0$, when $N - 2[z^*] - 2 \geq 1$, and $D''(0) = 0$, when $N - 2[z^*] - 3 \geq 1$.Therefore, $\mu_2'(0) = \frac{C(0)D'(0) - D(0)C'(0)}{[C(0)+2D(0)]^2}$ is equal to 0 whenever $N - 2[z^*] - 2 \geq 1$.

When $N - 2[z^*] - 1 = 1$, the derivative $D'(\varepsilon)$ sums from $j = 0$ and consequently $D''(\varepsilon)$ becomes $\binom{N}{[z^*]+2}\frac{2}{(1-\varepsilon)^3} + \sum_{j=1}^{N-[z^*]-2}\binom{N}{j+[z^*]+2}\left\{\frac{2(j+1)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j + \frac{j(j+1)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j-1}\right\} = \binom{N}{[z^*]+2}\frac{2}{(1-\varepsilon)^3} + \sum_{j=0}^{N-[z^*]-3}\binom{N}{j+[z^*]+3}\left\{\frac{2(j+2)}{(1-\varepsilon)^3}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^{j+1} + \frac{(j+1)(j+2)}{(1-\varepsilon)^4}\left(\frac{\varepsilon}{(1-\varepsilon)}\right)^j\right\}$. Thus, when $N - 2[z^*] - 1 = 1$, it follows that $D'(0) = \binom{N}{[z^*]+2} = \binom{N}{\frac{1}{2}N+1}$ and $D''(0) = 2\binom{N}{[z^*]+2} + 2\binom{N}{[z^*]+3} = 2\binom{N}{\frac{1}{2}N+1} + 2\binom{N}{\frac{1}{2}N+2}$.